# Non Linear Relationship between BMI and Sepsis among ICU Patients: Analysis of the MIMIC-III Real World Database

MK LINTU[1], AMITHA PURANIK[2], KALESH M KARUN[3], DAVID RAJ MICHEAL[4], ASHA KAMATH[5]

## ABSTRACT

**Introduction:** The relationship between occurrence of sepsis and Body Mass Index (BMI) has been studied by treating BMI as a categorical variable based on the standard categories developed by the World Health Organisation (WHO). This technique masks the actual form of the relationship that exist between BMI and sepsis.

**Aim:** To determine the shape of the association between BMI and probability of sepsis after adjusting the effect of gender, age, ethnicity and co-morbidities among critically-ill patients and to identify an appropriate technique to elucidate this association.

**Materials and Methods:** The study was based on the data extracted from the Medical Information Mart for Intensive Care III (MIMIC-III) database, a real-world dataset. A robust Poisson regression model and a Restricted Cubic Spline (RCS) regression model were applied to model the data.

**Results:** The study revealed a non linear relationship between BMI and occurrence of sepsis. From the RCS plot, an increased risk of sepsis was observed among underweight and obese males and a similar pattern was observed among those with hypertension. BMI, age, and hypertension were found to be significant in the robust Poisson model (RR 1.01, 95% CI 1.00-1.02, RR 1.01, 95% CI 1.00-1.02, and RR 0.76, 95% CI 0.69-0.84, p<0.05, respectively).

**Conclusion:** A non linear relationship was observed between BMI and occurrence of sepsis after adjusting for other factors in the critically-ill patients. Female patients in both underweight and obese category had lower occurrence of sepsis as compared to those with normal BMI. It is also worth noting that the status of diabetes did not display any association with BMI.

**Keywords:** Medical information mart for intensive care, Poisson model, Restricted cubic spline

## INTRODUCTION

Sepsis is a significant cause of mortality in Intensive Care Units (ICU) and has become an important public health problem. The early diagnosis of sepsis helps in reducing the risk and disease burden. Obesity is associated with the increased risk of infections including sepsis. There are studies suggesting clinical and pathophysiological connections between the obesity and development of sepsis [1-3]. Few studies have reported that obesity leads to diabetes, which further increases the risk of infection [4,5], whereas some studies have reported obesity as an independent risk factor associated with sepsis [2,6]. A large number of studies modeled the association between BMI and the risk of outcome by treating BMI as a categorical variable with different levels such as underweight, normal, overweight, and obese [7,8]. This approach fails to elucidate the non linear association that exists between BMI and risk of sepsis. Treating the BMI as continuous variable and choosing the appropriate model that account for the non-linearity improves the predictive ability of the model.

The present study aimed to determine the shape of the association between BMI and probability of sepsis after adjusting for the effect of gender, age, ethnicity and co-morbidities and identify a suitable model to fit this association on the data extracted from the MIMIC-III database.

## MATERIALS AND METHODS

The current study was conducted using the publicly available dataset MIMIC-III, maintained by the Massachusetts Institute of Technology (MIT)'s Laboratory for Computational Physiology which includes all the patients admitted to the ICU at the Beth Israel Deaconess Medical Center in Boston, Massachusetts from 2001 to 2012.

The data extraction process was done using Python version 3.7.6. The database consisted of 62722 ICU stay records of 49785 patients. The Ninth Revision, International Classification of Diseases (ICD-9) diagnosis codes for 'Sepsis' and 'Severe sepsis' [9] were used to identify the sepsis patients. A total of 12154 patients were considered in the present study based on the inclusion criteria: (i) the first ICU admission among the multiple ICU admissions; (ii) information on height, weight, age, and gender was not missing; (iii) age between 17 to 89 years; and (iv) BMI between 15 to 50 kg/m².

Out of the 12154 patients considered for further analysis, 346 patients belonged to the underweight (<18.5) category, 3667 patients had normal (18.5-24.9) weight, 4131 patients belonged to the overweight (25.0-29.9) category, and 4010 patients were found to be obese (>30). There were 5798 patients with hypertension, and 498 patients were diabetic.

A robust Poisson regression model was fit to estimate the Risk Ratio (RR) [10] of sepsis with respect to the associated factors. The model was chosen according to the study design as it was found to be the best regression model for cohort studies, providing a narrow confidence interval compared to any other model. Robust Poisson regression gives unbiased estimates of the relationship and tends to be robust to various model misspecifications [11,12].

Since there was evidence from the literature regarding the non linear relationship between BMI and RR of sepsis [1], a RCS regression model was fit using robust Poisson to visualise the shape and to estimate the association in the present data. Splines can be used in the regression models when the assumption of linearity is violated. As a solution, the explanatory variable is transformed by the splines by splitting the range into small intervals using knots. Separate curves are then fit in each interval which results in a smooth and continuous curve. The association between the independent variable and the

outcome variable in between any two knots is modelled using a polynomial of order three [13-15]. As the number of knots has an important role in the model, the performance of RCS model with different knots were fitted in the present study and the optimum number of knots was identified using the Akaike's Information Criterion (AIC). The locations of the knots were prespecified in quantiles of BMI. The analysis was done using the glm package in R version 4.0.2. The performance of RCS was compared to the robust Poisson model, to see the improvement in prediction.

## STATISTICAL ANALYSIS

Continuous variables were summarised using the mean {±standard deviation (SD)} and categorical variables using the frequency (percentage). Independent samples t-test (for continuous variables) and Chi-square test (for categorical variables) were used to compare the baseline characteristics between sepsis and non sepsis patients.

## RESULTS

A total of 12154 ICU admitted patients from the MIMIC registry were considered for the present study. The prevalence of sepsis was 15.56% (1891 out of 12154). There were 7243 (59.59%) male patients and 4911 (40.41%) female patients. The mean±SD age was 63.04±15.65 years. The descriptive statistics are given in [Table/Fig-1]. A significant difference (p <0.05) was found in age, hypertension, and diabetes between the sepsis and non sepsis groups.

| Variables | Sepsis | Non sepsis | p-value |
|---|---|---|---|
| n (%) | 1891 (15.56) | 10263 (84.44) | |
| Age in years (±SD) | 63.89±15.46 | 62.88±15.68 | 0.017* |
| BMI in kg/m² (±SD) | 28.46±6.52 | 28.18±6.19 | 0.092 |
| Gender-male (%) | 1128 (59.65) | 6115 (59.58) | 0.959 |
| Hypertension-yes (%) | 793 (41.94) | 5005 (48.77) | <0.001* |
| Diabetes-yes (%) | 98 (5.18) | 400 (3.90) | 0.006* |
| **Ethnicity** | | | |
| Asian (%) | 233 (81.20) | 54 (18.80) | |
| African American (%) | 945 (83.30) | 190 (16.70) | |
| Hispanic/Latino (%) | 395 (84.00) | 75 (16.00) | 0.270 |
| White (%) | 7625 (84.50) | 1394 (15.50) | |
| Others (%) | 1065 (85.70) | 178 (14.30) | |

[Table/Fig-1]: Baseline characteristics of the patients.
*Significant (p<0.05); SD-standard deviation

To visualise the shape of association between BMI and probability of sepsis, unadjusted RCS model was fit [Table/Fig-2]. The non linearity was observed between two factors and hence the model was further adjusted by adding other covariates. The shape of the association between BMI and probability of sepsis was non-linear in the unadjusted model and a similar shape of association was also observed when other covariates were included in the model. An increased risk was observed for the underweight and obese category compared to the normal BMI category [Table/Fig-2]. Similarly, the shape of association between BMI and probability of sepsis was also found to be non linear among different categories of gender, diabetes, hypertension [Table/Fig-3], and ethnicity [Table/Fig-4].

[Table/Fig-3a,b] depicts the association among the male and female populations respectively. Among the male patients, an increased risk of sepsis was observed in the underweight and obese category compared to the normal BMI category, however
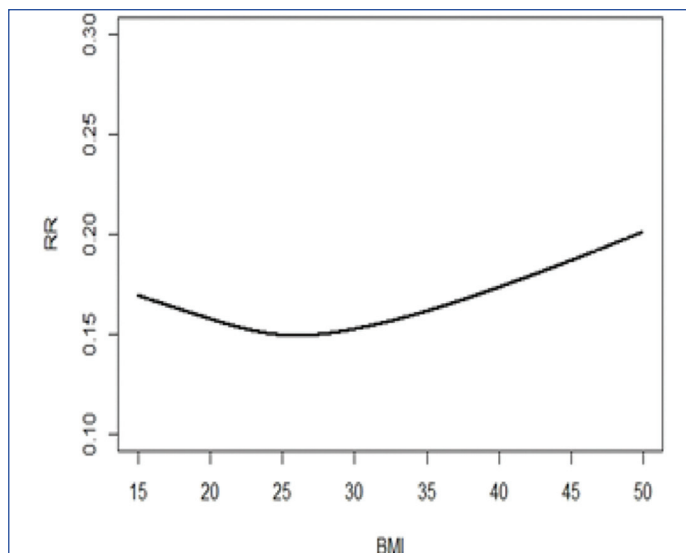


[Table/Fig-2]: Association of sepsis with BMI-Restricted Cubic Spline (RCS) regression model.
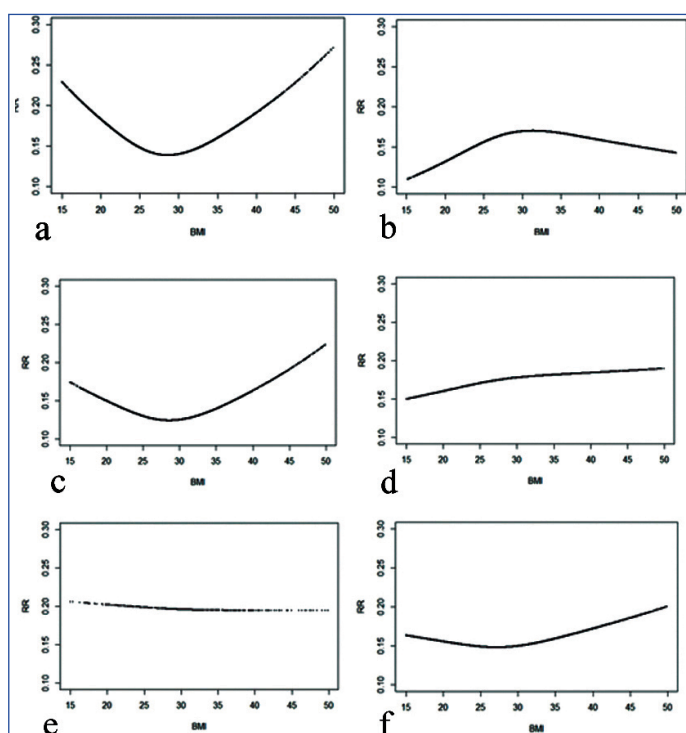


[Table/Fig-3]: Association of sepsis with BMI with respect to various factors-Restricted Cubic Spline (RCS) regression model. a) male; b) female; c) hypertensive; d) non hypertensive; e) diabetic; f) non diabetic.
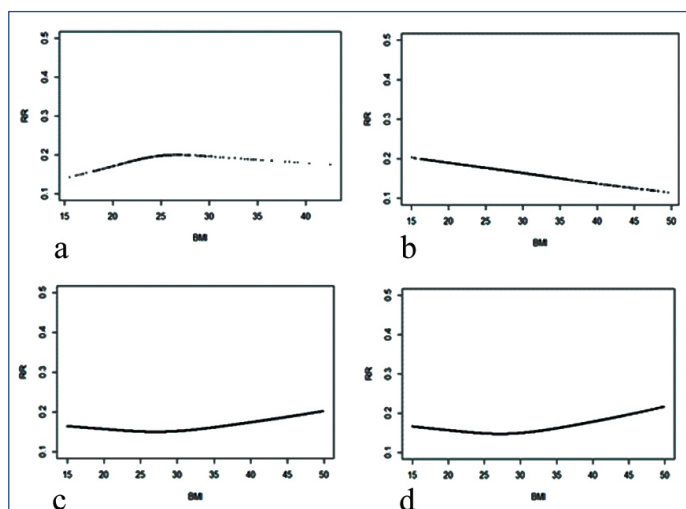


[Table/Fig-4]: Association of sepsis with BMI among different ethnicities-restricted cubic spline regression model. (a) Asian, (b) African American, (c) Hispanic/Latino, (d) White.

the females had lower risk. Underweight and overweight patients with hypertension had more risk compared to normal category [Table/Fig-3c,d]. Association of diabetic and non diabetic patients is depicted in [Table/Fig-3e,f]. Among the non diabetic patients, risk of sepsis was high in the obese category compared to the other two categories. However, the non linearity was not observed among those with diabetes.

The robust Poisson and RCS regression models were adjusted for age, gender, hypertension, diabetes, and ethnicity. The adjusted RCS model revealed a non linear association between the probability of sepsis and BMI. The RCS model were fitted with three, four, five, and six knots. The AIC values corresponding to the adjusted and unadjusted models are given in [Table/Fig-5].

| Model | AIC | |
| --- | --- | --- |
| | Unadjusted model | Adjusted model |
| Robust Poisson | 10820.06 | 10786.82 |
| RCS with 3 knots | 10819.22 | **10785.96** |
| RCS with 4 knots | 10821.04 | 10787.81 |
| RCS with 5 knots | 10822.82 | 10789.51 |
| RCS with 6 knots | 10821.67 | 10788.18 |

[Table/Fig-5]: Performance of the fitted models.
Models were adjusted for age, female gender, hypertension, diabetes, and ethnicity

The RCS model with three knots, adjusted for age, gender, hypertension, diabetes, and ethnicity was found to be best fit in the present study as it had the least AIC. The results of RCS model with three knots model along with robust Poisson model is provided in [Table/Fig-6].

BMI, age, and hypertension were found to be significant in the robust Poisson model (RR 1.01, 95% CI 1.00-1.02, RR 1.01, 95% CI 1.00-1.02, and RR 0.76, 95% CI 0.69-0.84, p<0.05 respectively). Those with diabetes had 22% more risk compared to those without (RR 1.22, 95% CI 0.99-1.49, p=0.05). Gender and ethnicity were not significant in either of the models. With reference to Asians, all other ethnic groups were at low risk (p>0.05) of sepsis in both the models.

The relationship between standard cut-offs of BMI (<18.5, 18.5 to <25, 25 to <30, and >30) and probability of sepsis was also examined using robust Poisson regression, which is the usual approach to handle BMI. [Table/Fig-7] gives the RR

| Model | Variables | Risk Ratio | 95% CI | p-value |
| --- | --- | --- | --- | --- |
| Robust poisson | BMI | 1.01 | (1.00, 1.02) | 0.02* |
| | Age | 1.01 | (1.00, 1.02) | <0.001* |
| | Gender | | | |
| | Female | Reference | | |
| | Male | 1.00 | (0.91, 1.10) | 0.78 |
| | Hypertension | | | |
| | No | Reference | | |
| | Yes | 0.76 | (0.69, 0.84) | <0.001* |
| | Diabetes | | | |
| | No | Reference | | |
| | Yes | 1.22 | (0.99, 1.49) | 0.05* |
| | Ethnicity | | | |
| | Asian | Reference | | |
| | African American | 0.86 | (0.64,1.18) | 0.34 |
| | Hispanic/Latino | 0.85 | (0.60,1.21) | 0.36 |
| | White | 0.79 | (0.61,1.05) | 0.07 |
| | Others | 0.75 | (0.56,1.03) | 0.09 |

| | Variable | Risk ratio | 95% CI | p-value |
| --- | --- | --- | --- | --- |
| Restricted cubic spline with three knots | BMI | 0.99 | (0.97, 1.01) | 0.44 |
| | BMI' | 1.02 | (0.99, 1.04) | 0.08 |
| | Age | 1.00 | (1.00, 1.00) | <0.001* |
| | Gender | | | |
| | Female | Reference | | |
| | Male | 1.02 | (0.92, 1.11) | 0.69 |
| | Hypertension | | | |
| | No | Reference | | |
| | Yes | 0.76 | (0.69, 0.84) | <0.001* |
| | Diabetes | | | |
| | No | Reference | | |
| | Yes | 1.22 | (0.98, 1.48) | 0.05* |
| | Ethnicity | | | |
| | Asian | Reference | | |
| | African American | 0.87 | (0.64, 1.18) | 0.36 |
| | Hispanic/Latino | 0.86 | (0.60,1.22) | 0.38 |
| | White | 0.80 | (0.61,1.05) | 0.10 |
| | Others | 0.76 | (0.56,1.04) | 0.08 |

[Table/Fig-6]: Results from the robust Poisson and Restricted Cubic Spline (RCS) model with three knots.
*significant (p<0.05)

of the different factors associated with sepsis by treating BMI as a categorical variable. It is worthy to note that the obese category of BMI showed a statistical significant relationship with sepsis.

| Variable | Risk ratio | 95% CI | p-value |
| --- | --- | --- | --- |
| **BMI** | | | |
| Underweight (<18.5) | 1.11 | (0.87,1.41) | 0.41 |
| Normal (18.5-24.9) | Reference | | |
| Overweight (25.0-29.9) | 0.95 | (0.85,1.06) | 0.34 |
| Obese (>30) | 1.11 | (1.01,1.23) | 0.04* |
| Age | 1.00 | (1.00,1.01) | <0.001* |
| **Gender** | | | |
| Female | Reference | | |
| Male | 1.02 | (0.92,1.11) | 0.75 |
| **Hypertension** | | | |
| No | Reference | | |
| Yes | 0.77 | (0.70,0.84) | <0.001* |
| **Diabetes** | | | |
| No | Reference | | |
| Yes | 1.22 | (0.99,1.49) | 0.05* |
| **Ethnicity** | | | |
| Asian | Reference | | |
| African American | 0.87 | (0.66,1.15) | 0.32 |
| Hispanic/Latino | 0.86 | (0.62,1.18) | 0.35 |
| White | 0.80 | (0.62,1.02) | 0.07 |
| Others | 0.76 | (0.58,1.00) | 0.05* |

[Table/Fig-7]: Risk Ratios (RR) of the different factors associated with sepsis by treating BMI as a categorical variable using robust Poisson model.
*significant (p<0.05)

Similar to the previous models, age and hypertension were significant factors (RR 1.00, 95% CI 1.00-1.01 and RR 0.77, 95% CI 0.70-0.84, p<0.05). Diabetes patients were at more risk of sepsis compared to non diabetic patients (RR 1.22, 95% CI 0.99-1.49, p=0.05). Underweight (RR 1.11, 95% CI 0.87-1.41, p=0.41) and obese (RR 1.11, 95% CI 1.01-1.23, p=0.04)

patients were at high risk with reference to the patients with normal BMI.

## DISCUSSION

The present study used the data extracted from the MIMIC-III database to determine the nature of association between BMI and occurrence of sepsis after adjusting for gender, age, ethnicity and co-morbidities such as hypertension and diabetes. The aim of the study was mainly to identify a suitable statistical approach that will not only reveal the actual shape of association but also improve the predictive ability of the model. A RCS regression model was fit using robust Poisson to visualise the shape and to estimate the association in the present data.

RCS has been efficiently used to model non-linear relationships between independent variables and outcomes. By defining k knots, the numerical values of the explanatory variables are transformed into k+3 new variables. The coefficients are calculated using conventional techniques for fitting multivariable regression models; however, they are usually not interpreted. Hence, it is typical to portray the findings graphically, showing the estimated shape of the spline function, in order to avoid the challenges associated with interpreting the spline regression coefficients [1,16,17].

The results of the present study revealed a non linear association between BMI and sepsis. The visual representation of RR with BMI based on RCS model displayed a higher occurrence of sepsis in the higher BMI range.

A study investigated the relationship of BMI on short-and long-term mortality in sepsis patients based on MIMIC database using Cox proportional hazards regression model by dividing the patients into four groups according to BMI standard categories. Overweight and obese patients had lower risk of both short and long-term mortality than those with normal weight [18]. A similar result was reported by Li S et al., based on multivariable logistic regression model [19]. A non-linear regression analyses using the Cox model to determine the effect of BMI on critically-ill patients with sepsis by Lin S et al., reported lowest 28-day mortality in the BMI range of 30-40 kg/m$^2$ [20].

BMI was not found to be a significant factor in the present study. The possible reason for this could be either not accounting for the other potential factors associated with sepsis or not treating BMI in categorical scale. The present work accounted for the cohort design of MIMIC database and the non linear nature of association between BMI and sepsis. The goal was to highlight the importance of treating BMI as a continuous variable rather than as a categorical variable, as this is beneficial for a better prediction of sepsis.

In the case of RCS model, the interpretation of coefficients are not always easy [21]. Royston P described the use of cubic smoothing splines to assess how well the functions of a given predictor fit the data [22]. The present study fit a cubic spline model for independent variable and tested whether it improves the prediction compared to a linear model. The goodness of fit of the RCS and linear model are assessed using AIC and it is observed that RCS is the best model as it has the least AIC. It is observed from the RCS model that it will be more appropriate for the populations with larger fractions of underweight individuals.

### Limitation(s)

As the present study was based on data from the MIMIC database, which is a secondary source, there were few limitations, such as selection bias due to a large proportion of missing observations in variables of interest and the inability to establish the cause-effect relationship between BMI and sepsis. Also, the sample size of underweight patients was 10 times smaller than that of other BMI categories, which might have an effect on generalisability. However, to our knowledge, this was the first study that demonstrates the usage of RCS regression to fit the association between BMI and sepsis on the MIMIC-III data.

## CONCLUSION(S)

In the present study, a non linear relationship was observed between BMI and occurrence of sepsis after adjusting for other factors in the critically-ill patients. The visual representation of RR with BMI based on RCS model displayed a higher occurrence of sepsis in the higher BMI range. Female patients in both underweight and obese category had lower occurrence of sepsis as compared to those with normal BMI. It is also worth noting that a non linear relationship was not found among those with diabetes, although the risk of sepsis was higher among obese non diabetic patients.

## REFERENCES

[1] Winter-Jensen M, Afzal S, Jess T, Nordestgaard BG, Allin KH. Body mass index and risk of infections: A Mendelian randomization study of 101,447 individuals. Eur J Epidemiol. 2020;35(4):347-54.

[2] Wang HE, Griffin R, Judd S, Shapiro NI, Safford MM. Obesity and risk of sepsis: A population-based cohort study. Obesity. 2013;21(12):E762-69.

[3] Vachharajani V, Russell JM, Scott KL, Conrad S, Stokes KY, Tallam L, et al. Obesity exacerbates sepsis-induced inflammation and microvascular dysfunction in mouse brain. Microcirculation. 2005;12(2):183-94.

[4] Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB. Annual deaths attributable to obesity in the United States. JAMA. 1999;282(16):1530-38.

[5] Klein S, Gastaldelli A, Yki-Järvinen H, Scherer PE. Why does obesity cause diabetes? Cell Metab. 2022;34(1):11-20.

[6] Zahr F, Genovese E, Mathier M, Shullo M, Lockard K, Zomak R, et al. Obese patients and mechanical circulatory support: Weight loss, adverse events, and outcomes. Ann Thorac Surg. 2011;92(4):1420-26.

[7] Pepper DJ, Demirkale CY, Sun J, Rhee C, Fram D, Eichacker P, et al. Does obesity protect against death in sepsis? A retrospective cohort study of 55,038 adult patients. Crit Care Med. 2019;47(5):643.

[8] Wang S, Liu X, Chen Q, Liu C, Huang C, Fang X. The role of increased body mass index in outcomes of sepsis: A systematic review and meta-analysis. BMC Anesthesiol. 2017;17:118.

[9] Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost. J Transl Med. 2020;18(1):01-04.

[10] Sistrom CL, Garvan CW. Proportions, odds, and risk. Radiology. 2004;230(1):12-19.

[11] Huang FL. Alternatives to logistic regression models in experimental studies. J Exp Educ. 2022;90(1):213-28.

[12] Knol MJ, Le Cessie S, Algra A, Vandenbroucke JP, Groenwold RH. Overestimation of risk ratios by odds ratios in trials and cohort studies: Alternatives to logistic regression. Can Med Assoc J. 2012;184(8):895-99.

[13] Qu Y, Niu H, Li L, Li M, Yan S, Li M, et al. Analysis of dose-response relationship between BMI and hypertension in Northeastern China using restricted cubic spline functions. Sci Rep. 2019;9(1):01-07.

[14] Feng X, Wu W, Zhao F, Xu F, Han D, Guo X, et al. Relationship between body mass index and kidney stones based on dose-response analyses using restricted cubic splines applied to NHANES 2011-2016 data. J Ren Nutr. 2021;31(3):263-69.

[15] Nieboer D, Vergouwe Y, Roobol MJ, Ankerst DP, Kattan MW, Vickers AJ, et al; Prostate Biopsy Collaborative Group. Nonlinear modeling was applied thoughtfully for risk prediction: The Prostate Biopsy Collaborative Group. J Clin Epidemiol. 2015;68(4):426-34.

[16] Lusa L, Ahlin Č Restricted cubic splines for modelling periodic data. PLoS One. 2020;15(10):e0241364.

[17] Harrell FE. Regression modeling strategies. Bios. 2017;330(2018):14.

[18] Dan-Ni WN, Li-Dian HU, Zhi-Guo PA, Qiang WE, Lei SU. The impact of body mass index on the prognosis in sepsis patients: A retrospective analysis on account of the large clinical database MIMIC-III. Jie Fang Jun Yi Xue Za Zhi. 2021;46(2):129-35.

[19] Li S, Hu X, Xu J, Huang F, Guo Z, Tong L, et al. Increased body mass index linked to greater short-and long-term survival in sepsis patients: a retrospective analysis of a large clinical database. Int J Infect Dis. 2019;87:109-16.

[20] Lin S, Ge S, He W, Zeng M. Association between body mass index and short-term clinical outcomes in critically ill patients with sepsis: A real-world study. Biomed Res Int. 2020.

[21] Shepherd BE, Rebeiro PF. Assessing and interpreting the association between continuous covariates and outcomes in observational studies of HIV using splines. J Acquir Immune Defic Syndr (1999). 2017;74(3):e60.

[22] Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med. 2000;19(14):1831-47.

PARTICULARS OF CONTRIBUTORS:
1. Research Scholar, Department of Data Science, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, Karnataka, India.
2. Senior Research Associate – Biostatistics, Centre for Health Informatics, Computing and Statistics, Lancaster Medical School, Lancaster University, United Kingdom.
3. Assistant Professor, Division of Biostatistics, MOSC Medical College, Kolenchery, Kerala, India.
4. Assistant Professor, Division of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
5. Associate Director PSPH, Professor and Head, Department of Data Science, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, Karnataka, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Dr. MK Lintu,
Research Scholar, Department of Data Science, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, Karnataka, India.
E-mail: lintu.mk@learner.manipal.edu

AUTHOR DECLARATION:
• Financial or Other Competing Interests:   None
• Was Ethics Committee Approval obtained for this study?     No
• Was informed consent obtained from the subjects involved in the study?   No
• For any images presented appropriate consent has been obtained from the subjects.   No