Chettinad Health City Medical Journal
Volume 15, Issue 2 - 2026, Pg. No. 4-9
Peer Reviewed & Open Access Journal

Research Article

Regression Modelling and Forecasting for Indian Suicide Cases

Dhanalakshmi Thangarasu[1], Thirunavukkarasu Murugesan[2], Arunkumar Radhakrishnan[3], Vedapriya Dande Rajasekar[4]

[1]Faculty of Allied Health Sciences, Chettinad Hospital and Research Institute, Chettinad Academy of Research and Education, Kelambkkam, Tamil Nadu, Tamil Nadu, India.
[2]Assistant Professor cum Statistician, Department of Community Medicine, Chettinad Hospital and Research Institute, Chettinad Academy of Research and Education, Kelambkkam, Tamil Nadu, India.
[3]Dean & Professor and Head, Department of Pharmacology, Chettinad Hospital and Research Institute, Chettinad Academy of Research and Education, Kelambkkam, Tamil Nadu, India.
[4]Department of Community Medicine, Chettinad Hospital and Research Institute, Chettinad Academy of Research and Education, Kelambkkam, Tamil Nadu, India.
DOI: https://doi.org/10.24321/2278.2044.202609

I N F O

Corresponding Author:
Thirunavukkarasu Murugesan, Department of Community Medicine
E-mail Id:
thirunavukkarasum@care.edu.in
Orcid Id:
https://orcid.org/0000-0003-1773-7239
How to cite this article:
Thangarasu D, Murugesan T, Radhakrishnan A, Rajasekar V D. Regression Modelling and Forecasting for Indian Suicide Cases. Chettinad Health City Med J. 2026;15(2):4-9.

Date of Submission: 2025-08-05
Date of Acceptance: 2026-06-011

A B S T R A C T

Objective: The issue of suicide has become a topic of public health concern in India. Mitigating the repercussions associated with suicide has become an imperative focus of public health initiatives. Proactive identification of suicide risks, informed by surveillance data, is crucial for managing the incidence of suicide fatalities in India. Our research endeavors were aimed at exploring the feasibility of employing a Quadratic model to forecast the incidence of suicide deaths in India.

Method: The dataset comprising annual suicide death cases in India from 1990 to 2016 was utilized to calibrate the Quadratic model. The discrepancy between the observed and predicted values was employed to assess the forecasting accuracy of the developed replicas.

Results: Mean suicide rate remained determined to be 20.11789, with a standard deviation of 1.194, a minimum of 17.8963, and a maximum of 21.4266 individuals. The Quadratic model demonstrated statistical significance ($p < 0.001$) and passed the correlation residual test ($p > 0.05$), exhibiting a high Adjusted R^2 and R^2 alongside low values of Mean Squared Error, Mean Absolute Error and Root Mean Squared Error and Average Forecast Error values at 0.961, 0.975, 0.004624, 0.322470, 0.39681, and 0.015475, respectively.

Conclusions: The Quadratic model emerged as the most suitable model for predicting suicide death cases in India.

Keywords: Suicide, Epidemiology, Regression, Quadratic, Curve, Prevention

Introduction

Self-harm represents one of the top three causes of death in young people around the world. According to the World Health Organization (WHO), nearly one million people take their own lives each year, while the number of those who attempt suicide is roughly twenty times higher; this results in a global death rate of 16 per 100,000 individuals,

which means one death occurs every 40 minutes and one attempt happens every 3 minutes on average. In 1998, it was estimated that self-harm would represent 1.8 percent of the global burden of disease; by 2020, this figure is expected to rise to approximately 2.4 percent in countries identified as having market economies and previously communist governments.

According to the most recent data from the World Health Organization as of 2011, suicide rates differ greatly, with figures as low as 0.7 per 100,000 in the Maldives and as high as 63.3 per 100,000 in Belarus. In 2009, India ranked 43rd globally in terms of suicide rates, with an incidence of 10.6 per 100,000 (WHO suicide rates).[1] India has experienced noticeable growth in suicide rates over the years, although there have been clear fluctuations of both increases and decreases during this time. Unlike the predominant high-risk demographic factors observed globally, marital status in India does not necessarily serve as a protective factor, and the female to male ratio concerning suicide rates is disproportionately higher. The motivations and methods of suicide also exhibit notable differences when compared to Western nations. Preventive measures enacted at the community level, along with the verification of identity of factor of risk individuals, may prove to be more efficacious than broad-based global strategies.[2] Over the past sixty years, according to the WHO, Japan, Hungary, and Lithuania have consistently ranked at the top in terms of suicide rates; however, should current trends persist, South Korea is anticipated to surpass all others within a few years. The epicenter of suicide-related deaths has transitioned from Western Europe to Eastern Europe, and it now appears to be migrating towards Asia. China and India represent the largest contributors to the absolute number of suicides on a global scale.[3] An Indian study revealed that the highest suicide rate was observed among the 15-29 age group (38 per 100,000 population), followed by the 30-44 age group (34 per 100,000 population). The suicide rate was documented at 18 per hundred thousand for individuals aged 45 to 59, and 7 per hundred thousand for those who are 60 years and older.[4]

The objective is to analyze the historical trends in suicide rates in India over a 50-year period (1969–2018) by utilizing longitudinal data from the NCRB, while also identifying the demographic and methodological traits of suicide victims. Additionally, the aim is to project national suicide rates for the next ten years through ARIMA time series modeling to better understand future suicide risks and aid in preventive planning.[16] To investigate the 26-year trend from 1995 to 2021 in suicide rates among children and teenagers in India, categorized by age and gender, in order to recognize patterns, methods, and reasons for suicide, as well as to project suicide rates for the upcoming decade utilizing ARIMA time series analysis, thus guiding age-specific prevention strategies.[17] The goal is to analyze and predict suicide incidents in India based on their specific causes using multivariate time series techniques VAR and VARMA, while also evaluating their predictive performance against the standard univariate ARIMA model, to determine targeted intervention strategies for suicide prevention.[18] This study aims to examine the relationship between various sociodemographic and economic factors at the state level such as religion, literacy rates, GDP, fertility rates, and urban unemployment and suicide rates in different Indian states. By employing ecological regression analysis, the goal is to pinpoint population level factors that can guide the development of targeted policies for suicide prevention.[19] The objective is to examine the changes over time in suicide incidents in India from 1967 to 2021 using ARIMA time series modeling, and to predict the number of suicides for the years from 2022 to 2026, with the aim of informing data-driven policy choices and prompt public health funding for suicide prevention.[20]

Materials And Methods

This research constitutes a retrospective inquiry aimed at obtaining the incidence of suicide fatalities in India, quantified as the number of suicide occurrences per 100,000 individuals for the period spanning from 1990 to 2016, sourced from the website https://ourworldindata.org/suicide, which is compiled by the Global Burden of Disease Study (GBD 2016).[5] The gathered data underwent analysis utilizing the (SPSS version 23) Statistical Package for Social Sciences, and the most appropriate curve estimation of the regression model for Indian suicide mortality cases was employed. The performance evaluation criteria were implemented to ascertain the accuracy of the model specification. Additionally, projections of the suicide rate were conducted leveraging the validity of the quadratic model.

Model Fitting

The Indian suicide cases showed a downward trend overall, with a few exceptions from 1991 to 1997 during the twenty-seven years of the study. For the suicide data, the Kolmogorov-Smirnov and Shapiro-Wilk tests yielded $p < 0.0001$, indicating that the data did not adhere to a normal distribution. The Spearman's Rho value of r is -.962, indicating a strong negative correlation, which suggests that as the years (X) increase, the number of suicide cases (Y) decreases. The R2 value stands at 93%, and the p-value is less than 0.0001. The y-axis represents suicide cases, while the x-axis indicates the relevant year; a curve fitting approach was utilized to determine the most appropriate model. Among all the models evaluated, the Quadratic model exhibited the strongest explanatory capability, achieving an R^2 value of 0.96 and the highest Adjusted R^2 value of 0.975. This suggests an exceptional

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202609

fit to the data when compared to the other models. The other models, such as Linear, Logarithmic, Inverse, Cubic, Compound, Power, Growth, Exponential, and Logistic models, displayed relatively lower R^2 and Adjusted R^2 values, which ranged from 0.877 to 0.886. The quadratic model is the most suitable and well-fitting option for the dataset examined, as it accounts for about 96% of the variability in the outcome variable and delivers the highest predictive performance among all models assessed. This indicates that the connection between the variables is more accurately depicted by a curvilinear quadratic pattern instead of a straightforward linear relationship. The F-test, along with R2 and Adjusted R2 values (greater than 85%), served as criteria for selecting the best-fitting curves and validating the model, with highly significant p-values when below 0.0001, as shown in [Table 1].

All regression models proved to be appropriate, with the cubic and quadratic models exhibiting the highest validity due to their minimal residual values. Among these, the quadratic model emerged as the most effective, accurately reflecting the trends in observed and predicted Indian suicide cases, as shown in [Figure 1 & 2]. The quadratic model's structural equation, which describes the influence of suicide data, the equation is represented as with the coefficients corresponding to and while the constant term is denoted by , In this formula, Y denotes the annual number of suicide cases, while X represents the corresponding years; for instance, 1990=1, 1991=2, 1992=3, 1993=4, and so forth.

Result

Suicide data from India were analyzed through the method of curve fitting estimation. The summary of the models and the parameters fitted for various models are shown in [Table 1] and [Figure 1]. The estimation curves of all models were deemed valid and the equations achieved the highest level of accuracy when the constant term was included, resulting in a p-value of (p = 0.001) for each model. This residual error between it's noted and fitted values was utilized to assess the predictive accuracy of the developed models.

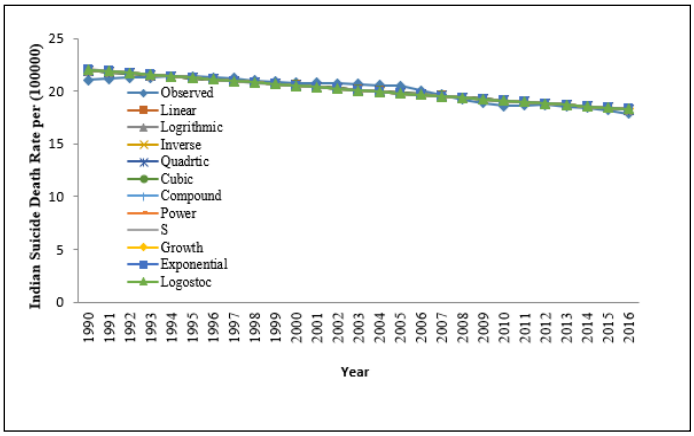

Figure 1.Fitted curves for India Suicide death rate

Table 1.Model and parameter estimate for different models for Suicide cases

Models	R^2	Adjusted R^2	p value	x_1	x_2	x_3
Linear	.889	0.884	0.000	-0.142		
Logarithmic	.888	0.883	0.000	-284.034		
Inverse	.887	0.882	0.000	568626.197		
Quadratic	.961	0.975	0.000	.000	-3.543E-005	
Cubic	.890	0.886	0.000	.000	.000	-1.180E-008
Compound	.883	0.878	0.000	0.993		
Power	.882	0.878	0.000	-14.325		
S	.881	0.877	0.000	28677.648		
Growth	.883	0.878	0.000	-0.007		
Exponential	.883	0.878	0.000	-0.007		
Logistic	.883	0.878	0.000	1.007		

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202609

Table 2.Model selection based on criteria

Models	MSE	MAE	RMSE	AFER
Linear	0.005668	0.334051	0.391202	0.016554
Logarithmic	0.005712	0.335634	0.392728	0.016634
Inverse	0.005757	0.337219	0.394258	0.016714
Quadratic	0.004624	0.322470	0.359681	0.015475
Cubic	0.005580	0.330891	0.388163	0.016395
Compound	0.006352	0.351117	0.414136	0.017366
Power	0.006400	0.352722	0.415722	0.017447
S	0.006449	0.354329	0.417312	0.017527
Growth	0.006352	0.351117	0.414136	0.017366
Exponential	0.006352	0.351117	0.414136	0.017366
Logistic	0.006352	0.351117	0.414136	0.017366

The cubic and quadratic models exhibited the lowest error values, with both models showing significant p-values as indicated in [Table 2]. I suggest that the quadratic model, which has the least parameters and residual error value, is the optimal choice for predicting suicide cases. The equations for the quadratic model utilize the independent variable X, representing the respective years, and the dependent variable Y, which denotes the incidence of suicide cases, along with the coefficients , and , derived from the observed suicide data. The equation for the chosen quadratic model concerning suicide cases is $Y = 162.277 - 3.543E\text{-}005\ X^2$, as outlined in Table 1, where Y signifies the annual number of suicide cases and X corresponds to the years; for instance, 1990=1, 1991=2, 1992=3, 1993=4, etc.

Lower Limit Confidence (LLC), Upper Limit Confidence (ULC),

In [Table 3] and [Figure 2], the quadratic equation forecasts suicide cases, along with their lower and upper confidence limits, projected through to the year 2026. The estimated suicide cases in India, built on data from the Global Burden of Disease Collaborative Training of the Network, for the years 2017 to 2026 are approximately 18.13068, 17.98873, 17.84679, 17.70484, 17.5629, 17.42095, 17.27901, 17.13707, 16.99512, and 16.85318. [Figure 2] illustrates the downward trend in Cases of suicide in India.

Table 3.Observed and forecasted Suicide death cases in India

Year	LCL	Quadratic	UCL
2017	17.26448	18.13068	18.99687
2018	17.12253	17.98873	18.85494
2019	16.98058	17.84679	18.71300
2020	16.83863	17.70484	18.57106
2021	16.69668	17.56290	18.42912
2022	16.55473	17.42095	18.28718
2023	16.41278	17.27901	18.14524
2024	16.27083	17.13707	18.00331
2025	16.12888	16.99512	17.86137
2026	15.98693	16.85318	17.71943

Table 4.Validation of model

Model	Residual		
	Mean	Min	Max
Quadratic	-0.087461	-.86743	.68494
Cubic	-0.090933	-.86448	.68306

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202609

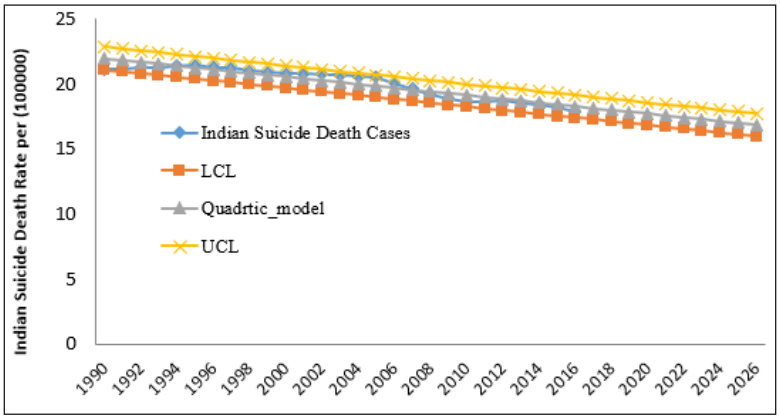

Figure 2.Predicting incidents of suicide on an annual basis

Validation of model

In [Table 4], the validation of cubic and quadratic models is presented; these two examples are solely related to the initial pattern of decreasing suicide rates observed. Both structural regression equations feature positive values that account for forecasts above the actual data, while negative values account for those below the forecasts, indicating that the extreme curves' forecasting characteristics of Indian suicide data will dissipate.

Quadratic and cubic models exhibit the lowest residual values in comparison to other regression models presented in [Table 2]. As shown in [Figure 1], all model curves have underestimated the outcomes. In the comparison between the quadratic and cubic models, the quadratic regression model has demonstrated a significance level of ($p < 0.0001$) and the residual test shows ($p > 0.05$), along with the highest Adjusted R^2 and R^2 and the lowest residual values for MSE, MAE, RMSE, and AFER. The forecasts generated by the quadratic model provide equally reliable estimates with fewer parameters when compared to the cubic model in [Table 2, Figure 2]. Consequently, the quadratic model has been identified as the most appropriate and validated model. The quadratic model yields ($p < 0.001$) and the residual tests show ($p > 0.05$), demonstrating high Adjusted R^2 and R^2 values along with low MSE, MAE, RMSE, and AFER at 0.961, 0.975, 0.004624, 0.322470, 0.359681, 0.015475, respectively.

Discussion

Suicide represents a significant public health issue that is largely preventable. However, addressing this issue is challenging, as highlighted by Gajalakshmi et al, due to the complex interplay of factors including depression, chronic physical ailments, poverty, unhappy marriages, low literacy rates, unemployment, domestic violence, crop failures, rising farming costs, the decline of traditional family structures, unfulfilled romantic expectations, generational conflicts, loss of employment or loved ones, substantial debt, mistreatment by in-laws and spouses, dowry challenges, alcoholism or drug abuse, and unlimited availability of suicide methods.[6] Strategies for primary prevention should focus on fostering Encouraging mental health and developing effective coping strategies in children; increasing awareness among parents, educators, then medical care services workers about effective approaches to parenting strategies and the execution of early strategies related to insufficient coping strategies. At the community level, this implementation of support initiatives for children and families, as well as programs targeting gender and socio-economic equality, could be beneficial. It is crucial to maintain a balance between press freedom and the media's responsibility to protect vulnerable individuals from harm.[7] The importance of advocacy and legislative action cannot be overstated. The WHO has recommended laws to limit access to lethal means such as firearms. Non-governmental organizations can significantly contribute to advocacy efforts, as shown by one NGO that found the highest rates of suicide were among students who had not excelled in a specific academic subject. Therefore, in 2002, government officials introduced a new policy that allows students who fail a single subject to retake the exam within a month, helping them continue their studies without the setback of losing an entire academic year.[8] The assessment of various curve fitting regression techniques has been applied to model and predict numerous communicable and non-communicable diseases.[9-16] Regarding HIV cases in Nepal, there has been an upward trend, and the cubic model provided the best fit, achieving an accuracy of R2 = 99% ($p < 0.001$).[9] For cancer cases at hospitals in Nepal, models that included a constant term were not suitable, while excluding the constant term in the cubic model yielded the best fit with an R2 of 95% and $p < 0.001$ for the overall cancer cases undergoing radiotherapy.[10] There was a noticeable increase in cancer cases receiving radiotherapy with a focus on palliative care during the entire duration of the study period. From this mathematical analysis, when constant term was excluded, the Logarithmic model emerged as the

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202609

best fit for forecasting cancer patients receiving palliative treatment, featuring an R2 of 72% and p = 0.002.[11]

The cubic model structural equation, which did not include a constant term, demonstrated the best fit with an R2 value of 0.95 and a value of p less than 0.0001, indicating an increasing trend in breast cancer cases in Nepal.[12] Conversely, a statistical model based on the Inverse model equation, which incorporated a constant period, was the best fit for forecasting the time trend of grade 2 disability among new leprosy cases in Nepal, yielding an R2 of 74% and a p-value of 0.062.[13] The prevalence of HIV/AIDS cases among Indian adults shows a declining trend based on statistical modeling and forecasting. After omitting the constant term from the mathematical equation and considering the curve's shape, the quadratic model emerged as the most suitable option (R2 = 99%, p = 0.0001).[14,15] In our current research, the quadratic model was again determined to be the best fit (R2 = 96%, p<0.001), exhibiting a lower residual value for MSE, MAE, RMSE and AFER. Compared to other regression models, quadratic regression model was identified as the closest fitting curve for suicide cases in India.

Conclusions

An Indian suicide case study shows a downward trend that is non-stationary. After fitting and comparing various regression curve estimation models, the quadratic model emerged as the best fitting final model due to its lowest residual value. In comparison to other regression model fittings, the quadratic model had fewer parameters than the cubic model. Therefore, this data suggests that the quadratic model is more suitable for predicting future suicide cases. This forecasting model can assist the Indian government in implementing proactive measures and prevention strategies for the anticipated increase in suicide cases.

Acknowledgements

The authors extend their sincere gratitude to the World Health Organization (WHO) and the Global Burden of Disease Collaborative Network.

References

1. World Health Organization. Suicide rates per 100,000 by country, year and sex. [Last accessed on 2012Mar27].
2. Radhakrishnan R, Andrade C. Suicide: An Indian perspective. Indian J Psychiatry 2012; 54:304-19.
3. Peeter Varnik. Suicide in the world. Int. J. Environ. Res. Public Health 2012; 9:760-771.
4. World Health Organization. Health statistics and information systems/Estimates for 2000-2012/ causespecificmortality.2014. http://www.who.int/ healthinfo/global_burden _disease /estimates/ en/ index1.html (accessed 20 Sept 2014).
5. Global Burden of Disease collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Results, Published by Seattle, United States: Institute for Health Metrics and Evaluation (IHME),2017. https:// ourworldindata.org
6. Gajalakshmi V, Peto R. Suicide rates in rural tamilnadu, South India: Verbal Autopsy of deaths in 1997-98. Int.J.Epidemiol 2007; 36:203-7.
7. Sharma BR, Gupta M, Sharma AK, Sharma S, Gupta N, Relhan N, et al. Suicides in northern India: Comparison of trends and review of literature. J Forensic Leg Med 2007; 14:318-26.
8. Vijayakumar L. Suicide prevention: The urgent need in developing countries. World Psychiatry 2004; 3:158-9.
9. Sathian B, Sreedharan J, Mittal A, Baboo NS, Chandrasekharan N, Devkota S, et al. Statistical Modelling and Forecasting of Reported HIV Cases in Nepal. Nepal Journal of Epidemiology 2011; 1:106-10.
10. Sathian B, Bhatt CR, Jayadevan S, Ninan J, Ninan J, Baboo NS, Sandeep G. Prediction of cancer cases for a hospital in Nepal: a statistical modelling. Asian Pac J Cancer Prev 2010; 11:441-445.
11. Sathian B, Sreedharan J, Sharan K, Baboo NS, Chawla R, Chandrasekharan N, et al. Statistical modelling Technique in Forecasting of Pallative Oncotherapy Load in Hospitals. Nepal Journal of Epidemiology 2010; 1:38-43.
12. Sathian B, Sreedharan J, Sharan K, Baboo NS, Ninan J, Joy T, Abhilash ES. Forecasting Breast Cancer Cases requiring Radiotherapy at a Teaching Hospital in Nepal. Journal of Clinical and Diagnostic Research 2010; 6(4):2378-2383.
13. Sathian B, Kumar A, Sreesharan J, Banerjee I, Roy B, Rajesh, et al. Time trend and prediction of grade 2 disability among new cases of leprosy in Nepal: A statistical modelling. Nepal Journal of Epidemiology 2014; 4(3): 378-83.
14. Sathian B et al. Forecasting of HIV/AIDS in India: A Statistical Modelling. Gulf Medical Journal 2014; 3(1):6-10.
15. Poduri PGS. Time series analysis of Indian suicides: Correlation with human development index (HDI). Acta Medica Internatinonal 2015; 2(1):122-124.
16. Swain, P. K., Tripathy, M. R., Priyadarshini, S., & Acharya, S. K. (2021). Forecasting suicide rates in India: An empirical exposition. PLOS ONE, 16(7).
17. Jena, S., Swain, P. K., Senapati, R. E., & Acharya, S. K. (2024). Trajectory of suicide among Indian children and adolescents: A pooled analysis of national data from 1995 to 2021. Child and Adolescent Psychiatry and Mental Health, 18, Article 123.
18. Anjali, B., & Kumar, R. K. (2024). Exploring cause-specific strategies for suicide prevention in India: