Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 22s (2025)
https://www.jneonatalsurg.com

Logistic Regression in Healthcare: Predictive Modeling for Enhanced Clinical Decision-Making

Prachi Saoji[1], Lakshmi Madireddy[2], Dr. Ajeet Saoji[*3]

[1]Assistant Professor, Department of Mathematics, Ramdeobaba University, previously Ramdeobaba College of Engineering and Management, Nagpur

[2]Assistant Professor, FEAT, DMIHER, Sawangi

[*3]Professor, Department of Community Medicine, NKP Salve Institute of Medical Sciences and Research Centre & LMH

***Corresponding Author**

Dr. Ajeet Saoji,

Email ID: saojiajeet@gmail.com

Cite this paper as: Prachi Saoji, Lakshmi Madireddy, Dr. Ajeet Saoji, (2025) Logistic Regression in Healthcare: Predictive Modeling for Enhanced Clinical Decision-Making. *Journal of Neonatal Surgery*, 14 (22s), 765-769

ABSTRACT

In clinical practice, logistic regression is becoming a key tool that allows health care practitioners to forecast binary outcomes associated with a wide range of diseases.

To identify risk factors and assist clinical decision making, this review aims to lime-light the significance of various models by looking at important factors including lifestyle, clinical conditions, and socio-demographic data.

Logistic regression has proven advantages like, interpretation and adaptation to different datasets, while it also has drawbacks like, it presumes that predictors and outcomes have linear correlations. Carefully validating models and knowledge of the context in which these models are used are mandatory to overcome these problems.

In order to overcome these constraints, the combination of logistic regression along with machine learning techniques shows promise in enhancing the clinical decision making. This review emphasizes the positive effect of logistic regression on decision making in the rapidly growing healthcare industry.

1. INTRODUCTION

Logistic regression is a commonly used statistical method in medical research for predicting binary outcomes, such as disease presence or absence. Its simplicity, ease of implementation, and ability to calculate odds ratio make it ideal for developing prediction models. Common applications include assessing the likelihood of cardiovascular events, diabetes, cancer recurrence, and Cesarean delivery. The technique's strength lies in its ability to adjust for confounding factors, offering actionable insights for clinical decision-making. Despite limitations in handling nonlinear relationships, logistic regression remains a powerful tool for understanding disease risks and informing medical practice. [1,2,3]

2. METHODOLOGY

Logistic regression is a popular technique used in research for modelling the probability of binarily distributed variables and also its practical and capability to offer relevant information makes it a favoured technique of forecasting health risks. Applying logistic regression involves several steps: in developing the model, choosing the predictors, assuring the outcomes' accuracy, and understanding the conclusions.

1. Formulating the Model: While logistic regression employs the use of a logistic function to work out the probability of an event occurring. For instance, it anticipates the probability that a certain disease is present depending on age or lifestyle or certain clinical measurements. This model gives odds ratios for each factor by use of coefficients for every factor. These ratios explain how much a certain factor (as smoking) is associated with the outcome. Mathematically, logistic regression is formulated as follows:

$P(Y=1|X) = e^{\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta nXn}/1+e^{\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta nXn}$

Here:

- Y represents the binary outcome variable.

- $X_1, X_2, \ldots, X_n$ denote the predictor variables.
- β0 is the intercept, while β1, β2, …, βn represent the regression coefficients for each predictor.

The coefficients provide odds ratios, which are inevitable in medical contexts as they indicate the strength and direction of the association between each predictor and outcome. A positive coefficient increases the likelihood of the outcome.

2. Selecting Predictors and Managing Overlap: Choosing the right predictors is essential. Researchers often use factors like age, blood pressure, or cholesterol levels because they're already known to influence certain conditions. Statistical methods, such as forward or backward selection, help refine the list of predictors. However, when predictors are too similar—like cholesterol and triglyceride levels—it can create problems (called multicollinearity) that reduce the model's reliability. Techniques like principal component analysis (PCA) help manage this overlap while preserving valuable information.

3. Ensuring Accuracy with Calibration and Validation: Once the model is built, it needs to be checked for accuracy. Calibration ensures the predicted probabilities match real-world outcomes. For example, if the model says there's a 70% chance of disease, this should align with observed data. Validation is equally important to ensure the model works well on new data, not just the dataset used to create it. Cross-validation or external testing with a different population helps achieve this.

4. Measuring Performance: To evaluate how well the model works, researchers use metrics like:

- AUC-ROC: Measures how well the model distinguishes between outcomes.
- Sensitivity and Specificity: Assess how accurately it identifies true positives and negatives.
- Predictive Values: Help gauge the likelihood that predictions are correct.

5. Interpreting and Reporting Results: The simplicity of logistic regression makes its findings easy to understand. Odds ratios show how much each factor increases or decreases the risk, giving clinicians actionable insights. Transparent reporting—such as listing predictors, adjustments for confounders, and confidence intervals—ensures the model's results can be trusted and compared across studies.

3. MODELS FOR DIFFERENT CONDITIONS

Logistic regression has been widely used to predict outcomes in various medical conditions, offering valuable insights for clinicians.

1. Cardiovascular Disease Prediction

Cardiovascular diseases (CVD), leading global causes of death, are often preventable through lifestyle changes and medical interventions. Logistic regression models, such as those from the Framingham Heart Study, use factors like age, cholesterol, smoking, and blood pressure to estimate CVD risk. These models provide odds ratios that allow clinicians to identify and modify risk factors, supporting early prevention. Similarly, hypertension risk models incorporate family history, BMI, and activity levels, enabling targeted strategies to improve cardiovascular health.[4,5]

2. Cancer Prognosis and Recurrence

In cancer care, logistic regression helps predict survival and recurrence. For instance, breast cancer models incorporating gene expression profiles allow personalized treatment plans, while lung cancer models use tumor stage and patient demographics to refine survival predictions. These tools guide treatment decisions and improve patient counseling.[6,7]

3. Infectious Disease Outcomes

During the COVID-19 pandemic, logistic regression models predicted mortality risk and ICU admissions based on factors like age and oxygen levels, aiding resource allocation. For sepsis, models incorporating vitals and lab values enable quick risk assessments in critical care, helping clinicians prioritize aggressive treatment.[8,9]

4. Neurological and Respiratory Conditions

Stroke recurrence and mortality prediction models use factors like hypertension and atrial fibrillation to guide preventative strategies. For COPD, logistic regression identifies patients at risk of severe disease progression, supporting timely interventions.[10,11]

5. Cesarean Section Risk Prediction

In obstetrics, logistic regression predicts Cesarean delivery needs using factors like maternal age, fetal position, and previous C-sections. These models assist in planning labor and minimizing complications, improving outcomes for mothers and babies.[12]

Across diverse medical fields, logistic regression remains a powerful tool for actionable, data-driven decision-making, improving prevention, treatment, and patient outcomes.

Condition	Key Predictors	Model Purpose	Reference
Cardiovascular Disease	Age, cholesterol levels, smoking status, blood pressure, diabetes	Predict risk of developing cardiovascular disease	*Wilson et al. (1998)*
	Family history, BMI, physical activity	Identify hypertension risk	*Garrison et al. (1987)*
Cancer Prognosis and Recurrence	Tumor stage, histology, patient demographics	Predict survival and recurrence risks in cancer	*Subramanian & Simon (2010)*
	Gene expression profiles	Tailor treatment approaches	*van't Veer et al. (2002)*
Infectious Disease Outcomes	Age, comorbidities, laboratory findings	Predict mortality risk and hospital needs	*Wynants et al. (2020)*
	Oxygen saturation levels	Assess COVID-19 severity	*Seymour et al. (2016)*
Neurological Conditions	Previous TIA, hypertension, atrial fibrillation	Prevent primary stroke	*Goldstein et al. (2001)*
Respiratory Conditions	Smoking history, age, environmental exposure, respiratory function tests	Assess COPD progression risk	*Duan RR. (2020)*
Cesarean Section Risk	Maternal age, previous C-sections, fetal position, comorbidities	Predict need for Cesarean delivery	*Ladfors L (2023)*

4. DISCUSSION

The use of logistic regression in medical research has revolutionized clinical decision-making, enabling healthcare professionals to identify risks, personalize care, and improve patient outcomes. Here, we explore its clinical implications, strengths, limitations, and future prospects.

1. Clinical Implications

Logistic regression integrates diverse risk factors into cohesive predictions, guiding early interventions. For example, in cardiovascular care, models combining age, cholesterol, and lifestyle factors help detect high-risk patients, reducing morbidity and mortality through preventative strategies. Similarly, diabetes prediction models allow early diagnosis and management, preventing complications.

In oncology, logistic models aid personalized medicine by identifying patients at higher risk of recurrence or poor outcomes. This enables tailored treatment plans and informed discussions with patients, fostering better care and resource allocation.

2. Strengths of Logistic Regression

A key strength is its interpretability. Odds ratios are intuitive, allowing clinicians to assess risk factors clearly and communicate findings effectively. The method accommodates diverse predictors and adjusts for confounders, ensuring reliable insights in observational studies. Its simplicity and accessibility have driven widespread use across fields, including cardiology, oncology, and obstetrics.

3. Limitations

However, logistic regression has limitations. The assumption of linearity between predictors and outcomes may fail to capture complex relationships. It is sensitive to outliers and multicollinearity, which can distort results and complicate interpretation. Additionally, reliance on retrospective data can introduce bias and limit generalizability, emphasizing the need for external validation across populations.

4. Future Directions

The future lies in hybrid approaches, combining logistic regression with machine learning to identify non-linear relationships and improve accuracy. Big data and wearable devices could enable dynamic, real-time predictions, transforming patient care. Incorporating patient-reported outcomes and qualitative factors may further enhance models, offering a holistic view of patient health and risks.

Logistic regression remains a cornerstone of medical prediction, with ongoing advancements promising even greater clinical impact.

In conclusion, Logistic regression has significantly advanced medical research and clinical decision-making by providing interpretable, data-driven insights. While it has limitations, ongoing innovations like hybrid models and real-time data integration promise to enhance its accuracy and applicability, empowering healthcare professionals to deliver personalized, proactive, and effective patient care.

REFERENCES

[1] Schober P, Vetter TR. Logistic Regression in medical research. Anesthesia and Analgesia. 2021 Feb 1;32(2):365-6

[2] Sperandei S. Understanding Logistic Regression Analysis. Biochemia Medica. 2014;24(1):12–8.

[3] Ayinde K, Apata EO, Alaba OO. Estimators of Linear Regression Model and Prediction under Some Assumptions Violation. Open Journal of Statistics. 2012;02(05):534–46

[4] Wilson PWF, D'Agostino RB, Sullivan L, Parise H, Kannel WB. Overweight and Obesity as Determinants of Cardiovascular Risk. Archives of Internal Medicine [Internet]. 2002 Sep 9;162(16):1867. Available from: https://pubmed.ncbi.nlm.nih.gov/12196085/

[5] Garrison, R. J., Kannel, W. B., Stokes, J., 3rd, & Castelli, W. P. (1987). Incidence and precursors of hypertension in young adults: the Framingham Offspring Study. *Preventive medicine*, *16*(2), 235–251. https://doi.org/10.1016/0091-7435(87)90087-9

[6] Van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature [Internet]. 2002 Jan [cited 2020 Jan 12];415(6871):530–6. Available from: https://www.nature.com/articles/415530a

[7] Subramanian, J., & Simon, R. (2010). An evaluation of resampling methods for assessment of survival risk prediction in high-dimensional settings. https://doi.org/10.1002/sim4106

[8] Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. 2020 Apr 7;369:m1328.

[9] Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., Rubenfeld, G., Kahn, J. M., Shankar-Hari, M., Singer, M., Deutschman, C. S., Escobar, G. J., & Angus, D. C. (2016). Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, *315*(8), 762–774.

[10] Goldstein, L. B., Adams, R., Alberts, M. J., & Appel, L. J. (2001). Primary prevention of ischemic stroke: A statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke, 32(1), 186-202. https://doi.org/10.1161/01.str.32.1.280

[11] Duan RR, Hao K, Yang T. Air pollution and chronic obstructive pulmonary disease. Chronic Diseases and Translational Medicine [Internet]. 2020 Jul 11;6(4). Available from: https://www.sciencedirect.com/science/article/pii/S2095882X20300438#bib4

[12] Keag, O. E., Norman, J. E., & Stock, S. J. (2018). Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. PLOS Medicine, 15(1), e1002494. https://doi.org/10.1371/journal.pmed.1002494