Iran J Public Health, Vol. 54, No.10, Oct 2025, pp.2395-2397

Letter to the Editor

A Statistical Model to Estimate the Potential Risk of Work-Related Stroke among Office Workers: A Methodological Description Article

Mahsa Farghadan [1], ***Nayef Shabbab Almutairi*** [2], ****Ardalan Shariat*** [3], ***Albert T. Anastasio*** [4]

1. *Department of Artificial Intelligence, Faculty of Computer Engineering, Islamic Azad University of South Tehran Branch, Tehran, Iran*
2. *Department of Public Health, Al-lith College of Health Sciences, Umm Al-qura University, Mecca, Saudi Arabia*
3. *Department of Digital Health, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran*
4. *Department of Orthopedic Surgery, Duke University, Durham, NC 27710, USA*

***Corresponding Author:** Email: a-shariat@tums.ac.ir

(Received 12 Apr 2025; accepted 25 Apr 2025)

Dear Editor-in-Chief

Work-related stress and sedentary life style, can affect the health of workers in various ways, such as increasing blood pressure, inflammation, and heart rate, as well as impairing mental health and well-being. These factors can increase the risk of developing cardiovascular diseases, such as ischemic heart disease and stroke (1,2).

This preliminary multivariable logistic regression model suggested by this paper (3) is described as a potential methodology to explore and estimate the individual incidence of a work-related stroke event using a statistical model, focusing on independent variables. Given the lack of similar models which have been employed in occupational settings, this proposed methodology could yield excellent insight into occupation-related factors which can be associated with stroke.

Using logistic regression to estimate the risk of stroke compared to other methods can have several advantages:

1. Probabilistic modeling
2. Binary modeling
3. Scalability
4. Nonlinear effects modeling
5. Interpretability (4–7).

To estimate the probability of a potential stroke, statistical and modeling methods are used. One of the common methods to estimate the probability of a potential stroke is use of logistic regression modeling. In this method, the probability of having a stroke could be calculated based on a set of risk factors and different variables.

The general steps for using logistic regression method are as follows (8):

1. Data collection
2. Data preprocessing
3. Building a logistic regression model
4. Training the model
5. Evaluating the model

In this suggested logistic regression model, the intent would be to estimate the probability of having a stroke. In this model, binary outcome-specific variables (high stroke risk or low stroke risk) which are dependent variables, and inde-

DOI: https://doi.org/10.18502/ijph.v54i10.20150

pendent variables (risk factors, such as characteristics of the workers and their jobs, such as occupation type, working hours, income level, education level, smoking status, alcohol consumption, physical activity, body mass index, blood glucose level, cholesterol level, history of hypertension or diabetes, and myriad other variables of interest) could be used to predict the occurrence of work-related stroke.

The logistic regression model uses these data to predict the probability of having a stroke based on independent variables.

The logistic regression model uses logistic function that its output is the probability of having a stroke. This function is defined as follows:

$P(Y=1|X) = 1 / (1 + \exp(-z))$ (A) (9)

In this formula (A):

$P(Y=1|X)$ indicates the probability of having a stroke and X is the set of independent variables. z is the weighted sum of independent variables and is calculated as follows:

$z = b0 + b1x1 + b2x2 + \ldots + bnxn$ (B) (9)

In this formula (B):

b0, b1, b2, …, bn are the weights of the model that are optimized by maximum likelihood estimation method. After training the model, it can be used to estimate the probability of having a stroke based on independent variable values. Researchers may need to use tools and programming languages related to logistic regression model (such as Python and related libraries) for accurate implementation of this methodology. Several additional considerations for the use of binary logistic regression are listed below.

1. The ability for use in binary research questions: Logistic regression is suitable for binary situations, where an event can be classified into one category.
2. Interpretability: Logistic regression often has the capacity to be easily interpreted. By analyzing the regression coefficients, we can examine the impact of each independent variable on the probability of occurrence of the event.
3. Fast computational capacity: Training and inference in logistic regression does not typically require substantial computational power.
4. Resistance to poorly balanced data: Logistic regression is resistant to poorly balanced data.
5. Generalizability: Logistic regression has a high generalizability and can be easily applied to more problems (10).

While a potential binary logistic regression analysis using the methodology described in this manuscript can be accurate and can control for a variety of confounding variables and issues with missing data points, each statistical method has its own advantages and limitations.

Use of a specific method should depend on the problem, the features of the data, and the specific conditions of the data set. Therefore, in choosing the appropriate method for each problem, a researcher should carefully and accurately evaluate the advantages and limitations of each method.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Teixeira LR, Pega F, Dzhambov AM, et al (2021). The effect of occupational exposure to noise on ischaemic heart disease, stroke and hypertension: A systematic review and meta-analysis from the WHO/ILO Joint Estimates of the Work-Related Burden of Disease and Injury. *Environ Int*, 154:106387.
2. Shariat A, Mohd Tamrin SB, Arumugam M, et al (2016). Musculoskeletal disorders and their relationship with physical activities among office workers: a review. *MJPHM*, 16(1):62-74.
3. Das A. Logistic regression. In: *Encyclopedia of quality of life and well-being research* (2024), Springer; p. 3985–6.
4. Lyu Y, Xu Q, Liu J (2024). Exploring the medical decision-making patterns and influencing factors among the general Chinese public: a binary logistic regression analysis. *BMC Public Health*, 24(1):887.
5. Schober P, Vetter TR (2021). Logistic

regression in medical research. *Anesth Analg*, 132(2):365-366.
6. Gomila R (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *J Exp Psychol Gen*, 150(4):700-709.
7. Kost S, Rheinbach O, Schaeben H (2021). Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling. *Geochemistry*, 81(4):125826.
8. El Sanharawi M, Naudet F (2013). Understanding logistic regression. *J Fr Ophtalmol*, 36(8):710–5.
9. Bishop C (2006). *Pattern recognition and machine learning*. Springer; 205–215 p. 183.
10. Harris JK (2021). Primer on binary logistic regression. *Fam Med Community Health*, 9(Suppl 1):e001290.