Neuro-Cell Molecular Research

Research Article

Neuro-Cell Mol Res 2025;2(3):80-83

Mediator Variable Analysis with the Hayes Process Method: An Application in the Field of Health

Mehmet Sıddık Çetinel[1], Sadi Elasan[1*]

[1]Van Yuzuncu Yil University, Faculty of Medicine, Department of Biostatistics, Van, Türkiye

*Corresponding Author:
Sadi Elasan
Van Yuzuncu Yil University, Faculty of Medicine, Department of Biostatistics, Van, Türkiye
Email: sadielasan@yyu.edu.tr
ORCID ID: 0000-0002-3149-6462

DOI: 10.5281/zenodo.18084340

Received: 24 September 2025
Accepted: 28 November 2025
Published: 30 December 2025

ABSTRACT

This study examines the role of mediators in the relationships between independent and dependent variables in health sciences, analyzing these relationships using the Hayes Process method. It also aims to demonstrate the methodological advantages of the Hayes Process over traditional methods like Baron and Kenny.

A quantitative research design was employed using an open-source dataset (Kaggle Inc: Peyman, 2020). Age was the independent variable (X), LDH was the dependent variable (Y), and CREA, KAL, ALT, NAT, PCR, GLU, AST were the mediators (M). Mediation analysis was conducted using Hayes Process Macro Model 4, with bootstrap confidence intervals (5000 samples, 95% CI) calculated for indirect effects.

The total effect of Age on LDH was significant (b=1.9449, p<0.001). When mediators were included, the direct effect of Age on LDH became non-significant (b=0.1869, p=0.5939), while the total indirect effect was significant (b=1.7580, BootLLCI=1.2689, BootULCI=2.4781). Among the mediators, PCR (b=1.1028) and AST (b=0.7110) had the strongest and most significant indirect effects.

The effect of Age on LDH is indirect, occurring through specific biochemical markers, particularly PCR and AST. The Hayes Process method with bootstrapping provided a reliable analysis without normal distribution assumptions. This study underscores the importance of mediator analysis in health sciences and demonstrates the applicability of the Hayes Process method.

Keywords: Mediator Analysis, Hayes Process, Bootstrap, Biochemical Markers.

Cite this article as: Çetinel MS, Elasan S. Mediator Variable Analysis with the Hayes Process Method: An Application in the Field of Health. Neuro-Cell Mol Res. 2025;2(3):80-83. doi:10.5281/zenodo.18084340

INTRODUCTION

Understanding the mechanisms that explain "how" or "why" relationships between independent and dependent variables occur is of paramount importance. Mediator variable analysis is a powerful statistical method that examines how the effect of an independent variable (X) on a dependent variable (Y) is transmitted, either partially or completely, through a third variable (M) [1].

Traditionally, Baron and Kenny's (1986) four-step model has been widely used for mediator variable analysis [1]. However, this model has significant limitations, including low statistical power, the assumption of normal distribution, and the inability to test indirect effects directly. To overcome these constraints, the Hayes Process method, developed by Andrew F. Hayes, along with bootstrapping techniques, allows for more reliable and flexible testing of indirect effects [2, 3]. Bootstrapping offers a significant advantage as it does not require the assumption of normal distribution and possesses high statistical power, even with small sample sizes.

The field of health is a highly fertile area for the application of mediator variable analysis. For instance, the effect of a treatment method (X) on patient outcomes (Y) might occur through variables such as patient compliance (M) or quality of life (M). In biochemical processes, the effect of a factor like age (X) on the level of an enzyme, such as LDH (Y), can be explained by mediators like creatinine (CREA) or inflammatory markers (PCR).

The primary aim of this study is to examine the mediating roles of various biochemical markers (CREA, KAL, ALT, NAT, PCR, GLU, AST) in the relationship between an independent variable (Age) and a dependent variable (LDH) in the health domain, and to analyze these relationships using the Hayes Process method. The study aims to demonstrate the methodological advantages of the Hayes Process method over traditional approaches while also contributing to the understanding of biochemical changes in the aging process from an applied perspective.

METHODS

Research Model and Data Source

This research was structured as a quantitative investigation, utilizing a relational screening model to elucidate the associations between variables. The analyses were conducted using a robust, publicly available dataset sourced from Kaggle Inc., as cited in Peyman (2020) [4]. This comprehensive dataset encompasses a wide array of biochemical parameters, providing a solid foundation for the examination of complex physiological relationships. The final analytical sample consisted of 1,170 individual records, a sample size that provides substantial statistical power for the mediation analyses performed in this study, thereby enhancing the reliability and generalizability of the obtained results.

Variables

Independent Variable (X): Age

Dependent Variable (Y): Lactate Dehydrogenase (LDH)

Mediator Variables (M): Creatinine (CREA), Potassium (KAL), Alanine Aminotransferase (ALT), Sodium (NAT), C-Reactive Protein (PCR), Glucose (GLU), Aspartate Aminotransferase (AST).

Statistical Analysis

Analyses were performed using IBM SPSS software (version 28) and the Hayes Process Macro extension. First, descriptive statistics (mean, standard deviation) for the variables are presented. Mediation analysis was conducted using Hayes Process Model 4. Model 4 is suitable for testing models involving a single independent variable, one dependent variable, and one or more mediator variables.

The bias-corrected bootstrap method was used to test the significance of indirect effects, and 95% confidence intervals (CI) were calculated. The bootstrap sample size was set at 5000. For an indirect effect to be considered statistically significant, its bootstrap confidence interval must not include zero [2, 3].

RESULTS

Descriptive Statistics

Descriptive statistics for all key variables analysed in this study, encompassing a total sample size of 1170 participants, are comprehensively detailed in Table 1. The analysis reveals that the mean age of the cohort was 52.4 years, with a standard deviation of ±14.1 years. This relatively wide dispersion indicates a substantial age range within the participant group, facilitating the examination of age-related associations. Furthermore, the mean baseline Lactate Dehydrogenase (LDH) level was determined to be 250.3 U/L, with a standard deviation of ±45.7 U/L. The distribution of LDH and other biochemical markers was assessed for normality to ensure the appropriateness of subsequent parametric statistical tests. These descriptive findings provide a crucial foundational overview of the study population's characteristics and the central tendencies of the primary variables of interest.

Hayes Process Model 4 Analysis Results

The results of the analysis conducted to examine the role of seven mediator variables (M1-M7) in the relationship between Age (X) and LDH (Y) are summarized.

First, the effect of Age on the mediator variables was examined. Age had a statistically significant effect on CREA, GLU, NAT, PCR, AST, and KAL ($p < 0.05$), but its effect on ALT was not significant ($p = 0.097$).

In the second stage, the effect of Age and all mediator variables on LDH was assessed. PCR, ALT, and AST had a significant effect on LDH ($p < 0.001$), CREA had a significant but negative effect ($p = 0.023$), while the effects of Age, GLU, NAT, and KAL were not significant.

Table 1. Descriptive Statistics of the Variables (n=1170)

Variable	Mean	Std. Deviation
Age (X)	52.4	14.1
LDH (Y)	250.3	45.7
CREA (M1)	0.9	0.2
GLU (M2)	110.5	23.5
NAT (M3)	140.2	3.5
PCR (M4)	5.2	2.1
ALT (M5)	28.4	10.3
AST (M6)	25.7	8.9
CAL (M7)	4.2	0.5

Total, Direct, and Indirect Effects

The total, direct, and indirect effects are summarized in **Table 2**.

According to Table 2, the total effect of Age on LDH is significant ($b = 1.9449$, $p < 0.001$). However, when the mediator variables are added to the model, the direct effect of Age on LDH loses its significance ($b = 0.1869$, $p = 0.594$), while the total indirect effect is statistically significant ($b = 1.7580$, Bootstrap CI does not include zero). This finding indicates that the mediator variables exhibit a full mediation role. In other words, the effect of Age on LDH occurs entirely through these biochemical markers.

When the indirect effects of individual mediator variables were examined, PCR ($b = 1.1028$, Bootstrap CI: 0.6809 - 1.6800) and AST ($b = 0.7110$, Bootstrap CI: 0.3710 - 1.2488) emerged as the variables with the strongest and most significant indirect effects.

DISCUSSION

This investigation employed the Hayes Process method to rigorously examine the potential mediating roles of seven distinct biochemical markers in the relationship between chronological age and LDH levels. The findings provide compelling evidence that the observed effect of age on LDH is not a direct pathway. Instead, the relationship is predominantly mediated by a subset of specific variables, with Procalcitonin (PCR) and AST emerging as the most statistically significant and influential mediators. This underscores a complex physiological model where aging exerts its influence on LDH indirectly, primarily through pathways linked to inflammatory processes and alterations in hepatic function.

The finding that the total effect of Age on LDH is significant, but the direct effect becomes non-significant when mediators are included, while the total indirect effect remains significant, points to a statistical full mediation effect [2, 5]. This suggests that the aging process does not directly increase LDH levels but exerts an indirect effect through changes it induces in biological processes such as inflammation (PCR), liver function (AST, ALT), and kidney function (CREA). The fact that PCR was the strongest mediator supports the notion that age-related chronic low-grade inflammation (inflammaging) may be a key mechanism in this relationship [6].

Methodologically, this study also highlights the advantages of using the Hayes Process and bootstrapping methods over the traditional Baron and Kenny approach. The bootstrapping technique, by accounting for the potential non-normal distribution of indirect effects, provided more reliable confidence intervals, thus enabling more robust results [3, 7]. Furthermore, the application of a multiple mediation model was a critical methodological strength, as it allowed for the examination of all seven biochemical markers within a single, cohesive analytical framework. This approach enabled the

Table 2. *Total, Direct, and Indirect Effects*

Effect Type	Coefficient (b)	Bootstrap 95% CI	Result
Total Effect (X → Y)	1.9449	(1.2055 - 2.6843)	Significant
Direct Effect (X → Y, controlling for M)	0.1869	(-0.5007 - 0.8745)	Not Significant
Total Indirect Effect (X → M → Y)	1.7580	(1.2689 - 2.4781)	Significant

quantification of the unique contribution of each potential mediator while statistically controlling for the presence of the others. Consequently, it was possible to determine the relative importance of each variable, identifying that PCR and AST were the most substantively significant pathways, rather than merely confirming their individual effects in isolation. This provides a more nuanced and accurate understanding of the complex biological interplay through which age influences LDH levels.

CONCLUSION

The findings of this study substantiate that the association between age and LDH levels is not one of direct causality. Instead, it appears to be mediated through complex biological pathways. Consequently, a comprehensive understanding of the biological effects of aging necessitates the consideration of critical intermediary mechanisms, particularly systemic inflammation and declining organ function.

In clinical practice, these insights underscore the importance of a nuanced approach. When evaluating elevated LDH levels in older adult patients, it is strongly recommended to thoroughly investigate potential underlying inflammatory processes, as measured by markers like PCR, and to assess overall liver function through established enzymes such as AST and ALT. This more holistic diagnostic strategy will aid in distinguishing age-related physiological changes from specific pathologies, thereby facilitating more effective and targeted planning of diagnosis and therapeutic interventions.

Methodologically, this study demonstrates the utility and encourages the broader adoption of the Hayes Process method for researchers investigating complex causal relationships in the health sciences. This analytical approach provides a more powerful and statistically reliable framework for unpacking indirect effects compared to traditional regression analyses.

Notwithstanding these contributions, it is important to acknowledge the limitations inherent in this study's cross-sectional design, which inherently restricts definitive causal inferences. Future longitudinal studies are imperative to examine the dynamics of these proposed relationships over time. Furthermore, validating these findings across more diverse populations and with larger sample sizes would be essential to confirm their generalizability and strengthen the evidence base.

Acknowledgement

This study was derived from the master's thesis of Mehmet Sıddık Çetinel.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper and that they have approved the final version.

Disclosure

The authors declare no conflicts of interest in the preparation and publication of this article.

Ethics Committee Approval

Ethics committee approval is not required in this study. The study was conducted following the international declaration, guidelines, etc.

Funding

The authors have no funding to disclose.

Referee Evaluation Process

Externally peer-reviewed

REFERENCES

1. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51(6):1173-82. https://doi.org/10.1037//0022-3514.51.6.1173
2. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. 2nd ed. New York: Guilford Press; 2017.
3. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods. 2008;40(3):879-91. https://doi.org/10.3758/BRM.40.3.879
4. Kaggle Inc: Peyman, MH. Covid19 blood Sample and biochemistry tests. 2020. Available from: https://www.kaggle.com/datasets/paymanmh/covid19-blood-sample-and-biochemistry-tests
5. Zhao X, Lynch Jr JG, Chen Q. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. J Consum Res. 2010;37(2):197-206. https://doi.org/10.1086/651257
6. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. Nat Rev Endocrinol. 2018;14(10):576-90. https://doi.org/10.1038/s41574-018-0059-4
7. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods. 2002;7(4):422-45. https://doi.org/10.1037//1082-989X.7.4.422