E-ISSN: 2636-7688

May 2026
volume 33
issue 5

ANNALS OF MEDICAL RESEARCH

The Official Journal of the Inonu University Faculty of Medicine

Original Article

Image

Erratum

İNÖNÜ UNIVERSITY PRESS

ANN MED RES

Ann Med Res

Annals of Medical Research

The Official Journal of Inonu University Faculty of Medicine

Editorial Board | Volume: 33 | Issue: 5 | May 2026

Ann Med Res

Annals of Medical Research

The Official Journal of Inonu University Faculty of Medicine

Editorial Advisory | Volume: 33 | Issue: 5 | May 2026

Ann Med Res

Annals of Medical Research

The Official Journal of Inonu University Faculty of Medicine

Table of Contents | Volume: 33 | Issue: 5 | May 2026

Original Articles

Image

Erratum

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Vortioxetine improves gastrointestinal motility and reduces intestinal inflammation in a cuprizone-induced multiple sclerosis model in mice

Ayse Ozkan [a,*], Pelin Koca [b]

[a] İzmir Bakırçay University, Faculty of Medicine, Department of Physiology, İzmir, Türkiye
[b] İzmir Bakırçay University, Faculty of Medicine, Department of Pharmacology, İzmir, Türkiye
*Corresponding author: ayse.ozkan@bakircay.edu.tr (Ayse Ozkan)

MAIN POINTS

- Vortioxetine treatment significantly improved gastrointestinal motility in cuprizone-induced demyelinated mice, as evidenced by increased fecal pellet output and wet fecal weight, restoring values close to controls.
- Intestinal and colonic pro-inflammatory cytokine levels (TNF-α, IL-1β) elevated by cuprizone exposure were markedly reduced with vortioxetine administration, demonstrating its peripheral anti-inflammatory effects.
- These findings suggest that vortioxetine may have therapeutic potential for gastrointestinal dysfunction associated with demyelinating and neuroinflammatory conditions.

Cite this article as: Ozkan A, Koca P. Vortioxetine improves gastrointestinal motility and reduces intestinal inflammation in a cuprizone-induced multiple sclerosis model in mice. *Ann Med Res*. 2026;33(5):192--198. doi: 10.5455/annalsmedres.2025.08.231.

ABSTRACT

Aim: To investigate the effects of vortioxetine on gastrointestinal motility and intestinal inflammation in a cuprizone-induced demyelination model in mice by evaluating pro-inflammatory cytokine responses.

Materials and Methods: Twenty-one C57Bl/6 mice (8 weeks old) were randomly allocated into three groups: Control, cuprizone, and cuprizone + vortioxetine (n = 7/group). Demyelination was induced by oral gavage of cuprizone (10 mg/kg) every other day for 5 weeks. Vortioxetine (10 mg/kg/day) was administered intraperitoneally for 5 weeks. Gastrointestinal motility was assessed by 24-hour fecal pellet count and wet fecal weight. At the end of the protocol, TNF-α and IL-1β levels in the stomach, intestine, and colon were measured by ELISA.

Results: Cuprizone administration significantly reduced fecal pellet output and wet fecal weight compared with controls ($p<0.0001$). Vortioxetine treatment restored both parameters toward control levels ($p<0.01$). Intestinal TNF α and IL 1β levels were markedly increased in the cuprizone group ($p<0.0001$) and were significantly decreased following vortioxetine treatment, reaching near-control values. Similar reductions were observed in colonic cytokine levels, whereas gastric cytokines were largely unaffected.

Conclusion: Vortioxetine improves gastrointestinal motility and attenuates intestinal and colonic pro-inflammatory cytokine expression in a cuprizone-induced demyelination model in mice. These findings suggest that vortioxetine has the potential to exert beneficial effects on the gastrointestinal function and highlight its potential as an adjunctive therapy for gastrointestinal comorbidities in multiple sclerosis.

Keywords: Cuprizone, Vortioxetine, Demyelination, Gastrointestinal motility, Inflammation

Received: Sep 09, 2025 **Accepted:** Nov 19, 2025 **Available Online:** May 22, 2026

INTRODUCTION

Multiple sclerosis (MS) is a chronic demyelinating disorder of the central nervous system (CNS), characterized by immune cell infiltration, myelin loss, and axonal damage, resulting in both neurological and systemic complications [1]. Beyond its classical neurological manifestations, there is increasing evidence suggesting that MS causes disturbances in the gut–brain axis, including gastrointestinal (GI) dysmotility and alterations in inflammatory signaling [2]. Shared genetic risk factors, immune dysregulation, and chronic low-grade inflammation have been reported between MS and inflammatory bowel disease (IBD), with comorbidity between these conditions consistently documented [3,4].

The cuprizone model of demyelination is extensively utilized as an experimental model to study MS physiopathology. It induces oligodendrocyte loss, demyelination, microglial activation, and astrocytosis. Notably, cuprizone exposure elicits systemic and peripheral inflammatory responses, characterized by alterations in gut permeability, immune signaling, and gastrointestinal motility [5-7]. Consequently, this model facilitates the investigation of gut-related changes secondary to CNS demyelination, making it highly relevant for exploring gut–brain axis interactions in neuroinflammatory diseases.

In addition to immune-mediated demyelination, mood dis-

turbances and depression are common comorbidities in MS. There is growing recognition that pro-inflammatory cytokines (e.g., TNF-α, IL-1β, IL-6) contribute not only in pathogenesis of MS but also to the development of depression and GI dysfunction through immune-mediated mechanisms [8-10]. Vortioxetine, a multimodal serotonergic antidepressant, has been shown to modulate serotonergic signaling both centrally and peripherally, polarize macrophages towards an anti-inflammatory phenotype, and positively influence gut microbiota composition [11-13]. Given that the majority of serotonin in the body is localized within the gastrointestinal tract, serotonergic antidepressants may directly affect gut motility and immune balance [14,15]. However, the effect of vortioxetine on gastrointestinal function in the context of neuroinflammatory demyelination such that is seen in MS has not yet been explored. Although the central antidepressant and anti-inflammatory actions of vortioxetine have been characterized, its effect on peripheral gastrointestinal inflammation and motility in demyelination models needs to be investigated.

Understanding these interactions could provide novel insights into bidirectional communication between the nervous and gastrointestinal systems and identify potential therapeutic targets for managing comorbid symptoms in neuroinflammatory diseases. In this study, we investigated the effects of vortioxetine on GI motility and pro-inflammatory cytokine expression (TNF-α and IL-1β) in a cuprizone-induced demyelination model. We hypothesized that vortioxetine, through its multimodal serotonergic and anti-inflammatory actions, would improve fecal excretion and attenuate intestinal inflammation in this model, thereby highlighting the contribution of the gut–brain axis to MS-related comorbidities.

MATERIALS AND METHODS

Animals and experimental groups

This was an experimental, randomized, controlled, parallel-group study designed to evaluate the effects of vortioxetine on gastrointestinal and inflammatory parameters in a cuprizone-induced demyelination model. A total of twenty-one C57Bl/6 mice (male and female, 8 weeks old, 19–21 g) were used in this study (Figure 1). Animals were randomly assigned to experimental groups using a simple randomization method based on a computer-generated random number sequence to ensure unbiased allocation (n = 7/group):

- Control: 1% methylcellulose solution (10 mg/kg, oral gavage) every other day for 5 weeks.
- Cuprizone (CUP): Cuprizone (10 mg/kg, oral gavage) dissolved in 1% methylcellulose solution, every other day for 5 weeks.
- Cuprizone (CUP) + Vortioxetine (VOR): Cuprizone (10 mg/kg, oral gavage, every other day for 5 weeks) plus vortioxetine (10 mg/kg/day, intraperitoneally) for 5 weeks, prepared in 5% 2-hydroxypropyl-β-cyclodextrin.

Only healthy C57Bl/6 mice with normal feeding and stool consistency were included in the study. Mice showing signs of gastrointestinal illness, dehydration, injury, or abnormal behavior prior to the experiment were excluded. During the experimental period, any animal exhibiting severe weight loss (>15% of initial body weight), persistent diarrhea, or signs of distress was removed from the study and excluded from subsequent analyses.

Animals were housed under controlled environmental conditions (21–24°C; 55–65% humidity; 12 h light/dark cycle) with ad libitum access to food and water. All procedures complied with the European Directive 2010/63/EU on the protection of animals used for scientific purposes and were approved by the Institutional Animal Care and Use Committee of Ege University (Approval No: 2024-017, Date: 2024-03-27). The study was conducted in accordance with the ARRIVE 2.0 (Animal Research: Reporting of In Vivo Experiments) guidelines.

All behavioral evaluations and biochemical analyses were performed by investigators who were blinded to the experimental group assignments to minimize observer bias.

Sample size was calculated using G*Power 3.1 software based on a One-way ANOVA model, with an effect size of 0.8, α = 0.05, and a statistical power (1–β) of 0.90. This analysis indicated that a minimum of seven animals per group would be sufficient to detect significant intergroup differences.

Induction of demyelination with cuprizone

Acute demyelination was induced by administering cuprizone (400 mg/kg/day) mixed in 1% methylcellulose and vortexed into a homogeneous suspension, given by oral gavage for 5 weeks, as previously described [16]. This experimental model produces prominent demyelination in the corpus callosum. Control and vortioxetine-only animals received the same vehicle without cuprizone.

Vortioxetine administration

Vortioxetine was dissolved in 5% 2-hydroxypropyl-β-cyclodextrin and administered intraperitoneally at 10 mg/kg daily for 4 weeks. This dose was selected based on previous studies on cognitive and depression-like behavior [17]. The vehicle (5% 2-hydroxypropyl-β-cyclodextrin) was administered to control and cuprizone group.

Measurement of fecal excretion

To evaluate gastrointestinal motility, mice were individually placed in a 40 × 40 × 40 cm open field for the final 3 days of the experiment to acclimate. On the last day, fecal pellets produced during a 24-h period were collected, counted, and weighed (wet and dry weights) as previously described [18].

Tissue collection and biochemical analyses

At the end of the 5-week protocol, mice were anesthetized via a high-dose intraperitoneal injection of ketamine (80–100

https://doi.org/10.5455/annalsmedres.2025.08.231

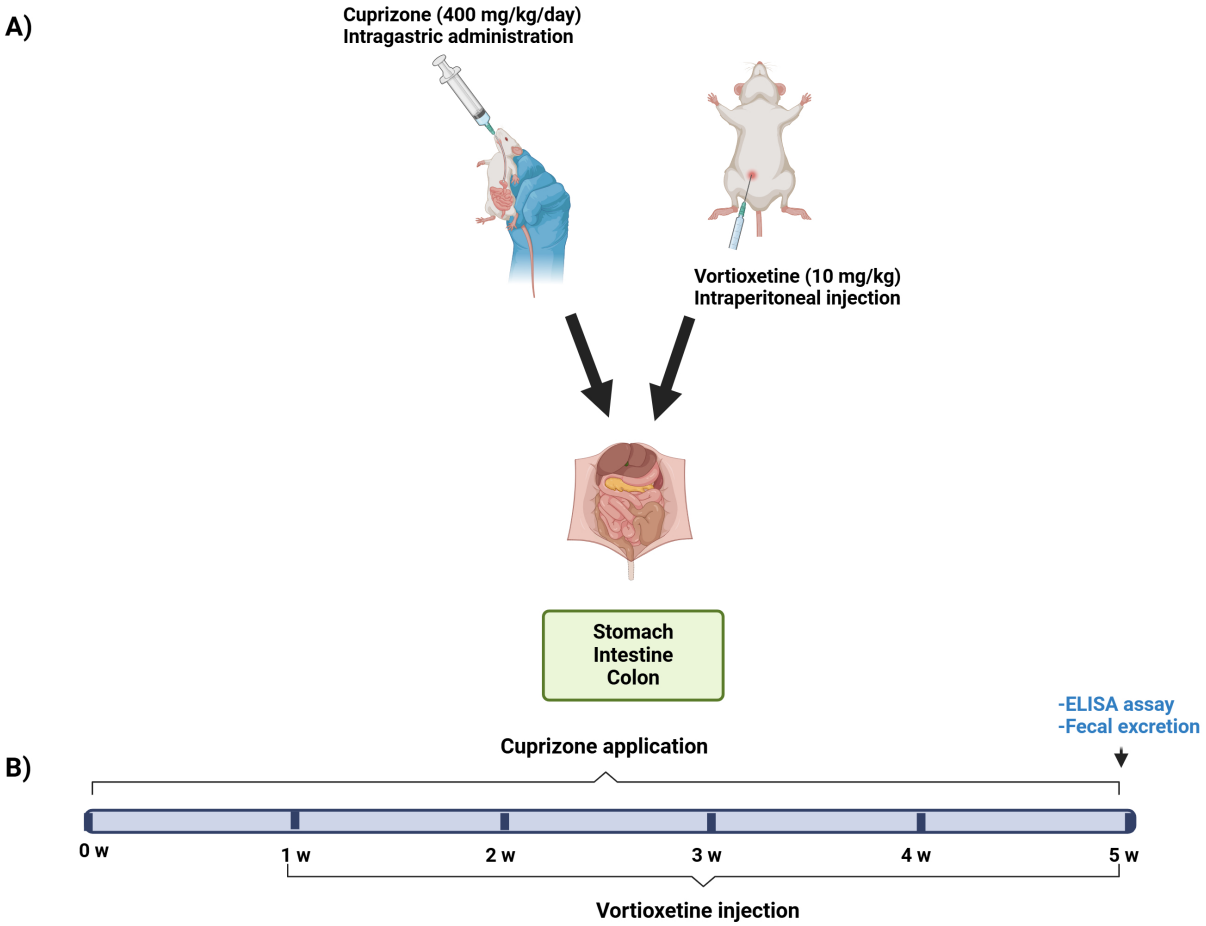

Figure 1. Experimental procedure. (A) Cuprizone (400 mg/kg/day) or vehicle was administered intraperitoneally. All murine tissues were extracted from the stomach, intestines, and colon. (B) Vortioxetine was administered intraperitoneally for 4 weeks, while cuprizone was administered intragastrically for 5 weeks. Upon completion of the experimental period, pro-inflammatory cytokines were assessed in all gastrointestinal tissues, and fecal excretion was quantified.

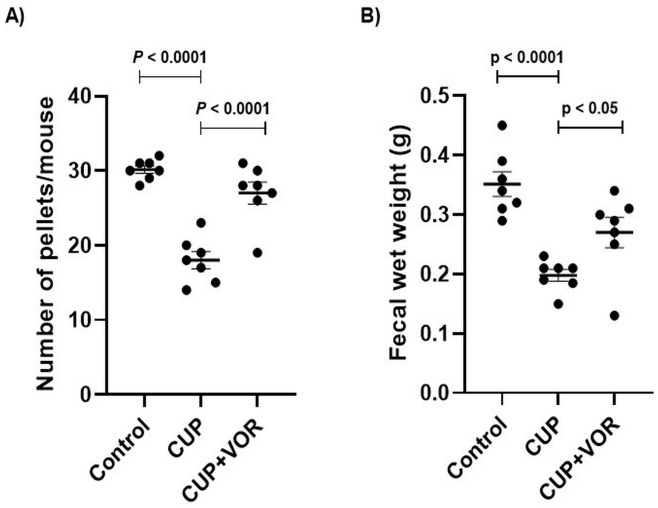

Figure 2. Fecal excretion analysis. (A) Number of pellets (24h) ($F(2, 18) = 31.51$; $P<0.0001$). (B) Fecal wet weight (g) ($F(2, 18) = 14.97$; $P<0.0001$). Values are mean ± SE of n = 7 animals. Statistical analysis was performed using One-way ANOVA followed by Sidak's post hoc test.

mg/kg) and xylazine (12 mg/kg), followed by cardiac perfusion with saline. Stomach, intestine, and colon tissues were collected, weighed, and homogenized in cold phosphate-buffered saline (PBS, pH 7.0–7.2). Homogenates were centrifuged at 5000 × g for 5 min, and supernatants were stored for cytokine analyses.

https://doi.org/10.5455/annalsmedres.2025.08.231

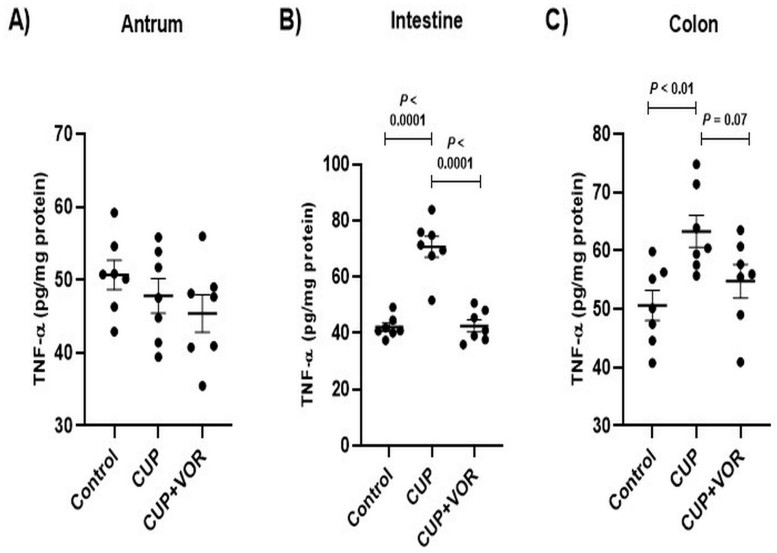

Figure 3. TNF-α levels in gastrointestinal tract tissues. (A) Antral TNF-α levels (F (2, 18) = 1.28; P>0.05). (B) Intestinal TNF-α levels (F (2, 18) = 38.96; P<0.0001). (C) TNF-α level in the colon (F (2, 18) = 5.62; P<0.05) Values are mean ± SE of n = 7 animals. Statistical analysis was performed using One-way ANOVA followed by Sidak's post hoc test.

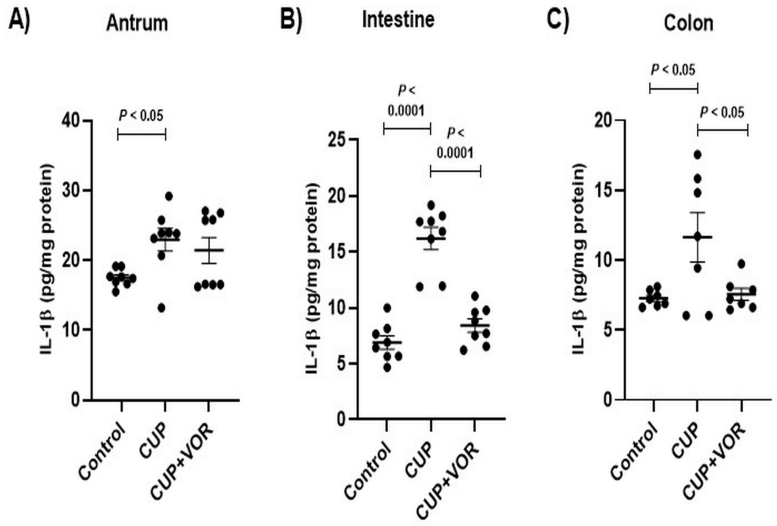

Figure 4. IL-1β levels in gastrointestinal tract tissues. (A) Antral IL-1β levels (F (2, 21) = 1.14; P>0.05). (B) Intestinal IL-1β levels (F (2, 21) = 44.35; P<0.0001). (C) IL-1β level in the colon (F (2, 18) = 5.35; P<0.05) Values are mean ± SE of n = 7 animals. Statistical analysis was performed using One-way ANOVA followed by Sidak's post hoc test.

ELISA for TNF-α and IL-1β

Concentrations of TNF-α and IL-1β in the stomach, intestine, and colon were determined using commercial ELISA kits (Elabscience, TNF-α: E-EL-M3063; IL-1β: E-EL-M0037) following the manufacturer's instructions. Results were expressed as pg/mg protein.

Statistical analysis

Statistical analysis were performed using GraphPad Prism 9 (Boston, USA). Data were first tested for normality using the Shapiro–Wilk test and for homogeneity of variances using Levene's test. For data that show normal distribution and equal variances, One-way ANOVA followed by Sidak's post hoc test was used to test intergroup differences. All tests were two-tailed, and differences were considered statistically significant at $p<0.05$. Results are presented as mean ± standard error (SE).

RESULTS

Fecal excretion

Quantification of fecal pellets over a 24 hour period showed a marked reduction in animals exposed to cuprizone com-

https://doi.org/10.5455/annalsmedres.2025.08.231

pared with controls (MS: 18.0±1.1 vs. Control: 30.1±0.5; P<0.0001; Figure 2A). Vortioxetine treatment significantly increased the number of fecal pellets in the CUP+VOR group (27.0±1.4; P<0.0001), approaching values observed in controls.

Similarly, wet fecal weight was significantly reduced in the cuprizone group (0.19 ± 0.009 g) compared to the control group (0.35±0.02 g; P<0.0001; Figure 2B). Daily intraperitoneal administration of vortioxetine for 5 weeks resulted in a significant increase in wet fecal weight (0.27±0.02 g; P<0.01 vs. cuprizone).

TNF α levels

TNF α concentrations in the gastric antrum did not differ significantly among the groups (Control: 50.6 ± 2.0; CUP: 47.8 ± 2.3; CUP+VOR: 45.4 ± 2.5 pg/mg protein; Figure 3A).

In the intestine, TNF α levels were significantly elevated in the cuprizone group (70.6±3.7 5 pg/mg protein) compared to the control group (42.1±1.4 5 pg/mg protein; P<0.0001). Vortioxetine treatment restored intestinal TNF α concentrations to near control levels (42.5±2.1 5 pg/mg protein; P< 0.0001 vs. cuprizone; Figure 3B).

In the colon, TNF α levels were also increased in the cuprizone group (63.3±2.7 5 pg/mg protein) compared with controls (50.6±2.5 5 pg/mg protein). Although vortioxetine administration reduced colonic TNF α concentrations, this reduction did not reach statistical significance (P=0.07; Figure 3C).

IL 1β levels

ELISA analysis revealed a significant elevation of IL 1β levels in the gastric antrum of cuprizone-induced dmyelinization model (22.9±1.63 5 pg/mg protein) compared with controls (17.5±0.43 5 pg/mg protein; P<0.05), whereas vortioxetine treatment did not produce a significant effect at this site (Figure 4A).

In contrast, intestinal IL 1β levels were markedly increased in the cuprizone group (16.2±0.9 5 pg/mg protein) compared to controls (6.8±0.5 5 pg/mg protein; P<0.0001). Vortioxetine administration significantly decreased intestinal IL 1β by approximately 50% (8.4±0.5 5 pg/mg protein; P<0.0001 vs. cuprizone; Figure 4B).

Similarly, colonic IL 1β concentrations were elevated in cuprizone mice (11.6±1.7 5 pg/mg protein) compared to controls and were significantly reduced following vortioxetine treatment (7.5±0.4 p5 pg/mg protein; P<0.05; Figure 4C).

DISCUSSION

This study demonstrates that cuprizone-induced demyelination disrupts gastrointestinal motility and increases intestinal pro-inflammatory cytokines. Importantly, vortioxetine treatment restored fecal output and attenuated TNF-α and IL-1β levels, suggesting a modulatory effect on gastrointestinal function and inflammation.

Gastrointestinal motility disorders are common in inflammatory and neurodegenerative diseases, particularly IBD and MS, and significantly affect quality of life [19-21]. Alterations in microbiota, visceral sensitivity, and CNS signaling contribute to these symptoms [22]. Serotonin which is the major neurotransmitter in the enteric nervous system, plays a key role in regulating intestinal motility and immune homeostasis [23,24]. The results of our study showed that vortioxetine improved motility and fecal consistency and they are supported by previous clinical observations, including Seddighnia et al. [25] who reported similar improvements in patients receiving vortioxetine. The majority of serotonin is produced in enterochromaffin cells of the gut, and specific receptors such as 5-HT7 and 5-HT3 modulate anti-inflammatory effects and motility [26,27]. These data suggest that serotonergic antidepressants, particularly multimodal agents such as vortioxetine, may exert benefits beyond mood regulation.

The gut–brain axis has gained attention as a key link between the CNS and gastrointestinal function. Experimental models, including experimental autoimmune encephalomyelitis (EAE) and cuprizone, have shown that demyelination is accompanied by gut microbiota changes, gliosis, and loss of enteric neurons, resulting in slowed intestinal motility [5,28,29]. Previous studies demonstrated that vortioxetine can modulate the gut microbiome and enhance synaptic plasticity [30]. These mechanisms may contribute to the improvements in motility observed in the present study.

There is a well-documented bidirectional association between MS and IBD, supported by shared genetic susceptibilities and immune mechanisms [3]. Dysregulation of cytokines, particularly TNF-α and IL-1β, plays a central role in the pathogenesis of both diseases [31,32]. Our findings that vortioxetine reduced intestinal cytokine levels highlight a potential role for serotonergic modulation in attenuating systemic inflammation in demyelinating disorders.

Limitations

This study has several limitations. First, histopathological analyses of gastrointestinal tissues were not performed; therefore, structural changes such as mucosal integrity and inflammatory cell infiltration could not be evaluated. Also, microbiota composition was not assessed, which would have provided valuable insights into the gut–brain axis. Finally, the study was limited to biochemical and motility endpoints and did not include behavioral or functional assessments of gastrointestinal activity. Future studies with larger cohorts, detailed histological and microbiome analyses, and additional behavioral assessments will be required to confirm and extend these findings.

https://doi.org/10.5455/annalsmedres.2025.08.231

CONCLUSION

In conclusion, this study demonstrates that vortioxetine ameliorates gastrointestinal dysmotility and attenuates intestinal pro-inflammatory cytokine expression in a cuprizone-induced demyelination model. These findings highlight the potential role of serotonergic modulation in restoring gut–brain axis integrity under neuroinflammatory conditions.

Based on these results, future research should focus on elucidating the underlying mechanisms through detailed histopathological, neurochemical, and microbiome analyses. In addition, studies using larger cohorts, different dosages, and long-term treatment protocols, as well as behavioral and functional assessments of gastrointestinal activity, are warranted. Translational studies in patients with multiple sclerosis, particularly those with gastrointestinal comorbidities, will be crucial to determine whether vortioxetine can offer therapeutic benefit beyond its antidepressant effects.

Acknowledgements: The authors would like to thank the staff of the Ege University Laboratory Animal Research Unit for their technical support.

Ethics Committee Approval: The study was approved by the Institutional Animal Care and Use Committee of Ege University (Approval No: 2024-017, Date: 2024-03-27).

Peer-review: Externally peer-reviewed.

Conflict of Interest: No potential conflicts of interest were reported by the authors in relation to the research, authorship, or publication of this article.

Author Contributions: A.O: Conceptualization, Methodology, Software, Data Curation, Investigation, Formal Analysis, Writing – Original Draft, Supervision; P.K: Investigation; Methodology.

Financial Disclosure: This research was supported by grants from the Coordination of Scientific Research Projects at İzmir Bakircay University, Izmir Independent Scientific Studies (BBAP.2023.004).

Artificial Intelligence Disclosure: The authors declare that an artificial intelligence based tool (ChatGPT, version 5.1) was used solely for language editing and for improving the clarity and readability of the manuscript. The AI tool was not used for study design, data collection, statistical analysis, interpretation of results, or the generation of scientific conclusions. All content was critically reviewed, validated, and approved by the authors, who take full responsibility for the accuracy, originality, and scholarly integrity of the manuscript.

REFERENCES

1. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. *Lancet.* 2018;391(10130):1622–36. doi: 10.1016/S0140-6736(18)30481-1.
2. Camara-Lemarroy CR, Metz LM, Yong VW. Focus on the gut-brain axis: Multiple sclerosis, the intestinal barrier and the microbiome. *World J Gastroenterol.* 2018;24(37):4217–23. doi: 10.3748/wjg.v24.i37.4217.
3. Yang Y, Musco H, Simpson-Yap S, Zhu Z, Wang Y, Lin X, et al. Investigating the shared genetic architecture between multiple sclerosis and inflammatory bowel diseases. *Nat Commun.* 2021;12(1):5641. doi: 10.1038/s41467-021-25768-0.
4. Nasirzadeh A, Jahanshahi R, Ghajarzadeh M, Mohammadi A, Moghadasi AN. The Prevalence of Inflammatory Bowel Disease (IBD) in Patients with Multiple Sclerosis (MS): A Systematic Review and Meta-Analysis. *Int J Prev Med.* 2023;14:81. doi: 10.4103/ijpvm.ijpvm_44_22.
5. Wang X, Eguchi A, Yang Y, Chang L, Wan X, Shan J, et al. Key role of the gut-microbiota-brain axis via the subdiaphragmatic vagus nerve in demyelination of the cuprizone-treated mouse brain. *Neurobiol Dis.* 2023;176:105951. doi: 10.1016/j.nbd.2022.105951.
6. Ghezzi L, Cantoni C, Pinget GV, Zhou Y, Piccio L. Targeting the gut to treat multiple sclerosis. *J Clin Invest.* 2021;131(13):e143774. doi: 10.1172/JCI143774.
7. Moles L, Egimendia A, Osorio-Querejeta I, Iparraguirre L, Alberro A, Suarez J, et al. Gut Microbiota Changes in Experimental Autoimmune Encephalomyelitis and Cuprizone Mice Models. *ACS Chem Neurosci.* 2021;12(5):893–905. doi: 10.1021/acschemneuro.0c00695.
8. Felger JC, Lotrich FE. Inflammatory cytokines in depression: neurobiological mechanisms and therapeutic implications. *Neuroscience.* 2013;246:199–229. doi: 10.1016/j.neuroscience.2013.04.060.
9. Wang C, Zhou Y, Feinstein A. Neuro-immune crosstalk in depressive symptoms of multiple sclerosis. *Neurobiol Dis.* 2023;177:106005. doi: 10.1016/j.nbd.2023.106005.
10. Harsanyi S, Kupcova I, Danisovic L, Klein M. Selected Biomarkers of Depression: What Are the Effects of Cytokines and Inflammation? *Int J Mol Sci.* 2022;24(1):578. doi: 10.3390/ijms24010578.
11. Talmon M, Rossi S, Pastore A, Cattaneo CI, Brunelleschi S, Fresu LG. Vortioxetine exerts anti-inflammatory and immunomodulatory effects on human monocytes/macrophages. *Br J Pharmacol.* 2018;175(1):113–24. doi: 10.1111/bph.14074.
12. Ye X, Wang D, Zhu H, Wang D, Li J, Tang Y, et al. Gut Microbiota Changes in Patients With Major Depressive Disorder Treated With Vortioxetine. *Front Psychiatry.* 2021;12:641491. doi: 10.3389/fpsyt.2021.641491.
13. Borgiani G, Possidente C, Fabbri C, Oliva V, Bloemendaal M, Arias Vasquez A, et al. The bidirectional interaction between antidepressants and the gut microbiota: are there implications for treatment response? *Int Clin Psychopharmacol.* 2025;40(1):3–26. doi: 10.1097/YIC.0000000000000533.
14. Guzel T, Mirowska-Guzel D. The Role of Serotonin Neurotransmission in Gastrointestinal Tract and Pharmacotherapy. *Molecules.* 2022;27(5):1680. doi: 10.3390/molecules27051680.
15. Terry N, Margolis KG. Serotonergic Mechanisms Regulating the GI Tract: Experimental Evidence and Therapeutic Relevance. *Handb Exp Pharmacol.* 2017;239:319–42. doi: 10.1007/164_2016_103.
16. Zhen W, Liu A, Lu J, Zhang W, Tattersall D, Wang J. An Alternative Cuprizone-Induced Demyelination and Remyelination Mouse Model. *ASN Neuro.* 2017;9(4):1759091417725174. doi: 10.1177/1759091417725174.
17. Nemutlu Samur D, Akcay G, Yildirim S, Ozkan A, Ceker T, Derin N, et al. Vortioxetine ameliorates motor and cognitive impairments in the rotenone-induced Parkinson's disease via targeting TLR-2 mediated neuroinflammation. *Neuropharmacology.* 2022;208:108977. doi: 10.1016/j.neuropharm.2022.108977.
18. Mule F, Amato A, Serio R. Gastric emptying, small intestinal transit and fecal output in dystrophic (mdx) mice. *J Physiol Sci.* 2010;60(1):75–9. doi: 10.1007/s12576-009-0060-8.
19. Bassotti G, Antonelli E, Villanacci V, Salemme M, Coppola M, Annese V. Gastrointestinal motility disorders in inflammatory bowel diseases. *World J Gastroenterol.* 2014;20(1):37–44. doi: 10.3748/wjg.v20.i1.37.

https://doi.org/10.5455/annalsmedres.2025.08.231

20. Khanna L, Zeydan B, Kantarci OH, Camilleri M. Gastrointestinal motility disorders in patients with multiple sclerosis: A single-center study. *Neurogastroenterol Motil.* 2022;34(8):e14326. doi: 10.1111/nmo.14326.
21. Spear ET, Holt EA, Joyce EJ, Haag MM, Mawe SM, Hennig GW, et al. Altered gastrointestinal motility involving autoantibodies in the experimental autoimmune encephalomyelitis model of multiple sclerosis. *Neurogastroenterol Motil.* 2018;30(9):e13349. doi: 10.1111/nmo.13349.
22. Mukhtar K, Nawaz H, Abid S. Functional gastrointestinal disorders and gut-brain axis: What does the future hold? *World J Gastroenterol.* 2019;25(5):552–66. doi: 10.3748/wjg.v25.i5.552.
23. Yabut JM, Crane JD, Green AE, Keating DJ, Khan WI, Steinberg GR. Emerging Roles for Serotonin in Regulating Metabolism: New Implications for an Ancient Molecule. *Endocr Rev.* 2019;40(4):1092–107. doi: 10.1210/er.2018-00283.
24. Banskota S, Khan WI. Gut-derived serotonin and its emerging roles in immune function, inflammation, metabolism and the gut-brain axis. *Curr Opin Endocrinol Diabetes Obes.* 2022;29(2):177–82. doi: 10.1097/MED.0000000000000713.
25. Seddighnia A, Tadayon Najafabadi B, Ghamari K, Noorbala AA, Ebrahimi Daryani N, Kashani L, et al. Vortioxetine effects on quality of life of irritable bowel syndrome patients: A randomized, double-blind, placebo-controlled trial. *J Clin Pharm Ther.* 2020;45(1):97–104. doi: 10.1111/jcpt.13032.
26. Chang WY, Yang YT, She MP, Tu CH, Lee TC, Wu MS, et al. 5-HT(7) receptor-dependent intestinal neurite outgrowth contributes to visceral hypersensitivity in irritable bowel syndrome. *Lab Invest.* 2022;102(9):1023–37. doi: 10.1038/s41374-022-00800-z.
27. Tao E, Zhu Z, Hu C, Long G, Chen B, Guo R, et al. Potential Roles of Enterochromaffin Cells in Early Life Stress-Induced Irritable Bowel Syndrome. *Front Cell Neurosci.* 2022;16:837166. doi: 10.3389/fncel.2022.837166.
28. Parodi B, Kerlero de Rosbo N. The Gut-Brain Axis in Multiple Sclerosis. Is Its Dysfunction a Pathological Trigger or a Consequence of the Disease? *Front Immunol.* 2021;12:718220. doi: 10.3389/fimmu.2021.718220.
29. Zhang W, Wang Y, Zhu M, Liu K, Zhang HL. Gut flora in multiple sclerosis: implications for pathogenesis and treatment. *Neural Regen Res.* 2024;19(7):1480–8. doi: 10.4103/1673-5374.387974.
30. Chen F, Danladi J, Ardalan M, Elfving B, Muller HK, Wegener G, et al. A Critical Role of Mitochondria in BDNF-Associated Synaptic Plasticity After One-Week Vortioxetine Treatment. *Int J Neuropsychopharmacol.* 2018;21(6):603–15. doi: 10.1093/ijnp/pyy022.
31. Aggeletopoulou I, Kalafateli M, Tsounis EP, Triantos C. Exploring the role of IL-1beta in inflammatory bowel disease pathogenesis. *Front Med (Lausanne).* 2024;11:1307394. doi: 10.3389/fmed.2024.1307394.
32. Souza RF, Caetano MAF, Magalhaes HIR, Castelucci P. Study of tumor necrosis factor receptor in the inflammatory bowel disease. *World J Gastroenterol.* 2023;29(18):2733–46. doi: 10.3748/wjg.v29.i18.2733.

https://doi.org/10.5455/annalsmedres.2025.08.231

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

The relationship between empathy and cognition in patients with tension type headache

Ugur Aylak[a], *, Ceyhun Sayman[b], Seyda Cankaya[b], Mehmet Ozansoy[c], Burak Yulug[b]

[a]Bahcesehir University, Graduate School, Department of Neuroscience, Istanbul, Türkiye
[b]Alanya Alaaddin Keykubat University, Faculty of Medicine, Department of Neurology, Antalya, Türkiye
[c]Medipol University, Faculty of Medicine, Department of Physiology, Istanbul, Türkiye
*Corresponding author: ugur.aylak@bahcesehir.edu.tr (Ugur Aylak)

MAIN POINTS

- Empathy alterations in tension-type headache (TTH) patients occur independently of global cognitive performance.
- Female patients show higher affective empathy compared to males.
- Age is associated with decreased cognitive function but increased affective empathy.
- Executive dysfunction is linked to heightened emotional empathy in TTH.

Cite this article as: Aylak U, Sayman C, Cankaya S, Ozansoy M. The relationship between empathy and cognition in patients with tension type headache. *Ann Med Res*. 2026;33(5):199–204. doi: 10.5455/annalsmedres.2025.08.243.

ABSTRACT

Aim: Tension-type headache (TTH) is the most common primary headache disorder, yet its impact on cognitive and empathic functions remains underexplored. This study investigates the relationship between cognitive domains and empathy in individuals with TTH, considering demographic influences.

Materials and Methods: A total of 71 TTH patients (49 females, 22 males) underwent cognitive assessment using the Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE), alongside the Empathy for Pain Scale (EPS) to evaluate cognitive and affective empathy. Statistical analyses included independent samples t-tests and Pearson correlation.

Results: No significant differences found were in terms of gender, general cognitive performance (MMSE, MoCA). However, females exhibited significantly higher affective distress, visceral pain empathy, and empathic concern scores. Males showed slower cognitive processing speed on the Trail Making Test. Age negatively correlated with cognitive function but positively with affective distress and empathic concern. Education was positively associated with cognitive scores but not with empathy measures. Executive dysfunction correlated with increased emotional empathy.

Conclusion: The findings reveal distinct gender-related differences in empathy among TTH patients, independent of global cognition, and highlight the complex interplay between executive function, age, education, and empathic processing. These insights underscore the need for integrative biopsychosocial approaches in TTH management.

Keywords: Headache, Cognitive impairment, Empathy, Pain

Received: Aug 28, 2025 **Accepted:** Dec 08, 2025 **Available Online:** May 22, 2026

INTRODUCTION

Tension-type headache (TTH) is the most prevalent primary headache disorder worldwide, characterized by bilateral, non-pulsating discomfort of mild to moderate intensity [1]. The etiology of TTH is considered multifactorial, involving a complex interplay between genetic predispositions and environmental influences [2]. For instance, Akcali et al. [3] suggested that polymorphisms in the serotonin transporter gene may be linked to TTH. Furthermore, musculoskeletal tension can be exacerbated by environmental stressors, poor urban living conditions, and suboptimal posture—all of which can prolong or aggravate headache symptoms [4, 5]. these multidimensional factors underscore the complex neurobiological and psychosocial underpinnings of TTH, highlighting the necessity of understanding how such influences affect higher-order psychological functions, including cognition and empathy.

While several studies have examined the cognitive and emotional consequences of migraine and neuropathic pain, TTH has received relatively little attention, particularly regarding its impact on empathic processes. This research gap may be attributed to a traditional focus on more debilitating headache forms, such as migraine, and the often-unrecognized long-term burden of TTH. Empathy is a critical social-cognitive ability that has been shown to be disrupted in various chronic pain conditions, potentially because pain-induced alterations

in cognitive processing interfere with these socio-emotional capacities [6].

In addition to these factors, the role of negative cognition—and the associated dysfunction of the anterior insula—is significant. Distorted thinking patterns, such as catastrophizing, hopelessness, and overgeneralization, are central to the cognitive domain of negative cognitions. Research suggests that such negative cognitions may generate as much distress as the subjective experience of chronic pain itself, emphasizing a profound emotional and cognitive dysregulation [7]. Although cognitive empathy is distinct from affective empathy—involving the mental construction or representation of another's experience—our study posits that disruptions to cognitive empathy may exist in TTH, potentially mediated by these maladaptive cognitive patterns.

Based on these findings, the present study aims to systematically explore the association between cognitive functions and empathy in individuals diagnosed with TTH. Specifically, we seek to determine whether alterations in executive functions, attention, and memory contribute to differences in cognitive and affective empathy compared to healthy controls. This study addresses the research problem that the cognitive–emotional interplay underlying TTH remains poorly understood, which may hinder the development of holistic treatment approaches. We hypothesize that: (1) individuals with TTH will demonstrate reduced cognitive empathy compared to controls; (2) these differences will be associated with impairments in specific cognitive domains; and (3) maladaptive cognitive patterns may mediate the relationship between headache severity and empathy performance.

Consequently, this study investigates the relationship between key cognitive domains—including executive functioning, attention, and memory—and types of empathy (cognitive and affective) using the Empathy for Pain Scale in TTH subjects [8]. By identifying potential cognitive mediators of empathy alterations in this population, we aim to broaden the understanding of TTH beyond its somatic manifestations toward a more comprehensive biopsychosocial model of care.

MATERIALS AND METHODS

Study design and participants

This is a prospective observational study. We included 71 patients diagnosed with tension-type headache (TTH) who were followed up at the Neurology Department of Alanya Training and Research Hospital between February 2023 and May 2024. The study was approved by the Ethics Committee of Alanya Alaaddin Keykubat University (ethics no: 16072023/289) and conducted in accordance with the Declaration of Helsinki and the guidelines outlined in the STROBE statement [9].

Diagnosis and grouping procedure

The diagnosis of TTH was established in accordance with the International Classification of Headache Disorders, 3rd edition (ICHD-3) criteria. Specifically, patients were required to present with bilateral location, a pressing or tightening (non-pulsating) quality, mild-to-moderate intensity, and a lack of aggravation by routine physical activity [10]. To ensure high diagnostic accuracy and minimize variability, all patients were evaluated by a neurology specialist with a minimum of five years of experience. This process involved structured clinical interviews and the application of standardized symptom criteria.

As this was an observational study, no randomization procedures were utilized for participant allocation; instead, all eligible patients meeting the predefined inclusion criteria were consecutively enrolled. Furthermore, blinding (masking) was not implemented, as the study design did not involve an interventional protocol or a specific control group comparison.

Assessment tools

The study employed the Montreal Cognitive Assessment (MoCA) [11] and the Mini-Mental State Examination (MMSE) [12] to evaluate global cognitive functioning. Both tools are well validated and frequently used in chronic pain populations to assess cognitive impairment [13].

As part of the visuospatial/executive subscale of the MoCA, the Trail Making Test Parts B (TMT) was administered to assess executive function, visual attention, and cognitive processing speed. Lower performance on the TMT reflects reduced mental flexibility and diminished executive control, which are often affected in patients with chronic pain or attentional deficits.

In addition, the Empathy for Pain Scale (EPS) was used to assess pain-specific cognitive and affective empathy [14]. Specific cognitive domains assessed included visual reconstruction, attention, sentence repetition, delayed recall, orientation, verbal fluency, naming, and abstract thinking [14,15]. Data related to age and educational status were recorded to control for potential confounding effects on cognitive and empathy outcomes [16]. State empathy, need-to-help, and compassion questions were used to evaluate empathic concern (EC), whereas affective distress (AD) was assessed using

Table 1. Demographic characteristics of patients.

		Percentage (%)
Age (Mean ± SD Range)	19-67 years (40.5 ± 13.2)	
Gender		
Male	22	30.9
Female	49	69
Education Status		
Primary School	24	33.8
Middle School	9	12.6
High School	21	29.5
University	17	23.9

https://doi.org/10.5455/annalsmedres.2025.08.243

Table 2. Summary of the analyses empathy and cognitive parameters according to gender.

Parameter	Female (Mean ± SD)	Male (Mean ± SD)	Levene's Test (p)	t-Test (p)
MMSE	28.6 ± 2.52	29 ± 1.57	0.201	0.508
MOCA	21.6 ± 4.53	23.5 ± 3.19	0.016	0.073
Visceral Pain	20 ± 10.8	11.4 ± 4.83	< 0.001	< 0.001
Affective Distress	86.2 ± 23.91	62.5 ± 13.84	0.014	< 0.001
Empathic Concern	50 ± 8.64	43.40 ± 10.7	0.162	0.007
Trail Making Test	0.59 ± 0.49	0.86 ± 0.35	< 0.001	0.024
Visuospatial	3.06 ± 1.10	3.45 ± 1.05	0.607	0.280
Naming	2.42 ± 0.67	2.50 ± 0.51	0.089	0.162
Attention	5.16 ± 1.10	5.54 ± 0.59	0.325	0.251
Sentence repetition	0.81 ± 0.83	0.77 ± 0.86	0.682	0.217
Verbal Fluency	0.61 ± 0.49	0.63 ± 0.49	0.694	0.126
Abstract Thinking	1.06 ± 0.82	1.18 ± 0.73	0.398	0.205
Memory	1.79 ± 1.51	2.63 ± 1.46	0.979	0.385
Orientation	5.71 ± 0.61	5.77 ± 0.68	0.672	0.163

This table presents mean and standard deviation values for neuropsychological test scores (MMSE, MOCA, MoCA subdomains), visceral pain, affective distress, and empathic concern between female and male participants. Levene's Test was used to assess homogeneity of variances, and the independent samples t-test was used to evaluate between-group differences. Values of $p < 0.05$ were considered statistically significant.

items addressing fear, distress, discomfort, disgust, restlessness, avoidance, and visceral sensations.

Primary outcome measures

The primary outcome variables were global cognitive scores (MoCA and MMSE) and empathy scores (EPS total, cognitive empathy, and affective empathy subscales). Secondary outcomes comprised Trail Making Test (TMT) scores, utilized as indicators of executive processing speed and cognitive flexibility, alongside the associations among these cognitive domains and various empathy parameters.

Statistical analysis

Statistical analyses were conducted using Jamovi software (version 2.3.28; GNU General Public License). Prior to analysis, all data were rigorously screened for completeness and accuracy. The normality of distribution for numerical variables was evaluated using the Shapiro–Wilk test. Continuous variables are presented as mean ± standard deviation (SD), while categorical variables are reported as frequencies and percentages.

For hypothesis testing, the independent samples t-test was employed to compare mean values between groups (e.g., gender), contingent upon the assumptions of normal distribution and homogeneity of variances, the latter of which was assessed via Levene's test. Pearson's correlation analysis was performed to examine the associations between empathy scores and cognitive measures. For all statistical evaluations, a p-value < 0.05 was established as the threshold for significance.

To validate the study's findings, a post hoc power analysis was conducted using G*Power (version 3.1.9.3, Universität Düsseldorf, Germany). With an alpha level of 0.05 and a large effect size (Cohen's d = 0.8), the analysis yielded a statistical power of 0.97, confirming that the sample size was sufficient to detect large effects with high reliability.

RESULTS

A total of 71 patients diagnosed with tension-type headache (TTH) were included in the study, comprising 49 females (69.0%) and 22 males (30.9%). The mean age of the participants was 40.5 ± 13.2 years (range: 19–67 years). Educational distribution was as follows: 33.8% were primary school graduates, 12.6% middle school, 29.5% high school, and 23.9% university graduates (Table 1).

Gender differences

Independent samples t-tests revealed significant gender-related differences in several empathy and executive parameters (Table 2). Before applying the t-tests, variance homogeneity was confirmed using Levene's test, ensuring that group comparisons met the assumption of equal variances.

Females demonstrated significantly higher Affective Distress (AD) scores (86.2 ± 23.91 vs. 62.5 ± 13.84, $p < 0.001$), higher Visceral Pain (VP) scores (20.0 ± 10.8 vs. 11.4 ± 4.83, $p < 0.001$), and greater Empathic Concern (EC) (50.0 ± 8.64 vs. 43.4 ± 10.7, $p = 0.007$) compared with males.

No significant gender differences were found in global cognitive performance, as measured by MMSE ($p = 0.508$) or MoCA ($p = 0.073$). However, a significant difference was detected in TMT scores, with males exhibiting higher mean values (0.86 ± 0.35) compared to females (0.59 ± 0.49, $p = 0.024$), suggesting slower executive processing in females.

No statistically significant differences were observed in other cognitive subdomains, including visuospatial ability ($p = 0.280$), naming ($p = 0.162$), attention ($p = 0.251$), sentence repetition ($p = 0.217$), verbal fluency ($p = 0.126$), abstract thinking ($p = 0.205$), memory ($p = 0.385$), and orientation ($p = 0.163$) (Table 2).

Correlation analysis

As presented in Table 3, age demonstrated a negative correlation with both MoCA ($r = -0.406$, $p < 0.001$) and MMSE

https://doi.org/10.5455/annalsmedres.2025.08.243

Table 3. Correlation between parameters such as education year, age, cognitive tests and empathic parameters.

According to Sociodemographic Variables			
Variable 1	Variable 2	r	p
Age	MOCA	-0.406	< .001
Age	MMSE	-0.250	0.035
Age	Tracking Test	-0.363	0.002
Age	Visuospatial	-0.226	0.058
Age	Naming	-0.027	0.825
Age	Attention	-0.202	0.091
Age	Repeat Sentences	-0.235	0.049
Age	Verbal Fluency	-0.073	0.546
Age	Abstract Thinking	-0.315	0.007
Age	Memory	-0.412	< .001
Age	Orientation	-0.229	0.055
Age	AD	0.285	0.016
Age	EC	0.290	0.014
Age	VP	0.126	0.297
Education	MOCA	0.596	< .001
Education	MMSE	0.347	0.003
Education	Tracking Test	0.369	0.002
Education	Visuospatial	0.585	< .001
Education	Naming	0.348	0.003
Education	Attention	0.327	0.005
Education	Repeat Sentences	0.346	0.003
Education	Verbal Fluency	0.478	< .001
Education	Abstract Thinking	0.388	< .001
Education	Memory	0.361	0.002
Education	Orientation	0.221	0.063
Education	AD	-0.135	0.260
Education	EC	-0.098	0.417
Education	VP	-0.119	0.323
According to Cognitive and Empathic Scores			
Variable 1	Variable 2	r	p
MOCA	MMSE	0.684	< .001
MOCA	AD	-0.148	0.218
MOCA	EC	0.104	0.386
MOCA	VP	−0.244	0.040
MMSE	Tracking Test	0.373	0.001
MMSE	Visuospatial	0.462	< .001
MMSE	Naming	0.298	0.012
MMSE	Attention	0.667	< .001
MMSE	Repeat Sentences	0.326	0.006
MMSE	Verbal Fluency	0.294	0.013
MMSE	Abstract Thinking	0.237	0.046
MMSE	Memory	0.284	0.017
MMSE	Orientation	0.799	< .001
MMSE	AD	-0.249	0.036
MMSE	EC	0.104	0.388
MMSE	VP	−0.293	0.013
Tracking Test	AD	-0.382	0.001
Tracking Test	EC	-0.211	0.077
Tracking Test	VP	−0.425	<.001

*MOCA: Montreal Cognitive Assessment Scale, MMSE: Minimental State Examition, AD: Affective Distress, VP: Visceral Pain, EC: Emphatic Concerne r: correlation coefficient.

(r = −0.250, p = 0.035) scores, indicating that cognitive performance declined with advancing age. Conversely, age correlated positively with Affective Distress (r = 0.285, p = 0.016) and Empathic Concern (r = 0.290, p = 0.014), reflecting increased emotional empathy in older individuals.

Education level was positively associated with cognitive test performance, including MoCA (r = 0.596, p < 0.001) and MMSE (r = 0.347, p = 0.003), but not significantly related to empathy parameters (p > 0.05).

Cognitive-executive measures showed inverse associations with emotional empathy: TMT scores correlated negatively with Affective Distress (r = −0.382, p = 0.001) and Visceral Pain (r = −0.425, p < 0.001), indicating that lower executive efficiency was associated with stronger emotional reactivity. Subgroup analyses further suggested that lower education intensified the relationship between cognitive impairment and empathic distress (Table 3).

DISCUSSION

This study investigated the relationships among empathy, cognition, and demographic factors in patients with Tension-type headache (TTH), revealing significant gender-based disparities, age-related patterns, and the moderating effects of education. Regarding empathic dimensions, female participants scored significantly higher in Vicarious Pain (VP), Affective Distress (AD), and Empathic Concern (EC). These findings align with previous literature suggesting that females generally demonstrate higher emotional responsiveness and empathic concern [8, 17]. Decety and Jackson [17] proposed a social-neuroscience framework wherein empathy integrates cognitive and affective processes, while Christov-Moore et al. [8] provided empirical evidence for gender-related variations in both neural activity and behavioral expression. Together, these perspectives suggest that while empathy relies on shared neurobiological systems, it is modulated by sex-dependent mechanisms.

Hormonal influences may also play a role; estrogen has been linked to enhanced emotional processing through prefrontal–limbic modulation [18], whereas testosterone may attenuate empathic distress [19]. Interestingly, while gender differences in empathy were pronounced, they did not extend to global cognitive performance as measured by MMSE and MoCA. However, males demonstrated superior performance on the Trail Making Test (TMT), reflecting more efficient executive processing speed and flexibility.

Age emerged as another significant influential factor in our cohort. Older participants exhibited higher AD and EC but lower MMSE and MoCA scores, aligning with findings that affective empathy tends to remain stable or even increase with age, whereas cognitive functions typically decline [20]. The inverse association between cognitive scores and empathy observed here suggests a potential trade-off, possibly reflecting diminished prefrontal regulation of emotional responses in older populations. The finding that reduced executive functioning was associated with heightened affective empathy further supports the hypothesis that decreased cognitive control may enhance emotional reactivity. Consequently, interventions such as cognitive training or mindfulness-based practices may be beneficial in regulating empathic distress in TTH populations [21].

https://doi.org/10.5455/annalsmedres.2025.08.243

Furthermore, educational attainment correlated positively with cognitive outcomes but not with empathy, supporting the cognitive reserve hypothesis, which suggests that higher education mitigates the impact of cognitive decline on broader social cognition [22]. From a neurobiological perspective, empathy involves the mirror neuron system and the prefrontal cortex [23]. These systems may be altered in TTH due to functional connectivity changes within pain-processing networks [24]. Chronic pain may further modulate the hypothalamic–pituitary–gonadal (HPG) axis, thereby influencing empathic behavior through specific neuroendocrine pathways.

The clinical implications of these findings are substantial. Enhancing cognitive flexibility and executive control through structured cognitive or empathy-focused interventions may mitigate empathic distress and improve emotional regulation in patients with TTH. Programs such as Cognitive-Behavioral Therapy (CBT), Mindfulness-Based Stress Reduction (MBSR), or compassion training could alleviate the psychosocial burden associated with chronic pain. Additionally, psychoeducation regarding empathy regulation may assist patients in developing adaptive coping strategies, ultimately improving quality of life. Future research should utilize larger, more diverse cohorts and employ longitudinal designs to establish causal relationships. Incorporating hormonal assays and neuroimaging assessments is warranted to further elucidate the underlying neurobiological mechanisms, and intervention-based studies are needed to test the efficacy of cognitive and empathic training protocols in TTH populations.

Limitations

This study has several limitations that should be acknowledged. First, the relatively small sample size may limit the generalizability of the findings to broader populations. Second, the cross-sectional design precludes the ability to draw causal inferences regarding the relationships between cognitive and empathic measures. Third, while validated instruments were utilized, the reliance on self-report scales introduces the potential for reporting bias and social desirability bias.

Furthermore, comorbid psychiatric symptoms—such as depression, anxiety, and other affective disorders—were not specifically assessed or statistically controlled in this study. Given that these conditions are known to influence both empathy and cognitive performance, their omission represents a potential confounding factor. Finally, hormonal fluctuations and medication use were not systematically monitored; these variables could further influence emotional and cognitive outcomes and should be considered in future research.

FUTURE DIRECTIONS

Future studies should overcome this by including healthy controls and employing a longitudinal design to track the fluctuation of empathy and cognition. Advanced neuroimaging techniques and hormonal profiling could give further mechanistic insights. Further, investigating the role of culture and psychosocial modulation may yield valuable perspectives in empathic responses.

Ethics Committee Approval: The study was approved by the Alanya Alaaddin Keykubat University Non-Interventional Clinical Research Ethics Committee (Decision number: 16072023/289).

Informed Consent: This prospective study was approved by the institutional ethics committee. All participants provided written informed consent in accordance with human ethics regulations before data collection began.

Peer-review: Externally peer-reviewed.

Conflict of Interest: No conflict of interest was declared by the authors.

Author Contributions: U.A: Concept and Design, Manuscript Writing; C.S: Data collection, Manuscript Writing; Ş.Ç: Analysis and Interpretation; M.O: Literature search; B.Y: Critical Review.

Financial Disclosure: The authors declared that this study has received no financial support.

Data Access Statement: Data supporting this study cannot be made available due to personal data protection law.

Artificial Intelligence Disclosure: The authors declare that no artificial intelligence tools were used in the preparation of this article.

REFERENCES

1. The International Classification of Headache Disorders, 3rd edition (ICHD-3). *Cephalalgia.* 2018;38(1) 1–211. doi: 10.1177/0333102417738202.
2. Pellesi L, Yangjeh A, Hajjaj I, Lababidi M, Sarwar F, et al. Neurotransmitter Imbalance in Tension-Type Headache: A Systematic Review of Mechanisms and Therapeutic Targets. *Pain Ther.* 2025;14(4):1279-1291. doi: 10.1007/s40122-025-00761-3.
3. Aylin A, Cengiz T, Emin E, Neriman A, Sacide P. Serotonin transporter gene polymorphisms in patients with chronic tension-type headache: a preliminary study. *Neurol India.* 2008;56(2):156-60. doi: 10.4103/0028-3886.41993.
4. Saylor D, Steiner TJ. The Global Burden of Headache. *Semin Neurol.* 2018;38(2):182-190. doi: 10.1055/s-0038-1646946.
5. Decety J, Jackson PL. The functional architecture of human empathy. *Behav Cogn Neurosci Rev.* 2004;3(2):71-100. doi: 10.1177/1534582304267187.
6. Fan Y, Duncan NW, de Greck M, Northoff G. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neurosci Biobehav Rev.* 2011 Jan;35(3):903-11. doi: 10.1016/j.neubiorev.2010.10.009.
7. Rusu AC, Gajsar H, Schlüter MC, Bremer YI. Cognitive Biases Toward Pain: Implications for a Neurocognitive Processing Perspective in Chronic Pain and its Interaction With Depression. *Clin J Pain.* 2019;35(3):252-260. doi: 10.1097/AJP.0000000000000674.
8. Christov-Moore L, Simpson EA, Coudé G, Grigaityte K, Iacoboni M, Ferrari PF. Empathy: gender effects in brain and behavior. *Neurosci Biobehav Rev.* 2014;46 Pt 4(Pt 4):604-27. doi: 10.1016/j.neubiorev.2014.09.001.

https://doi.org/10.5455/annalsmedres.2025.08.243

9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4(10):e296. doi: 10.1371/journal.pmed.0040296.
10. Olesen J. Classification of migraine and tension-type headache. *Cephalalgia.* 2023;43(4):3331024221139238. doi: 10.1177/03331024221139238.
11. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005;53(4):695-9. doi: 10.1111/j.1532-5415.2005.53221.x.
12. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12(3):189-98. doi: 10.1016/0022-3956(75)90026-6.
13. Moriarty O, McGuire BE, Finn DP. The effect of pain on cognitive function: a review of clinical and preclinical research. *Prog Neurobiol.* 201;93(3):385-404. doi: 10.1016/j.pneurobio.2011.01.002.
14. Preis MA, Kroener-Herwig B. Empathy for pain: the effects of prior experience and sex. *Eur J Pain.* 2012;16(9):1311-9. doi: 10.1002/j.1532-2149.2012.00119.x.
15. Tsai CF, Lee WJ, Wang SJ, Shia BC, Nasreddine Z, Fuh JL. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: validation of the Taiwanese version of the MoCA and an item response theory analysis. *Int Psychogeriatr.* 2012;24(4):651-8. doi: 10.1017/S1041610211002298.
16. Rossetti HC, Lacritz LH, Cullum CM, Weiner MF. Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology.* 2011;77(13):1272-5. doi: 10.1212/WNL.0b013e318230208a.
17. Decety J, Jackson PL. A social-neuroscience perspective on empathy. *Current Directions in Psychological Science.* 2006;15(2):54-58. doi: 10.1111/j.0963-7214.2006.00406.x.
18. Wu Y, Zhang Y, Yuan X, Guo J, Gao X. Influence of education level on MMSE and MoCA scores of elderly inpatients. *Appl Neuropsychol Adult.* 2023;30(4):414-418. doi: 10.1080/23279095.2021.1952588.
19. Rizzolatti G, Craighero L. The mirror-neuron system. *Annu Rev Neurosci.* 2004;27:169-92. doi: 10.1146/annurev.neuro.27.070203.144230.
20. Koenigs M, Tranel D. Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *J Neurosci.* 2007;27(4):951-6. doi: 10.1523/JNEUROSCI.4606-06.2007.
21. Zeidan F, Emerson NM, Farris SR, Ray JN, Jung Y, et al. Mindfulness Meditation-Based Pain Relief Employs Different Neural Mechanisms Than Placebo and Sham Mindfulness Meditation-Induced Analgesia. *J Neurosci.* 2015;35(46):15307-25. doi: 10.1523/JNEUROSCI.2542-15.2015.
22. Stern Y. What is cognitive reserve? Theory and research application of the reserve concept. *J Int Neuropsychol Soc.* 2002;8(3):448-60. PMID: 11939702.
23. Chung YS, Poppe A, Novotny S, Epperson CN, Kober H, et al. A preliminary study of association between adolescent estradiol level and dorsolateral prefrontal cortex activity during emotion regulation. *Psychoneuroendocrinology.* 2019;109:104398. doi: 10.1016/j.psyneuen.2019.104398.
24. Fernández-de-Las-Peñas C, Palacios-Ceña M, Valera-Calero JA, Cuadrado ML, Guerrero-Peral A, et al. Understanding the interaction between clinical, emotional and psychophysical outcomes underlying tension-type headache: a network analysis approach. *J Neurol.* 2022;269(8):4525-4534. doi: 10.1007/s00415-022-11039-5.

https://doi.org/10.5455/annalsmedres.2025.08.243

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Explainable deep learning for the automated classification of macular diseases in OCT images

Murat Firat [a], Ilknur Tuncer Firat [b,*], Taner Tuncer [c]

[a] Malatya Turgut Özal University, Faculty of Medicine, Department of Ophthalmology, Malatya, Türkiye
[b] İnönü University, Faculty of Medicine, Department of Ophthalmology, Malatya, Türkiye
[c] Fırat University, Faculty of Engineering, Department of Computer Engineering, Elazığ, Türkiye
*Corresponding author: ilknurtuncer89@gmail.com (Ilknur Tuncer Firat)

MAIN POINTS

- A ResNet50-based deep learning model was developed to automatically classify 7 macular diseases using only OCT images, achieving a high overall accuracy of 95%.
- The model demonstrated strong interclass discrimination, especially distinguishing with high precision between morphologically similar conditions such as DME, RVO, and RAO.
- Grad-CAM heat maps were used to visualize the decision-making areas of the model, enhancing the transparency and clinical interpretability of the AI system.
- The system is highly effective in correctly identifying normal OCT scans (99.5%), which is crucial for community screening and reducing unnecessary clinical workload.
- The model showed promising performance in detecting surgically managed diseases such as vitreomacular interface disorders (VIDs), supporting its integration into triage and referral systems.

Cite this article as: Firat M, Tuncer Firat I, Tuncer T. Explainable deep learning for the automated classification of macular diseases in OCT images. *Ann Med Res*. 2026;33(5):205–214. doi: 10.5455/annalsmedres.2025.07.206.

ABSTRACT

Aim: Optical coherence tomography (OCT) is a widely used, noninvasive, rapid, and high-resolution imaging technique for diagnosing and monitoring macular diseases. Despite its clinical value, OCT image interpretation is time-consuming and requires expert knowledge, which may lead to inconsistencies in diagnosis. The objective of this study is to create an AI-based model that reliably and effectively categorizes macular diseases from OCT images, offering a workable solution in environments with restricted access to ophthalmology specialists.

Materials and Methods: A convolutional neural network model based on ResNet50 architecture was developed to classify OCT images into seven categories: age-related macular degeneration (AMD), diabetic macular edema (DME), epiretinal membrane (ERM), retinal artery occlusion (RAO), retinal vein occlusion (RVO), vitreomacular interface disease (VID), and normal (NO) controls. Grad-CAM was employed to enhance the interpretability of the model and support clinical usability.

Results: The model's macro-averaged precision, recall, and F1-score were 0.943 (95% confidence interval [CI]: 0.941–0.960), 0.940 (95% CI: 0.941–0.960), and 0.940 (95% CI: 0.941–0.960), respectively, with an overall accuracy of 0.950 (95% CI: 0.941–0.960). Grad-CAM visualizations confirmed the model's focus on relevant retinal regions, thus supporting diagnostic reliability and interpretability.

Conclusion: The explainable model demonstrated strong diagnostic performance and potential as a clinical decision-support tool, especially in environments with limited resources. The integration of explainable AI techniques, such as Grad-CAM, enhances trust in automated decision-making and offers significant potential in supporting non-expert users and early detection strategies.

Keywords: Retinal diseases, Artificial intelligence, Optical coherence tomography, Maculopathy

Received: Aug 05, 2025 **Accepted:** Dec 16, 2025 **Available Online:** May 22, 2026

INTRODUCTION

The diagnosis and follow-up of retinal diseases in ophthalmology practice requires a high level of expertise and time. During this process, specialists frequently utilize various imaging techniques, including visual acuity measurement, biomicroscopy, fundus photography, fundus fluorescein angiography (FFA), fundus autofluorescence (FAF), optical coherence tomography (OCT), and OCT angiography (OCTA) [1,2]. However, the clinical interpretation of these methods is time-consuming and creates a significant workload for specialists working under high patient volumes. Reducing the burden on specialists through preliminary assessment systems is crucial for the early and accurate diagnosis of RDs. To achieve this, we need reliable, easily trainable, and accessible diagnostic support systems that can be used by general physicians. Such systems, which offer high accuracy, do not re-

quire advanced expertise, and are low-cost, have become possible through AI-based approaches [3,4,5]. This has led to a growing interest in automated classification systems that aim to reduce specialist dependency and increase diagnostic reliability and speed. AI-based decision support systems stand out for their potential to standardize the process and alleviate the diagnostic burden, especially in regions with limited specialist access [6]. Today, optical coherence tomography (OCT) images have become one of the gold standard methods for assessing retinal diseases and play a critical role in macular disease diagnosis [7]. AI-based OCT data analysis improves clinical performance and offers large-scale applicability in community screening and remote monitoring scenarios. Recently, diagnostic support systems based on ML algorithms have been proposed for identifying retinal diseases [8, 9]. These studies employed classical algorithms, such as support vector machine (SVM), k-nearest neighbor (K-NN), decision tree (DT), and ensemble model (EM), by manually extracting features from fundus or OCT images. However, because manual feature extraction is time-consuming, labor-intensive, and has limited generalization capacity, deep learning-based decision support systems that perform automatic feature extraction and classification have become more popular.

Numerous studies have focused on the detection of diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration. Choudhary et al. used a VGG19-based classifier to distinguish choroidal neovascularization, drusen, and diabetic macular edema (DME) from normal OCT images. Their study achieved 99.17% accuracy with 84,568 OCT images [10]. Kabataş et al. used 1397 fundus images to distinguish DR and macular hole (MH) from normals. They achieved the highest accuracy of 93.97% with the ResNet50, InceptionV3, and Xception models [11]. Sunil et al developed a computer-aided diagnosis (CAD) system using fundus images for the diagnosis of DR stages. The proposed CNN model achieved an accuracy of 98.33% [12]. In a study involving 29,800 OCT images, Eghtedar et al. achieved 97.6% accuracy using a deep neural network (DNN) [13]. Rajan et al used 3 different models—OctDeepNet, OctDeepNet1, and OctDeepNet2—to classify normal, CNV, DME, and drusen. They achieved 98% accuracy by adjusting the model architecture for factors such as the number of layers, kernel size, and pooling [14]. Wang et al classified 8 retinal diseases using the ViT model and compared the results with those of existing methods, such as ResNet, VGG, DenseNet, and MobileNet. The ODIR-2019 dataset consists of fundus images and has a class imbalance. Therefore, the F1 score and AUC values were examined, and 0.932 ± 0.06 and 0.950 ± 0.03 were obtained, respectively [15]. Laouarem et al. combined the strengths of CNNs and visual transformers to classify 7 retinal diseases. They achieved accuracies of 99.40%, 97.00%, and 99.77% on the OCT-2017, OCT-C8, and OCT-2014 datasets, respectively [16]. To distinguish CNV, drusen, DME, and normal samples from OCT images, Akça et al. used the Vision Transformer (ViT), Token-to-Token Vision Transformer (T2T-ViT), and Mobile Vision Transformer (Mobile-ViT) algorithms. Accuracy rates of 95.14%, 96.07%, and 99.17% was achieved with the ViT, T2T-ViT, and Mobile-ViT models, respectively [17].

Maculopathies and OCT findings

Age-related macular degeneration (AMD) is one of the most common degenerative pathologies affecting the macula and is the primary cause of central vision loss, particularly in individuals aged >50 years. AMD develops because of the disruption of the interaction between the RPE, Bruch's membrane, and the choroid. In the dry form, drusen accumulate beneath the RPE, gradually weakening macular function. Abnormal vessels originating from the choroid (choroidal neovascularization) migrate into the subretinal space in the wet form, causing bleeding and fluid leakage [18]. This condition can be observed in OCT images as subretinal and intraretinal fluid accumulation, PED, RPE irregularities, and increased retinal thickness [19].

The macula is highly susceptible to edema, degeneration, and ischemic processes due to its high metabolic activity and structural characteristics. Macular edema develops due to increased capillary permeability, particularly in diabetic retinopathy and retinal vein occlusions. Edema affects the macula more severely than the peripheral retina due to its higher photoreceptor density and different vascular structure. DME is the most important complication of diabetic retinopathy, leading to vision loss. On OCT, it is characterized by increased macular thickness, cystic spaces, hyperreflective areas indicating exudate in the inner retinal layers, and sometimes subretinal fluid [20]. RVO, which is associated with hypertension and systemic vascular disease, is the second most common retinal vascular disease. Macular edema is the primary cause of vision loss. OCT reveals macular edema, increased retinal thickness, and intraretinal cystic spaces [21]. RAO is one of the most severe forms of retinal ischemia and is often associated with embolic events. OCT reveals marked hyperreflectivity and thickening of the inner retinal layers in the early stage and atrophy in the late stage [22].

Disruptions in the interactions between the vitreous cortex and the retinal surface cause vitreoretinal interface diseases (VIDs). Abnormally tight adhesions, incomplete vitreous separation, and the resulting fibroglial proliferation characterize these pathologies [23]. Examples of these conditions include VMT, VMA, macular hole, and epiretinal membrane (ERM). The most common complaints in these conditions include decreased visual quality, metamorphopsia, and blurred vision. OCT findings may reveal persistent adhesion of the vitreous cortex to the macula (vitreomacular adhesion) or traction (vitreomacular traction), as well as full-thickness (full-thickness macular hole) or partial (lamellar macular hole) defects in the macula [24]. ERM, a condition of this group of diseases, is characterized by the formation of a fibroglial mem-

https://doi.org/10.5455/annalsmedres.2025.07.206

brane on the inner surface of the retina. OCT imaging reveals a hyperreflective membrane, flattening of the inner retinal layers, and macular distortion [25].

Contributions of this study to the literature:

- Reduces inter-expert interpretation differences,
- It reduces the burden of manual evaluation,
- It enables the differentiation of multiple diseases,
- Provides explainable and interpretable disease information.

In this study, we developed an explainable ResNet50-based OCT classifier that simultaneously distinguished 7 macular conditions and provided Grad-CAM visualizations to support clinical interpretability.

MATERIALS AND METHODS

OCT is a valuable tool for creating detailed retinal microstructure images and diagnosing ocular conditions. This study used an open-access dataset consisting of 2064 OCT images labeled by experts according to retinal pathology, with predefined disease classes that eliminate the need for additional subjective clinical interpretation [26]. This dataset was published to provide researchers with access to a large set of labeled images that contribute to the development and improvement of algorithms for automated processing and analysis of OCT images for the early diagnosis and monitoring of eye diseases. The dataset consists of seven classes and includes OCT images of age-related macular degeneration, diabetic macular edema, epiretinal membrane, retinal artery occlusion, retinal vein occlusion, vitreomacular interface disease, and normal macula. The dataset exhibits an imbalanced class distribution. To address this issue, each image in the classes with fewer samples (DME, ERM, RAO, RVO, VID, and normal macula) was rotated 3° to the left and right, and symmetry operations were applied. As a result, in each minority class, the final sample size was simply determined by multiplying the original sample size by the fixed augmentation scheme, which reduced but did not completely eliminate the original class imbalance. This yielded an augmented dataset of 3 824 images (original n = 2 064). Table 1 shows the class distributions of the resulting dataset. After random augmentation, the dataset was divided into 2 parts: 80% training and 20% testing.

Table 1. Class distribution of the dataset.

Class	Original	Augmented
Age-related macular degeneration	1231	1231
Diabetic macular edema	147	441
Epiretinal Membrane	155	493
Normal	332	996
Retinal artery occlusion	22	132
Retinal venous occlusion	101	303
Vitreomacular Interface Disease (VMID)	76	228

Explainable ResNet50 model

The ResNet50 model was chosen for the classification of retinal diseases in this study because of its ability to recognize complex patterns and reduce overfitting, particularly in small and detailed medical datasets. Heat maps generated by the Grad-CAM method were used to better understand the obtained results, facilitate the interpretation of OCT images in the clinical setting, and increase the explainability of the model's decisions. The details of the model are provided below.

ResNet50 is a residual architecture with a depth of 50 layers. Its primary layers are convolution, max pooling, sequential residual blocks, global average pooling, and a fully connected layer. Taking an $X \in R^{H \times W \times 3}$ image as input, the ResNet50 model uses a residual network structure to address the degradation problem (vanishing gradient). In this residual network block, the input information x is passed through the convolution → batch normalization → ReLU chain and combined with x again via the shortcut connection (Eq. 1).

$$y = F(x, \{W_i\}) + x \tag{1}$$

In the model, ($R^{2048} \rightarrow R^{c}$) was used instead of the original fully connected layer ($R^{2048} \rightarrow R^{1000}$) to represent the number of classes c. The difference between the model's output probabilities and the true labels was calculated using the cross-entropy loss function. The model output $z = f(x, \theta) \in R^c$ was used to calculate the softmax classification probabilities and the loss function (eq.2 and 3).

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{c} e^{z_j}} \tag{2}$$

$$L = -\sum_{i=1}^{c} y_i \log(\hat{y}_i) \tag{3}$$

Here, $y_i \in \{0, 1\}$ is one-hot coded.

Gradient-weighted Class Activation Mapping (Grad-CAM) was used to ensure the model's explainability of the classification results. In ResNet50, the activation maps in the final convolutional layer, which have the highest-level features for class distinction, were used for Grad-CAM. Activations indicate the pixel's importance in the classification.

For class c, weights are calculated as in Equation 4 using $y_c = f_c(x)$, the activation map $A^k \in R^{u \times v}$ obtained from the last convolution layer, and the gradients $\frac{(\partial y^c)}{(\partial A^k)}$ obtained by backpropagation.

$$\alpha_k^c = \frac{1}{Z} \sum_{i,j} \frac{\partial y^c}{\partial A_{i,j}^k} \tag{4}$$

https://doi.org/10.5455/annalsmedres.2025.07.206

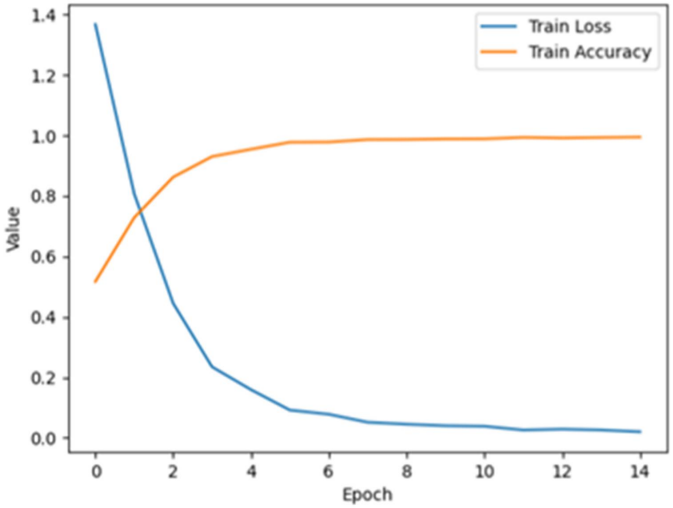

Figure 1. Training and validation of accuracy/loss curves across epochs for the ResNet50 model. These curves demonstrate the learning behavior and generalization performance of the model throughout the training process.

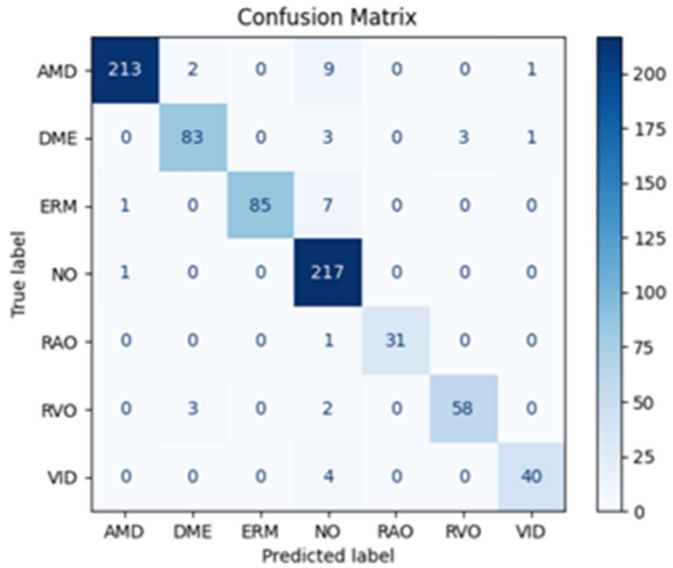

Figure 2. Confusion matrix summarizing the performance of the model across all 7 disease classes. Diagonal values indicate correct classifications, whereas off-diagonal values indicate misclassifications, enabling a visual analysis of class-wise accuracy.

Using the weights, the Grad-CAM heat map to be superimposed on the image is calculated as in eq.5.

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{k} \alpha^{c}_{k} A^{k}\right) \tag{5}$$

RESULTS

A computer environment with an Intel(R) Core(TM) i7-9750H CPU @ 2.60 GHz (2.59 GHz), 8 GB RAM, and Python programming language was used to train and test the ResNet50 model. The random seed was fixed to 42 in all experiments; the same seed was used for Python/NumPy/PyTorch, and data splitting and shuffling were performed with

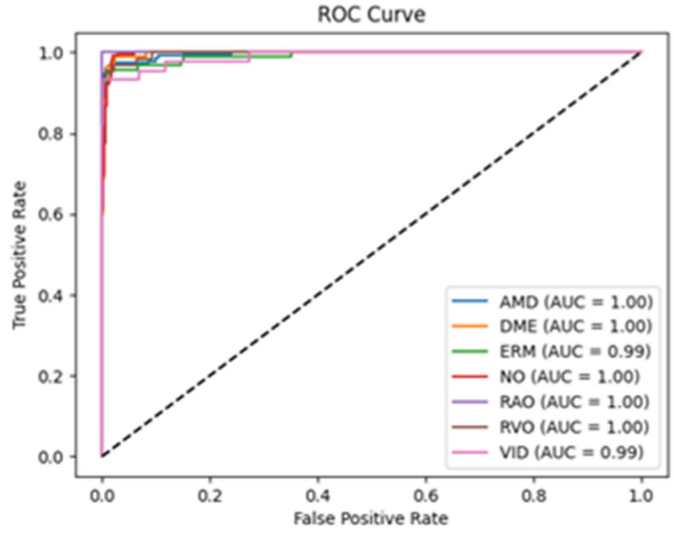

Figure 3. ROC curves for each class, illustrating the trade-off between sensitivity and specificity. AUC values provide a comparative metric for class-level performance.

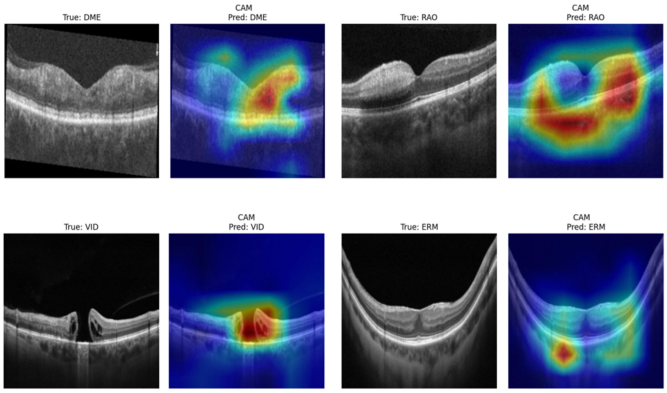

Figure 4. Examples of OCT images correctly classified with Grad-CAM–based attention heat maps. These visualizations demonstrate that the model accurately focuses on disease-relevant anatomical regions when making accurate predictions.

torch.Generator().manual_seed(42), CuDNN was set to deterministic and benchmark=False. Model performance parameters were analyzed using the accuracy, F1 score, precision, and sensitivity parameters obtained from the confusion matrix (eq.6-9). Additionally, ROC-AUC was used to determine model performance due to class imbalance in the dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = \frac{2TP}{2TP + FN + FP} \tag{9}$$

https://doi.org/10.5455/annalsmedres.2025.07.206

Table 2. Training parameters.

Parameter	Value	Definition
Pretrained	True	ImageNet loaded pre-trained weights
Num Classes	7	Outputs equal the number of classes in the dataset
Loss Function	CrossEntropyLoss	Suitable for multiclass classification
Optimizer	Adam	The Adam optimization algorithm
Learning Rate	1e-4	Initial learning rate (%)
Epochs	15	Total number of training cycles
Batch Size	32	Training and validation
Train/Test Split	80%/20%	Splitting the ratio into training and test sets
Input Size	(224, 224)	All images were resized to this size.
Grad-CAM Layer	model.layer4[-1]	Final convolution layer used for the Grad-CAM

Grad-CAM: Gradient-weighted Class Activation Mapping.

Table 3. Precision, recall, and F1 score by class.

Class	Precision	Recall	F1 Score
AMD	0.995	0.947	0.970
DME	0.932	0.922	0.927
ERM	0.867	0.914	0.890
NO	0.943	0.995	0.969
RAO	0.911	0.969	0.940
RVO	1.000	0.921	0.959
VID	0.952	0.909	0.930

AMD: Age-related Macular Degeneration, DME: Diabetic Macular Edema, ERM: Epiretinal Membrane, NO: Normal, RAO: Retinal Artery Occlusion, RVO: Retinal Vein Occlusion, VID: Vitreomacular Interface Disease.

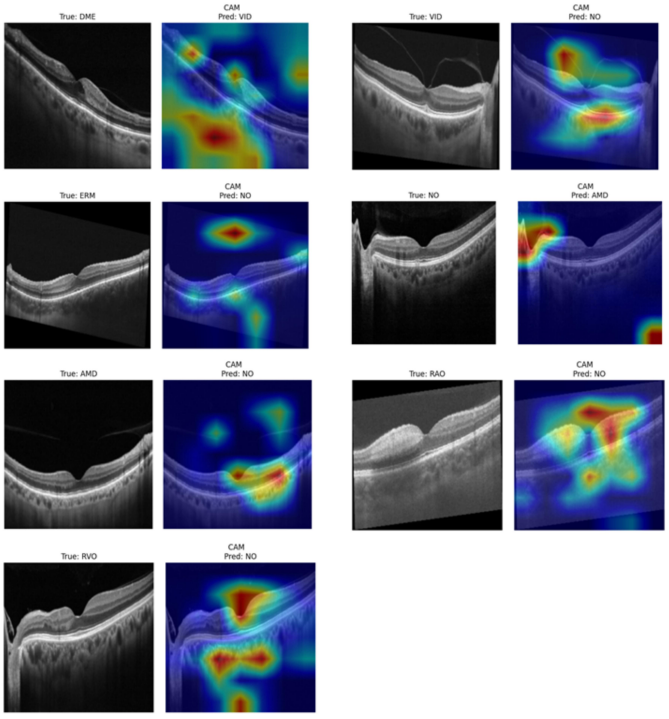

Figure 5. Examples of misclassified OCT images with Grad-CAM–based attention heat maps These cases illustrate the difficulty of the model in identifying subtle or ambiguous features and highlight the importance of explainability in evaluating model limitations.

Table 2 shows the parameters used for training the model. The dataset was first classified using the Vision Transformer, a widely used model in the literature, to demonstrate the superiority of the proposed model. Classification accuracy and f1-score values were 92.15% and 92%, respectively. Figure 1 shows the accuracy and loss function changes for training and validation as a result of the classification performed with ResNet50, Figure 2 shows the confusion matrix, and Figure 3 shows the ROC change.

The accuracy and loss graphs show that the loss value and accuracy are stable after the eighth epoch. This change indicates that the training process was successful. Furthermore, the receiver operating characteristic (ROC) curve was used to determine the discriminatory power of the model for each class. Since the change in all classes is close to the upper left corner, the model performed well and distinguished between classes. The AUC 95% confidence interval values were as follows: AMD, 0.996–1.000; DME, 0.994–1.000; ERM, 0.976–1.000; NO, 0.996–1.000; RAO, 0.991–1.000; RVO, 0.993–1.000; and VID, 0.969–1.000. The precision, recall, and F1 score values obtained from class after the elimination of the relative class imbalance are shown in the Table 3. The F1 score of 0.94, which is particularly relevant for class imbalance, indicates that the model successfully learned the dataset.

Heat maps (Grad-CAM) are created to visualize the regions to which the deep learning model pays attention during classification. This makes the decision-making process of the model more transparent and contributes significantly to clinical validation and explainability. Warm colors (red, orange, and yellow) represent diagnostically important regions during classification, whereas cool colors (blue tones) represent areas that the model ignores or deems unimportant for classification. Figures 4 and 5 present the correctly classified and misclassified images by the model, respectively, along with their corresponding heat maps of attention.

As shown in Figure 4, the Grad-CAM attention maps in correctly classified cases focused on the pathological regions. In the correctly classified DME OCT image, which is characterized by small cystic spaces and increased thickness due to fluid accumulation in the inner layers of the retina, the attention map is concentrated in the central and thickened inner layers of the retina. In the OCT image of a correctly classified macular hole, which is a type of VID characterized by a full-

https://doi.org/10.5455/annalsmedres.2025.07.206

Table 4. Comparison of AI-based studies using OCT images to classify macular disease.

Reference	Model	Disease	Dataset	Performance
Wang et al, 2016 [27]	SMO	Dry AMD (n =15), DME (n =15), and NO (n = 15)	3 000+ OCT images	Accuracy: 0.993
Li et al, 2019 [28]	ResNet50	CNV, DME, Drusen, and NO (n = 2 796)	15 573 OCT images	Accuracy: 0.973 AUC: 0.995 Sensitivity: 0.963 Specificity: 0.985
Li et al, 2019 [29]	VGG-16	CNV, DME, Drusen, and NO	109 312 OCT images	Accuracy: 0.986 Sensitivity: 0.978 Specificity: 0.994 AUC: 1.000
Kuwayama et al, 2019 [30]	CNN (Caffe)	Wet AMD (n = 136), DR (n = 104), ERM (n = 90), NO (n = 570), and other diseases (n = 19)	1 200 OCT images	Precision/recall: Wet AMD 1.000/0.770; DR 0.780/1.000; ERM 0.750/0.750; NO 0.850/0.970
Bhatia et al, 2020 [31]	VGG-16	Dry AMD, wet AMD, DME, and NO	5 588 OCT volumes (162 721 B-scans)	AUC: ≥0.980 AUC of dry AMD: 0.850–0.990 Wet AMD area under the curve: 0.860–0.980 DME AUC: 0.930–0.990
Esfahani et al, 2023 [32]	CNN	Drusen, CNV, DME, and NO	3 133 OCT images	Accuracy: 0.987 Precision: 0.99 Recall: 0.962 F1: 0.971
Choudhary et al, 2023 [10]	VGG-19	CNV, DME, Drusen, and NO	84 568 OCT images	Accuracy: 0.992 Specificity: 0.995 Sensitivity: 0.990
Kayadibi et al, 2023 [33]	FD-CNN	CNV, DME, Drusen, NO (n = 4696)	84 495 OCT images (Duke + UCSD)	Accuracy: 0.996 (UCSD) Accuracy: 0.975 (Duke)
Proposed Method	ResNet50	AMD, DME, ERM, RAO, RVO, VID, and NO	3 824 OCT images (after augmentation; original n= 2 064)	Accuracy: 0.950; Precision: 0.943; Recall: 0.940;

SMO: Sequential Minimal Optimization, AMD: Age-related Macular Degeneration, DME: Diabetic Macular Edema, CNV: Choroidal Neovascularization, DR: Diabetic Retinopathy, ERM: Epiretinal Membrane, VID: Vitreomacular Interface Disease, RAO: Retinal Artery Occlusion, RVO: Retinal Vein Occlusion, NO: Normal, OCT: Optical Coherence Tomography, AUC: Area Under the Curve, CNN: Convolutional Neural Network, FD-CNN: Fully Dense Convolutional Neural Network, UCSD: University of California San Diego, F1: F1 Score, ROC: Receiver Operating Characteristic, VGG: Visual Geometry Group, ResNet50: 50-layer Residual Network.

thickness defect in the central fovea, the attention map focuses on VMT and the central defect. Significant hyperreflectivity and thickening of the inner retinal layers are observed in the OCT image of a correctly classified RAO. The attention map focuses on retinal thickening. In the correctly classified ERM example, the foveal contour is flattened. The attention map focuses on the flattened foveal contour.

Notably, there are mild cases where the features of the misclassified images are not evident. Mild cystic changes in the inner retinal layers typical of DME are observed at the center of the macula in the DME case that was misclassified as VID. It is noteworthy that there is slight retinal thickening. The attention map is shifted to the peripheral retinal regions. In the ERM OCT image that was misclassified as a normal macula, a thin hyperreflective line was observed adjacent to the inner retinal surface. The attention map is focused on the vitreous over the retina. In the OCT image of AMD that was misclassified as a normal macula, a slightly irregular RPE structure is observed, consistent with AMD. The attention map focuses primarily on the inner retinal layers. In the RVO OCT image that was misclassified as a normal macula, irregularity and mild retinal thickness increase are present in the inner retinal layers affecting the nasal fovea due to RVO. Cystoid macular edema was not evident. The attention map focuses on the peripheral lines rather than the central pathological area. In the VID OCT image that was misclassified as a normal macula, the vitreous cortex has a very small base, causing mild macular shrinkage, resulting in VID. The attention map focused on the adherent vitreous cortex but was misclassified. In the OCT image of a healthy macula that was misclassified as AMD, the optic disc was mistakenly evaluated as a sign of AMD. In the OCT image of an RAO case, hyperreflectivity and minimal thickening of the inner retinal layers are observed. The model successfully focused on retinal thickening but still misclassified it as normal.

https://doi.org/10.5455/annalsmedres.2025.07.206

DISCUSSION

The ResNet50 model demonstrated high performance in distinguishing 7 different macular diseases using only OCT images. The model achieved an overall accuracy of 0.95 (95% CI: 0.9408–0.9600), an F1-score of 0.94 (95% CI: 0.9405–0.9600), a precision of 0.943 (95% CI: 0.9408–0.9600), and a recall of 0.94 (95% CI: 0.9405–0.9600). These consistently high values across all metrics highlight the model's strong inter-class discrimination capability and its potential for clinical decision support system integration.

Unlike many previous studies in the literature that focused primarily on common retinal pathologies, our study includes a broader classification spectrum by incorporating less frequently addressed conditions, such as RAO, RVO, and VID. Although early studies, such as those by Wang et al. [27] and Li et al. [28, 29], achieved high accuracy with limited disease classes, their clinical applicability remains restricted. Kuwayama et al. [30] and Bhatia et al. [31] expanded class diversity but faced limitations in sample size or reliance on commercial systems, which may reduce generalizability. Similarly, Choudhary [10], Esfahani [32] and Kayadibi [33] reported high accuracy but were limited to more balanced and prevalent disease classes. In contrast, our study offers a more comprehensive approach by successfully classifying seven retinal conditions using real-world data, achieving robust performance (accuracy: 0.95, F1 score: 0.94), and demonstrating compatibility with clinical scenarios—highlighting its originality and practical relevance. We extend the typical four-class OCT taxonomy by adding RAO, RVO, and VID, resulting in a seven-class scheme that is aligned with clinical referral and optimized for triage. We also implemented a ViT baseline; although the ViT study (17) reported 99.17% accuracy on the four-class OCT2017 benchmark, our seven-class, referral-oriented task—including RAO/RVO/VID—is inherently harder, and our model achieves Accuracy 0.95 and F1 0.94. Furthermore, considering the prevalence of these diseases [34-36], providing early diagnosis opportunities with a low-cost and accessible AI-based screening system could offer significant gains for public health [37]. Previous studies similar to ours are summarized and presented in Table 4 for comparison.

In terms of model architecture, while many existing works employed VGG-16/19 or custom-built architectures (e.g., FD-CNN) [10, 29, 33], we utilized the ResNet50 model, which balances network depth, generalizability, and computational efficiency through residual connections. Furthermore, the integration of Grad-CAM–based visual explanations enhances interpretability and supports clinical trust in the model's outputs—an aspect often lacking in other studies [27,32]. Compared with related work, our model performs reliably even with imbalanced datasets, offers transparent decision-making through explainable AI, and uses an open-source architecture that promotes practical implementation. Collectively, these features position our model as a technically robust and clinically applicable contribution to the current literature.

The 99.5% accuracy of normal samples is a critical advantage, particularly for community screening and preliminary assessment systems. Low false-positive rates in normal individuals increase patient confidence and prevent unnecessary burden on the health care system. This supports the utility of AI-based systems as triage tools. Additionally, high precision in detecting normal cases can improve system efficiency in large-scale screening programs by reducing the number of cases referred for unnecessary specialist evaluation.

AMD detection with 94.67% accuracy is highly valuable for early diagnosis and timely intervention. Early-stage vitamin supplementation, risk factor management, and timely initiation of anti-vascular endothelial growth factor therapies, which can be applied in wet cases, are key to reducing vision loss [38]. This aligns with global efforts to reduce preventable vision impairment due to AMD through earlier therapeutic interventions.

The ability of the model to distinguish morphologically similar vascular pathologies such as DME (92.22%), RVO (96.67%), and RAO (96.88%) is crucial for triage systems. Despite the structural similarities between DME and RVO on macular OCT, the model's misdiagnosis of these 2 diseases is only 2.0%. This is a particularly valuable distinction in clinical decision-making processes, such as anti-VEGF treatment planning and systemic assessment differentiation [39]. The model may aid in prompt therapeutic direction by providing automated differentiation of these pathologies, especially in settings with limited retinal specialists.

The ability of the model to distinguish VID with 90.91% accuracy is a significant achievement from both technical and clinical perspectives. The VID class includes pathologies, such as VMT, VMA, ERM, and macular hole, that may require surgical management. The correct identification of these cases is essential for timely surgical referral, potentially preventing irreversible visual deterioration in progressive cases.

The ERM, a subtype of the VID class, was classified by the model with 91.4% accuracy. This result demonstrates that ERM is highly successful in identifying cases with isolated ERM, as it frequently accompanies other groups [24].

The most significant difference between VID and other macular diseases is that many cases are managed surgically rather than medically. In particular, macular holes and advanced ERM require vitreoretinal surgery, such as vitrectomy [40]. Accurate identification of VID can reduce time lost not only in terms of disease diagnosis but also by directing patients to the appropriate specialist (vitreoretinal surgeon). This can minimize unnecessary examinations, re-evaluations, and surgical interventions. Therefore, the model's VID performance may improve referral efficiency and reducing time to interven-

https://doi.org/10.5455/annalsmedres.2025.07.206

tion.

Attention heat maps visually represent the OCT image regions that the model focuses on when making classification decisions. The model was found to effectively identify anatomical regions with disease-specific structural abnormalities in correctly classified examples. In misclassified cases, the pathologies that reveal class characteristics in the OCT images were subtle, and the attention maps were generally concentrated in scattered or irrelevant regions. Beyond low disease severity, attenuation of acute signs over time (e.g., subacute/chronic phases) may also hinder feature extraction from a single B-scan, and weak cues can divert attention toward artifacts, extramacular structures (such as vitreous hyperreflectivity), or disk-containing slices, as illustrated by the Grad-CAM examples. Methodologically, these errors could be mitigated by incorporating phase metadata (acute vs. chronic), applying lesion-centric attention regularization with penalties for off-target regions, re-weighting borderline cases via focal loss, and enforcing slice quality control (excluding peripapillary, tilted, or low signal-to-noise scans). These observations that ambiguous or borderline findings may challenge the generalizability of the model, emphasizing the potential role of human-AI collaboration to improve diagnostic safety in such cases. Consistent with these observations, misdiagnoses in our cohort mainly arose from confusion between diseased classes with overlapping OCT appearances (particularly DME, ERM, VID, and RVO), rather than between diseased and normal eyes, and were exacerbated by the limited sample size of some categories.

Limitations

However, certain limitations must be explicitly acknowledged. First, this study used a single-center dataset, which may limit the model's generalizability across different populations and imaging devices. Second, the retrospective nature of the dataset could introduce selection bias, as the data may not reflect the full clinical spectrum of MDs. Third, class imbalance in the dataset may still impact model sensitivity in underrepresented disease groups, although partially addressed through augmentation. Accordingly, prospective validation using multicenter, balanced datasets is necessary to confirm performance across diverse settings. In addition, we did not perform calibration analysis (e.g., ECE, Brier score, and reliability diagrams) or ablation analyses, which limited our assessment of probability calibration and component effects. Finally, the study did not incorporate multimodal data or clinical parameters, which could further enhance classification accuracy and clinical relevance in future models. Moreover, because data augmentation was performed before the train–test split, augmented variants of the same original image may have been present in both sets, potentially leading to a slight overestimation of test performance. Therefore, future work with larger, more balanced multicenter datasets, multimodal inputs, and class-balanced or cost-sensitive training strategies may help reduce such errors.

CONCLUSION

In this study, the ResNet50 model was evaluated as an effective AI approach for the high-accuracy classification of macular diseases from OCT images. In particular, the accurate identification of normal individuals, successful differentiation of common diseases such as AMD, and effective classification of VID requiring surgical intervention hold great potential for clinical decision-making. The high interclass discrimination power of the model enabled selective results even among diseases with similar OCT morphology. To the best of our knowledge, our seven-class design—adding RAO, RVO, and VID—goes beyond four-class OCT benchmarks and better mirrors referral-driven workflows, enabling triage and timely surgical referral. These findings support the utility of AI-supported systems in population-based screening, primary care, and clinical triage. Future studies should test the model's generalizability and clinical applicability with larger and more heterogeneous datasets.

Ethics Committee Approval: Our study did not involve invasive procedures on humans or animals and used a publicly available open-access optical coherence tomography dataset. Therefore, approval from the ethics committee was not necessary.

Informed Consent: Not necessary for this manuscript.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflicts of interest to declare.

Author Contributions: M.F: Conseption, Design, Supervision, Data Collection and/or Processing, Analysis and/or Interpretation, Writing, Critical Review; İ.T.F: Conseption, Design, Supervision, Analysis and/or Interpretation, Writing, Critical Review; T.T: Design, Materials, Data Collection and/or Processing, Analysis and/or Interpretation, Literature Review, Writing, Critical Review.

Financial Disclosure: This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

Artificial Intelligence Disclosure: The authors declare that an artificial intelligence–based tool (ChatGPT 5) was used solely for language editing and improving the clarity and readability of the manuscript. The AI tool did not contribute to the study design, data analysis, interpretation of results, or scientific conclusions. All content was critically reviewed and approved by the authors, who take full responsibility for the manuscript.

REFERENCES

1. Wong TY, Sabanayagam C. Strategies to Tackle the Global Burden of Diabetic Retinopathy: From Epidemiology to Artificial Intelligence. *Ophthalmologica*. 2020;243(1):9-20. doi: 10.1159/000502387.

https://doi.org/10.5455/annalsmedres.2025.07.206

2. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for the detection of DR in primary care offices. *NPJ Digit Med.* 2018;1:39. doi: 10.1038/s41746-018-0040-6.
3. Stankiewicz A, Marciniak T, Budna N, Chwałek R, Dziedzic M. Classification of OCT Images of the Human Eye Using Mobile Devices. *Appl Sci.* 2025;15(6):2937. doi: 10.3390/app15062937.
4. Akinniyi O, Rahman MM, Sandhu HS, El-Baz A, Khalifa F. Multi-Stage Classification of Retinal OCT Using Multi-Scale Ensemble Deep Architecture. *Bioengineering (Basel).* 2023;10(7):823. doi: 10.3390/bioengineering10070823.
5. Khan A, Pin K, Aziz A, Han JW, Nam Y. Optical Coherence Tomography Image Classification Using Hybrid Deep Learning and Ant Colony Optimization. *Sensors (Basel).* 2023;23(15):6706. doi: 10.3390/s23156706.
6. Heydon P, Egan C, Bolter L et al. Prospective evaluation of an AI-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol.* 2021;105(5):723-728. doi: 10.1136/bjophthalmol-2020-316594.
7. Song Y, Wang Q, Du Y, et al. Prognostic biomarkers of treatment-naïve central retinal vein occlusion with macular edema. *Eur J Med Res.* 2025;30(1):594. doi: 10.1186/s40001-025-02884-x.
8. Özdaş MB, Uysal F, Hardalaç F. Classification of Retinal Diseases in Optical Coherence Tomography Images Using Artificial Intelligence and Firefly Algorithm. *Diagnostics (Basel).* 2023;13(3):433. doi: 10.3390/diagnostics13030433.
9. Abd El-Khalek AA, Balaha HM, Alghamdi NS, et al. A concentrated machine learning-based classification system for age-related macular degeneration (AMD) diagnosis using fundus images. *Sci Rep.* 2024;14(1):2434. doi: 10.1038/s41598-024-52131-2.
10. Choudhary A, Ahlawat S, Urooj S, Pathak N, Lay-Ekuakille A, Sharma N. A Deep Learning-Based Framework for Retinal Disease Classification. *Healthcare (Basel).* 2023;11(2):212. doi: 10.3390/healthcare11020212.
11. Kabataş B, Ölmez E. Deep learning-based decision support system for the classification of retinal diseases: diabetic retinopathy and macular hole. *Artif Intell Theory Appl.* 2025;5(1):51-62.
12. Sunil SS, Shrivindhya A. Computer-aided detection of diabetic retinopathy on fundus images via statistical classification. 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS). doi: 10.1109/IC3SIS54991.2022.9885359.
13. Alizadeh Eghtedar R, Vard A, Malekahmadi M, Peyman A. A new computer-aided diagnosis tool based on deep learning methods for automatic detection of retinal disorders from OCT images. *Int Ophthalmol.* 2024;44(1):110. doi: 10.1007/s10792-024-03033-9.
14. Rajan R, Kumar SN. Deep learning architectures for OCT images retinal disease classification. *SN Comput Sci.* 2025;6(2):155. doi: 10.1007/s42979-025-03715-w.
15. Wang D, Lian J, Jiao W. Multilabel classification of retinal disease via a novel vision transformer model. *Front Neurosci.* 2024;17:1290803. doi: 10.3389/fnins.2023.1290803.
16. Laouarem A, Kara-Mohamed C, Bourennane EB, Hamdi-Cherif A. HTC-retina: A hybrid retinal diseases classification model using transformer-Convolutional Neural Network from optical coherence tomography images. *Comput Biol Med.* 2024;178:108726. doi: 10.1016/j.compbiomed.2024.108726.
17. Akça S, Garip Z, Ekinci E, Atban F. Automated classification of choroidal neovascularization, diabetic macular edema, and drusen from retinal OCT images using vision transformers: a comparative study. *Lasers Med Sci.* 2024;39(1):140. doi: 10.1007/s10103-024-04089-w.
18. Sheth JU, Stewart MW, Narayanan R et al. Macular neovascularization. *Surv Ophthalmol.* 2025;70(4):653-675. doi: 10.1016/j.survophthal.2024.08.003.
19. Spaide RF, Jaffe GJ, Sarraf D, et al. Consensus Nomenclature for Reporting Neovascular Age-Related Macular Degeneration Data: Consensus on Neovascular Age-Related Macular Degeneration Nomenclature Study Group. *Ophthalmology.* 2020;127(5):616-636. doi:10.1016/j.ophtha.2019.11.004.
20. Das R, Spence G, Hogg RE, Stevenson M, Chakravarthy U. Disorganization of Inner Retina and Outer Retinal Morphology in Diabetic Macular Edema. *JAMA Ophthalmol.* 2018;136(2):202-208. doi: 10.1001/jamaophthalmol.2017.6256.
21. Song Y, Wang Q, Du Y, et al. Prognostic biomarkers of treatment-naïve central retinal vein occlusion with macular edema. *Eur J Med Res.* 2025;30(1):594. doi: 10.1186/s40001-025-02884-x.
22. Ahn SJ, Woo SJ, Park KH, Jung C, Hong JH, Han MK. Retinal and choroidal changes and visual outcome in central retinal artery occlusion: an optical coherence tomography study. *Am J Ophthalmol.* 2015;159(4):667-676. doi: 10.1016/j.ajo.2015.01.001.
23. Phillips JD, Hwang ES, Morgan DJ, Creveling CJ, Coats B. Structure and mechanics of the vitreoretinal interface. *J Mech Behav Biomed Mater.* 2022;134:105399. doi: 10.1016/j.jmbbm.2022.105399.
24. Duker JS, Kaiser PK, Binder S, et al. The International Vitreomacular Traction Study Group classification of vitreomacular adhesion, traction, and macular hole. *Ophthalmology.* 2013;120(12):2611-2619. doi: 10.1016/j.ophtha.2013.07.042.
25. Besagar S, Chennupati S, Ji X, Chen Q, Thomas AS, Finn AP. Optical coherence tomography biomarkers in the epiretinal membrane. *Ophthalmic Surg Lasers Imaging Retina.* 2025;56(7):422-428. doi: 10.3928/23258160-20250416-01.
26. Kulyabin M, Zhdanov A, Nikiforova, et al. OCTDL: Optical coherence tomography dataset for image-based deep learning methods. *Sci Data.* 2024;11(1):365. doi: 10.1038/s41597-024-03182-7.
27. Wang Y, Zhang Y, Yao Z, Zhao R, Zhou F. Machine learning-based detection of age-related macular degeneration and diabetic macular edema from optical coherence tomography images. *Biomed Opt Express.* 2016;7(12):4928-4940. doi: 10.1364/BOE.7.004928.
28. Li F, Chen H, Liu Z, et al. Deep learning-based automated detection of retinal diseases using OCT images. *Biomed Opt Express.* 2019;10(12):6204-6226. doi: 10.1364/BOE.10.006204.
29. Li F, Chen H, Liu Z, Zhang X, Wu Z. Fully automated detection of retinal disorders using image-based deep learning. *Graefes Arch Clin Exp Ophthalmol.* 2019;257(3):495-505. doi: 10.1007/s00417-018-04224-8.
30. Kuwayama S, Ayatsuka Y, Yanagisono D, et al. Automated detection of macular diseases using optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images. *J Ophthalmol.* 2019;2019:6319581. doi: 10.1155/2019/6319581.
31. Bhatia KK, Graham MS, Terry L, et al. Disease classification of macular optical coherence tomography scans using deep learning software: validation on independent, multicenter data. *Retina.* 2020;40(8):1549-1557. doi: 10.1097/IAE.0000000000002640.
32. Riazi Esfahani P, Reddy AJ, Nawathey N, et al. Deep learning classification of drusen, choroidal neovascularization, and diabetic macular edema in OCT images. *Cureus.* 2023;15(7):e41615. doi: 10.7759/cureus.41615.
33. Kayadibi İ, Güraksın GE. An explainable fully dense fusion neural network with deep support vector machine for retinal disease determination. *Int J Comput Intell Syst.* 2023;16(1):28. doi: 10.1007/s44196-023-00210-z.
34. GBD 2021 Global AMD Collaborators. Global burden of vision impairment due to age-related macular degeneration, 1990–2021, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Glob Health.* 2025;13(7):e1175-e1190. doi: 10.1016/S2214-109X(25)00143-3.
35. Candan Ö, Orman G, Ünlü N, Üney G, Burcu A. Prevalence of retinal vascular diseases in a Turkish tertiary care hospital in Türkiye: a hospital-based epidemiologic study. *Turk J Ophthalmol.* 2025;55(1):16-23. doi: 10.4274/tjo.galenos.2025.88262.

https://doi.org/10.5455/annalsmedres.2025.07.206

36. McKibbin M, Farragher T, Shickle D. Vitreoretinal interface abnormalities in middle-aged adults with visual impairment in the UK Biobank study: prevalence, impact on visual acuity and associations. *BMJ Open Ophthalmol.* 2017;1(1):e000057. doi: 10.1136/bmjophth-2016-000057.
37. Heloterä H, Viita AM, Laine J. The evolution of workload associated with anti-VEGF treatments for AMD, DME, RVO, and mCNV in the Hospital District of Southwest Finland. *Clin Ophthalmol.* 2024;18:3645-3655. doi: 10.2147/OPTH.S479816.
38. Guymer RH, Rosenfeld PJ, Ehlers JP, et al. Designing the next generation of clinical trials in intermediate AMD—a consensus driven, pragmatic, proof-of-concept early intervention study. *Exp Eye Res.* 2025;255:110340. doi: 10.1016/j.exer.2025.110340.
39. Padilla-Pantoja FD, Sanchez YD, Quijano-Nieto BA, Perdomo OJ, Gonzalez FA. Etiology of Macular Edema Defined by Deep Learning in Optical Coherence Tomography Scans. *Transl Vis Sci Technol.* 2022;11(9):29. doi: 10.1167/tvst.11.9.29.
40. Er-Reguyeg Y, Doukkali S, Hébert M et al. Anatomical and functional outcomes of lamellar macular hole and epiretinal membrane foveoschisis surgery: predictive factors and associated complications: a retrospective interventional study. *Clin Ophthalmol.* 2025;19:1365-1376. doi: 10.2147/OPTH.S499493.

https://doi.org/10.5455/annalsmedres.2025.07.206

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

One-year cardiovascular risk assessment after laparoscopic sleeve gastrectomy: A retrospective cohort study

Adnan Gundogdu[a], *, Sangar Abdullah[a], Mustafa Kagan Basdogan[a], Guney Ozkaya[a], Ibrahim Aydin[a]

[a]Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital, Department of General Surgery, Istanbul, Türkiye
*Corresponding author: adnangundogdu89@gmail.com (Adnan Gundogdu)

MAIN POINTS

- Sleeve gastrectomy reduced Framingham cardiovascular risk score by 52.7% at one year, independent of BMI change magnitude.
- HDL cholesterol demonstrated a paradoxical biphasic pattern with initial decrease followed by increase above baseline levels.
- Age was a strong negative predictor of risk reduction, emphasizing the importance of optimal surgical timing.
- Patients with higher baseline cardiovascular risk achieved greater absolute risk reduction.
- The lack of correlation between BMI change and cardiovascular risk reduction suggests benefits beyond simple weight loss.

Cite this article as: Gundogdu A, Abdullah S, Basdogan MK, Ozkaya G, Aydin I. One-year cardiovascular risk assessment after laparoscopic sleeve gastrectomy: A retrospective cohort study. *Ann Med Res*. 2026;33(5):215–221. doi: 10.5455/annalsmedres.2025.08.239.

ABSTRACT

Aim: Bariatric surgery effectively reduces cardiovascular risk, though the mechanisms remain incompletely understood. This study aimed to evaluate changes in Framingham cardiovascular risk scores and identify predictors of risk reduction following sleeve gastrectomy.

Materials and Methods: This retrospective cohort study included 111 patients (73% female; median age, 33 years [range 18–63]; median BMI, 43 kg/m² [range 35–77]) who underwent laparoscopic sleeve gastrectomy between January 2021 and June 2024. Framingham 10-year cardiovascular disease risk scores were calculated at baseline and 12 months postoperatively. Multivariable linear regression identified independent predictors of absolute risk reduction.

Results: The Framingham risk score decreased from 6 (-10–25) to 3 (-9–26) at 12 months ($p<0.001$), representing a 50% relative risk reduction. BMI decreased from 43 (35–77) to 29 (21–55) kg/m² ($p<0.001$). Notably, change in BMI was not correlated with cardiovascular risk reduction ($r = 0.032$, $p = 0.735$). HDL cholesterol demonstrated a biphasic trend—initially decreasing at 3 months, then exceeding baseline values at 12 months ($p<0.001$). In multivariable analysis, age ($\beta = -0.223$, $p<0.001$); baseline Framingham score ($\beta = 0.532$, $p<0.001$); and HDL change ($\beta = 0.279$, $p = 0.047$) were independent predictors of the absolute risk reduction. Interestingly, an increase in HDL showed a negative association in univariable analysis but a positive association in multivariable analysis.

Conclusion: Sleeve gastrectomy achieved a significant reduction in cardiovascular risk independent of the magnitude of weight loss. The HDL paradox and age-dependent response suggest that complex mechanisms beyond simple weight reduction are involved. These findings support incorporating cardiovascular risk assessment, not just BMI, when determining surgical indications.

Keywords: Sleeve gastrectomy, Framingham risk score, Cardiovascular risk, HDL paradox, Bariatric surgery, Weight loss

Received: Aug 25, 2025 **Accepted:** Dec 18, 2025 **Available Online:** May 22, 2026

INTRODUCTION

Obesity has become a major global health problem and is recognized as an independent risk factor for cardiovascular disease (CVD) [1, 2]. The prevalence of obesity continues to rise worldwide; in the United States, it increased from 30.5% to 42.4% between 2000 and 2018, while severe obesity nearly doubled from 4.7% to 9.2% [2]. Obesity is strongly associated with metabolic disorders such as hypertension, diabetes, and dyslipidemia, thereby contributing to increased CVD morbidity and mortality [2, 3].

The effects of obesity on the cardiovascular system are multifactorial. Increased adipose tissue elevates metabolic demand, blood volume, and cardiac workload, potentially leading to left ventricular hypertrophy and diastolic dysfunction [2]. Moreover, obesity promotes endothelial dysfunction via oxidative stress and the release of pro-inflammatory cytokines [3].

Laparoscopic sleeve gastrectomy is one of the most widely used and effective bariatric procedures for morbid obesity. In addition to significant weight loss, this surgery has been

shown to improve cardiovascular risk factors [4]. According to Framingham data, each 1-unit increase in BMI is associated with a 5% higher risk of heart failure in men and a 7% higher risk in women [5]. A recent study found that sleeve gastrectomy reduces visceral adiposity, improves retinal microvascular parameters, and significantly decreases the Framingham Risk Score [6].

The aim of this study was to investigate changes in the Framingham cardiovascular risk score and its determinants following sleeve gastrectomy. Understanding these changes may contribute to the development of personalized treatment strategies in bariatric surgery.

MATERIALS AND METHODS

Study design and population

This single-center, retrospective cohort study evaluated changes in cardiovascular risk among patients who underwent laparoscopic sleeve gastrectomy between January 2021 and June 2024. Only patients with at least 12 months' follow-up were included. Data were extracted from electronic medical records at baseline, 3, 6, and 12 months postoperatively.

This study was approved by the Ethics Committee of Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital (Decision No: 2025/263, Date: 18.07.2025).

Inclusion and exclusion criteria

Inclusion criteria

- Age 18–65 years, BMI $\geq$ 40 kg/m^2 or $\geq$ 35 kg/m^2 with obesity-related comorbidities
- Underwent primary laparoscopic sleeve gastrectomy
- Complete 12–month follow-up data available.

Exclusion criteria

- History of cardiovascular disease (CAD, MI, revascularization, stroke, TIA, PAD)
- Previous or revision bariatric surgery
- Incomplete baseline or 12–month laboratory/blood pressure data
- Additional procedures beyond sleeve gastrectomy
- Inability to calculate Framingham risk score.

Data collection

- Demographics: Age, sex
- Anthropometrics: BMI at baseline and 12 months
- Biochemistry: Total cholesterol, HDL cholesterol at all time points
- Cardiovascular: Systolic blood pressure at baseline, 3, and 12 months; antihypertensive use
- Lifestyle: Smoking status at all visits.

Cardiovascular risk assessment

The primary endpoint was the change in the Framingham 10–year CVD risk score [7]. Absolute and relative risk reductions were calculated.

Surgical technique

A standardized laparoscopic sleeve gastrectomy was performed using a 36–38-French bougie, starting 4 cm from the pylorus, with complete fundus mobilization. Staple line reinforcement and leak testing were routinely performed.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics version 25.0 (Armonk, NY: IBM Corp.). Continuous variables were expressed as median (min–max) and categorical variables as n (%). Between-group comparisons of continuous and categorical variables were performed using the Mann–Whitney U test and the chi-square test, respectively. Within-group comparisons between two time points were analyzed using the Wilcoxon signed-rank test, while comparisons across three or more time points were evaluated with the Friedman test followed by Bonferroni-adjusted post hoc pairwise comparisons. Changes in binary outcomes, such as smoking status, were assessed using Cochran's Q test with Bonferroni-adjusted McNemar tests where applicable.

Correlations between continuous variables were assessed using Spearman's rank correlation coefficient (ρ). Independent predictors of absolute cardiovascular risk reduction were determined through multivariable linear regression analysis. Model assumptions—including linearity, independence, and homoskedasticity—were verified through residual plots, Q–Q plots, and the Shapiro–Wilk test. Heteroskedasticity-robust (HC3) standard errors were applied, and multicollinearity was assessed using the variance inflation factor (VIF < 2.5). Model performance was reported with R^2 and adjusted R^2 values.

As this was a retrospective study that included all eligible patients during the study period, no a priori sample size calculation was performed. However, an exploratory post hoc power analysis for the multivariable regression model was conducted using G*Power version 3.1.9.2 (Universität Düsseldorf, Germany). Based on the observed model fit (R^2 = 0.413, Cohen's $f^2 \approx 0.70$), with n = 111, k = 5, and α = 0.05, the achieved power (1–β) was estimated to be to exceed 0.90, indicating adequate statistical power to detect large effects.

A two-tailed p-value < 0.05 was considered statistically significant.

RESULTS

Baseline characteristics

A total of 111 patients [81 females (73%) and 30 males (27%)] were included in the final analysis. The median age of the study population was 33 years (range, 18–63), and the median

https://doi.org/10.5455/annalsmedres.2025.08.239

Table 1. Baseline patient characteristics and differences between groups.

Characteristic	All Patients	Female	Male	p-value
Age (years) median (min--max)	33 (18--63)	34 (18--57)	30 (19--63)	0.705
BMI (kg/m²) median (min--max)	43 (35--77)	43 (35--77)	43 (36--53)	0.651
Total cholesterol (mg/dL) median (min--max)	191 (127--316)	191.5 (127--305)	191 (160--316)	0.700
HDL cholesterol (mg/dL) median (min--max)	46 (26--70)	46 (26--66)	44.5 (27--70)	0.735
Systolic blood pressure (mmHg) median (min--max)	120 (80--170)	120 (80--170)	110 (90--160)	0.799
Antihypertensive use n(%)	22 (19.8)	18 (22.2)	4 (13.3)	0.297
Current smoking n(%)	53 (47.7)	39 (48.1)	14 (46.7)	0.890
Framingham score (%) median (min--max)	6 (-10--25)	6 (-8--23)	6 (-10--25)	0.239

Values are presented as median (min–max) or n (%); BMI:body mass index; HDL: high-density lipoprotein; $p < 0.05$ was considered statistically significant; †Mann–Whitney U test was used for continuous variables and Chi-square test for categorical variables.

Table 2. Temporal changes in cardiovascular risk parameters.

Parameter	Preop	3 months	6 months	12 months	p-value
BMI (kg/m²) median (min--max)	43 (35--77)			29 (21--55)	<0.001**†
Total cholesterol (mg/dL) median (min--max)	191 (127--316)[a,b]	187 (110--320)[e]	183(114--300)	178 (124--337)	<0.001**‡
HDL cholesterol (mg/dL) median (min--max)	46 (26--70)	44 (26--65)	48 (28--72)	50 (33--70)	<0.001**‡
Blood pressure (mmHg) median (min--max)	120 (80--170)			110 (80--140)	<0.001**†
Framingham score (%)median (min--max)	6 (-10--25)			3 (-9--26)	<0.001**†
Current smoking n(%)median (min--max)	53 (47.7)	42 (37.8)	35 (31.5)	34 (30.6)	<0.001**§

[a]: Preop–3 months; [b]: Preop–6 months; [c]: Preop–12 months; [d]: 3–6 months; [e]: 3–12 months; [f]: 6–12 months; †Wilcoxon signed-rank test; §Cochran's Q test; ‡Friedman test with Bonferroni-adjusted post hoc analysis; Values are presented as median (min–max) or n (%); $p < 0.05$ was considered statistically significant.

Table 3. Gender-based comparison of bmı and framingham risk score changes.

Parameter	Female (n=81)	Male (n=30)	p-value
BMI measurements			
Preoperative BMI (kg/m²) (kg/m²), median (min--max)	43 (35--77)	43 (36--53)	0.651
Postoperative 12-month BMI (kg/m²), median (min--max)	29 (21--55)	28.5 (26--37)	0.338
BMI change (kg/m²), median (min--max)	15 (4--23.6)	14.55 (10--19)	0.963
BMI change (%), median (min--max)	31.8 (10--52.91)	32.9 (27.03--40.43)	0.106
Framingham risk score			
Preoperative score (%), median (min--max)	6 (-8--23)	3.5 (-10--25)	0.239
Postoperative 12-month score(%) median (min--max)	3 (-8--19)	2 (-9--26)	0.724
Absolute risk reduction(%)‡ median (min--max)	3 (-9--15)	2 (-7--8)	0.159
Relative risk reduction(%)δ median (min--max)	37.4 (-128.57--500)	14.4 (-500--500)	0.105

‡Absolute risk reduction = Preoperative – 12-month value, δRelative risk reduction = (Absolute reduction / Preoperative value) × 100. †Mann–Whitney U test, Values are presented as median (min–max), $p < 0.05$ was considered statistically significant.

preoperative BMI was 43 kg/m² (range, 35–77). The median baseline Framingham 10-year cardiovascular risk score (FRS) was 6% (range: -10 to 25). There were no significant differences in baseline characteristics between genders ($p>0.05$; Table 1).

Changes in cardiovascular risk parameters

At 12 months, significant improvements were observed across all cardiovascular risk parameters (Table 2). BMI decreased from 43 (35–77) to 29 (21–55) kg/m² ($p<0.001$), systolic blood pressure from 120 (80–170) to 110 (80–140) mmHg ($p<0.001$), and total cholesterol.

The values progressively declined from 191 (127–316) mg/dL at baseline to 178 (124–337) mg/dL at 12 months ($p<0.001$). HDL cholesterol demonstrated a biphasic response, decreasing to 44 (26–65) mg/dL at 3 months and subsequently rising above baseline to 50 (33–70) mg/dL at 12 months ($p<0.001$).

Framingham risk score changes

The FRS decreased significantly from 6 (-10–25) preoperatively to 3 (-9–26) at 12 months ($p < 0.001$; Table 2). When analyzed by gender, no significant differences were observed in absolute risk reduction between females and males: females, 3% (−9–15); males, 2% (−7–8); $p = 0.159$ (Table 3). The temporal trajectory of risk reduction is shown in Figure 1.

Predictors of risk reduction

In the correlation analysis, absolute risk reduction was negatively associated with change in HDL ($r = -0.235$, $p = 0.013$) and positively associated with reduction in systolic blood pressure ($r = 0.399$, $p<0.001$), but not associated with change in BMI ($r = 0.032$, $p = 0.735$) (Table 4). As shown in Figure

https://doi.org/10.5455/annalsmedres.2025.08.239

Table 4. Univariable and multivariable analyses of predictors of absolute cardiovascular risk reduction

(a) Univariable analysis of factors associated with absolute risk reduction.

Variable	Correlation with Absolute Risk Reduction	
	r	p
Change in BMI (kg/m²)	0.032	0.735
Change in HDL cholesterol (mg/dL)	-0.235*	0.013
Change in systolic blood pressure (mmHg)	0.399**	<0.001
Antihypertensive discontinuation	Median (min - max)	
Discontinued (n=104)	6 (-9--15)	0.738
Continued (n=7)	11 (-4--10)	

(b) Multivariable linear regression analysis for predictors of absolute risk reduction.

Variable	Unstandardized Coefficients		Standardized Coefficients	t	p
	B	Se	β		
(Constant)	4.598	2.692		1.708	0.091
Age	-0.223	0.045	-0.519	-4.926	<0.001**
Gender	-0.856	0.865	-0.075	-0.990	0.325
Preoperative Framingham score	0.532	0.068	0.848	7.835	<0.001**
Change in HDL cholesterol (mg/dL)	0.279	0.139	0.154	2.006	0.047*
Change in systolic blood pressure (mmHg)	-0.073	0.046	-0.119	-1.575	0.118

Table 4a: BMI, body mass index; HDL, high-density lipoprotein; r, Spearman's correlation coefficient. **p<0.05; **p<0.01. Mann-Whitney U test was used for antihypertensive discontinuation comparison.
Table 4b: HDL, high-density lipoprotein; SE, standard error. **p<0.05; **p<0.01. Model summary: R^2=0.413, Adjusted R^2=0.385, F(5.105)=14.769, p<0.001.

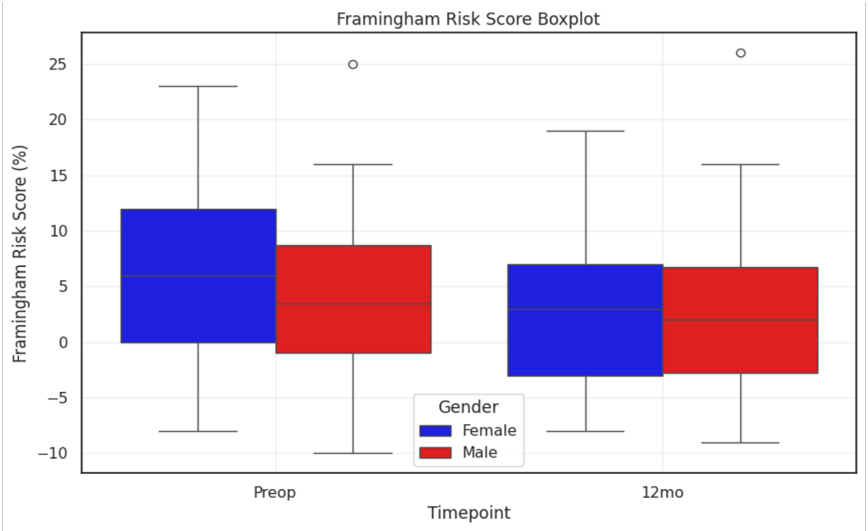

Figure 1. Clustered boxplot illustrating Framingham 10-year cardiovascular risk scores at the preoperative and 12-month postoperative timepoints, stratified by gender (female = blue; male= red). Boxes represent medians and interquartile ranges; whiskers indicate the full range; circles denote outliers.

2, BMI reduction was not significantly correlated with cardiovascular risk reduction. Multivariable linear regression analysis (R^2 = 0.413, Adjusted R^2 = 0.385, F = 14.769, p<0.001) identified independent predictors of absolute risk reduction:

- Age (β= - 0.223, p<0.001) - older age associated with less risk reduction.
- Baseline FRS (β = 0.532, p<0.001) - higher baseline risk associated with greater absolute reduction.
- Change in HDL cholesterol (β = 0.279, p = 0.047).

Gender (β = -0.856, p = 0.325) and change in systolic blood pressure (β = -0.073, p = 0.118) were not significant predictors in the multivariable model.

Smoking cessation

Smoking prevalence decreased from 47.7% (n=53) at baseline to 30.6% (n=34) at 12 months (p<0.001).

https://doi.org/10.5455/annalsmedres.2025.08.239

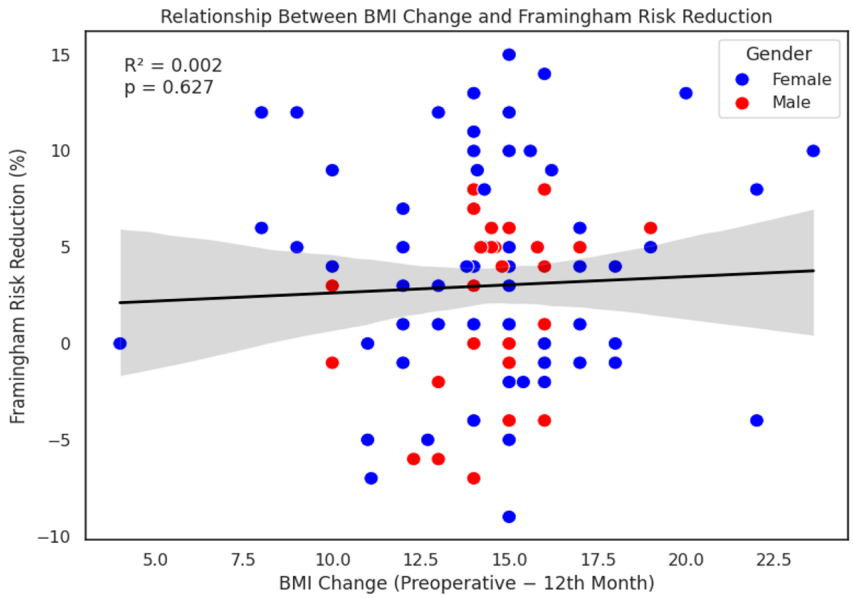

Figure 2. Scatter plot showing the relationship between BMI change and absolute cardiovascular risk reduction. Each point represents an individual patient (female = blue; male = red). The solid line represents the linear regression model with its 95% confidence band.

DISCUSSION

In this retrospective cohort study, the findings from a one-year follow-up of 111 patients who underwent laparoscopic sleeve gastrectomy highlight the need to re-evaluate the cardiovascular effects of bariatric surgery. Although the 50% reduction in the Framingham cardiovascular risk score is clinically significant, one of the most striking observations is that this improvement did not correlate with the 31.8% reduction in BMI ($r = 0.032$, $p = 0.735$). This suggests that the reduction in cardiovascular risk may occur independently of the magnitude of weight loss. Moreover, multivariable analysis identified age, baseline Framingham score, and changes in HDL cholesterol as independent predictors, indicating that the cardiovascular benefits of sleeve gastrectomy are mediated through multifactorial mechanisms rather than weight loss alone.

The lack of correlation between BMI change and cardiovascular risk reduction is consistent with recent literature emphasizing the complexity of the obesity–cardiovascular disease relationship. The joint position statement of the World Obesity Federation and the World Heart Federation underscores that this link is mediated by metabolic, endocrinological, immunological, structural, humoral, and hemodynamic mechanisms [8]. In their large cohort, Katsoulis et al. also demonstrated that the effect of weight loss on cardiovascular outcomes is nonlinear and varies by BMI category [9]. Our findings, showing risk reduction despite the absence of correlation with BMI change, support the notion that sleeve gastrectomy confers benefits beyond adipose tissue loss.

The biphasic HDL cholesterol pattern observed in our study has been previously described in the bariatric literature [10-12]. HDL typically shows an early postoperative decline, followed by recovery to levels above preoperative levels. In our series, HDL levels decreased to 44 mg/dL at 3 months, then increased to 50 mg/dL at 12 months. Interestingly, univariable analysis showed a negative correlation between HDL increase and risk reduction ($r = -0.235$, $p = 0.013$), this association became positive in the multivariable model ($\beta = 0.279$, $p = 0.047$).

This paradoxical finding, which may be termed the "HDL paradox", supports the hypothesis that bariatric surgery alters HDL functionality rather than simply its quantity [11,12].

Age was identified as a strong negative predictor in our study ($\beta = -0.223$, $p<0.001$), consistent with reports that metabolic plasticity decreases with age. Athanasiadis et al. reported that younger patients achieved greater weight loss and resolution of comorbidities [13], and Tabasi et al. similarly observed greater reductions in BMI in younger cohorts [14]. Our findings in a population with a median age of 33 years (range, 18–63) underscore the importance of early surgical intervention. Although older patients also experienced significant risk reduction, the greater adaptive capacity of younger patients underscores the importance of optimal timing for surgery.

Additionally, patients with higher baseline Framingham scores experienced greater absolute risk reduction ($\beta = 0.532$, $p<0.001$). This aligns with van Veldhuisen et al.'s meta-analysis, which reported a 41% reduction in cardiovascular mortality after bariatric surgery [15], and with El Ansari et al., who observed Framingham score reductions of up to 32% in high-risk patients [16]. These findings suggest that surgical indications should incorporate cardiovascular risk assessment alongside BMI.

In our cohort, no statistically significant sex differences were found in cardiovascular risk reduction after sleeve gastrectomy, although women showed a numerically greater improvement. This supports prior evidence that bariatric

https://doi.org/10.5455/annalsmedres.2025.08.239

surgery provides cardiometabolic benefits largely independent of sex [17-19]. Despite the underrepresentation of men in bariatric cohorts, accumulating evidence—including our results—indicates that short- and mid-term outcomes are comparable across sexes. Consistent with this, we observed significant improvements in systolic blood pressure, from 120 (80–170) to 110 (80–140) mmHg, in total cholesterol, from 191 (127–316) to 178 (124–337) mg/dL, as well as a reduction in smoking prevalence from 47.7% to 30.6%. Although smoking cessation was not an independent predictor in multivariable analysis, its potential long-term contribution to cardiovascular outcomes remains important.

Limitations

This study has several limitations. The retrospective design precludes causal inference, and the 12-month follow-up may not capture long-term cardiovascular events. Medication use and HbA1c levels, which could influence the Framingham score, were not systematically assessed. The relatively small proportion of men (27%) limited statistical power for sex comparisons, and lifestyle factors such as dietary compliance and physical activity were not measured. Despite these limitations, our findings provide valuable insights into the complex interplay among bariatric surgery, HDL dynamics, and cardiovascular risk.

CONCLUSION

Sleeve gastrectomy produced a significant reduction in cardiovascular risk, independent of the magnitude of weight loss. Change in BMI was not correlated with improvements in the Framingham Risk Score, indicating that weight loss alone does not confer cardiovascular benefit. The HDL paradox—negative in univariable analysis but positive in multivariable modeling—suggests that functional changes in HDL may be more important than absolute concentrations in mediating cardiovascular protection after bariatric surgery.

Ethics Committee Approval: The study protocol was approved by the Ethics Committee of Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital (Decision No: 2025/263, Date: 18.07.2025).

Informed Consent: Due to the retrospective nature of this study and the use of de-identified data extracted from medical records, the requirement for informed consent was waived by the ethics committee. Patient anonymity was carefully protected throughout the study; all personal identifiers were removed from the dataset, and data were analyzed using coded identification numbers.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Author Contributions: Conception and Design: A.G., I.A.; Supervision: I.A.; Materials: None; Data Collection and/or Processing: A.G., S.A., M.K.B., G.Ö.; Analysis and/or Interpretation: A.G., S.A., M.K.B., G.Ö.; Literature Review: S.A., M.K.B., G.Ö.; Writing: A.G., S.A.; Critical Review: A.G., I.A.

Financial Disclosure: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Artificial Intelligence Disclosure: During the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-4) solely for the purpose of English language editing, including improving grammatical accuracy, sentence structure, and overall readability of the text. This tool was not involved in the study design, data collection, data analysis, interpretation of results, or the formulation of scientific conclusions. All AI-assisted edits were critically reviewed, verified, and approved by the authors. The authors take full responsibility for the content of this publication.

REFERENCES

1. Lee YC, Christensen JJ, Parnell LD, Smith CE, Shao J, McKeown NM, et al. Using Machine Learning to Predict Obesity Based on Genome-Wide and Epigenome-Wide Gene-Gene and Gene-Diet Interactions. *Front Genet.* 2021;12:783845. doi: 10.3389/fgene.2021.783845.
2. Tutor AW, Lavie CJ, Kachur S, Milani RV, Ventura HO. Updates on obesity and the obesity paradox in cardiovascular diseases. *Prog Cardiovasc Dis.* 2023;78:2-10. doi: 10.1016/j.pcad.2022.11.013.
3. Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, Lee A, et al. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med.* 2017;377(1):13-27. doi: 10.1056/NEJMoa1614362.
4. Jamaly S, Carlsson L, Peltonen M, Jacobson P, Karason K. Surgical obesity treatment and the risk of heart failure. *Eur Heart J.* 2019;40(26):2131-8. doi: 10.1093/eurheartj/ehz295.
5. Kenchaiah S, Evans JC, Levy D, Wilson PW, Benjamin EJ, Larson MG, et al. Obesity and the risk of heart failure. *N Engl J Med.* 2002;347(5):305-13. doi: 10.1056/NEJMoa020245.
6. Kıvrak U, Karahan M, Ertuğrul İ, Boğazlıyan ÖF, Tuna MK, Şenol Y, et al. Evaluation of changes in posterior segment parameters and cardiovascular risk score following laparoscopic sleeve gastrectomy in obese patients. *Laparosc Endosc Surg Sci.* 2024;31(4):205-215. doi: 10.14744/less.2024.19794.
7. D'Agostino Sr RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation.* 2008;117(6):743-53. doi: 10.1161/CIRCULATIONAHA.107.699579.
8. Lopez-Jimenez F, Almahmeed W, Bays H, Cuevas A, Di Angelantonio E, le Roux CW, et al. Obesity and cardiovascular disease: mechanistic insights and management strategies. A joint position paper by the World Heart Federation and World Obesity Federation. *Eur J Prev Cardiol.* 2022;29(17):2218-37. doi: 10.1093/eurjpc/zwac187.
9. Katsoulis M, Stavola BD, Diaz-Ordaz K, Gomes M, Lai A, Lagiou P, et al. Weight Change and the Onset of Cardiovascular Diseases: Emulating Trials Using Electronic Health Records. *Epidemiology.* 2021;32(5):744-55. doi: 10.1097/EDE.0000000000001393.
10. Asztalos BF, Swarbrick MM, Schaefer EJ, Dallal GE, Horvath KV, Ai M, et al. Effects of weight loss, induced by gastric bypass surgery, on HDL remodeling in obese women. *J Lipid Res.* 2010;51(8):2405-12. doi: 10.1194/jlr.P900015.

https://doi.org/10.5455/annalsmedres.2025.08.239

11. Aminian A, Zelisko A, Kirwan JP, Brethauer SA, Schauer PR. Exploring the impact of bariatric surgery on high density lipoprotein. *Surg Obes Relat Dis.* 2015;11(1):238-47. doi: 10.1016/j.soard.2014.07.017.
12. Piché M-E, Tardif I, Auclair A, Poirier P. Effects of bariatric surgery on lipid-lipoprotein profile. *Metabolism.* 2021;115:154441. doi: 10.1016/j.metabol.2020.154441.
13. Athanasiadis DI, Hernandez E, Monfared S, Kubicki N, Ninad N, Karim A, et al. Bariatric surgery outcomes: is age just a number? *Surg Endosc.* 2021;35(6):3139-46. doi: 10.1007/s00464-020-07752-9.
14. Tabasi M, Farsimadan M, Yazdannasab M, Elyasinia F, Siadat SD, Soroush A. One-year outcomes of laparoscopic sleeve gastrectomy in morbidly obese patients regarding the age, gender, and postoperative follow-ups. *Adv Dig Med.* 2022;10(3):159-70. doi: 10.1002/aid2.13341.
15. Van Veldhuisen SL, Gorter TM, van Woerden G, de Boer RA, Rienstra M, Hazebroek EJ, et al. Bariatric surgery and cardiovascular disease: a systematic review and meta-analysis. *Eur Heart J.* 2022;43(20):1955-69. doi: 10.1093/eurheartj/ehac071.
16. El Ansari W, Lock M, Elgenaied I, Elhag W. From Outcomes to Long-Term Impact: Do Improvements in Individual Cardiometabolic Risk Factors After Sleeve Gastrectomy Translate to Reductions in Composite Indicators of Cardiovascular Risk? *Obes Surg.* 2025;35(8):2926-2939. doi: 10.1007/s11695-025-07942-1.
17. Hider AM, Bonham A, Carlin A, Finks J, Ghaferi A, Varban O, et al. Association of sex differences on weight loss and complications following bariatric surgery. *J Surg Res.* 2024;299:359-65. doi: 10.1016/j.jss.2024.04.050.
18. Mousapour P, Tasdighi E, Khalaj A, Mahdavi M, Valizadeh M, Taheri H, et al. Sex disparity in laparoscopic bariatric surgery outcomes: a matched-pair cohort analysis. *Sci Rep.* 2021;11(1):12809. doi: 10.1038/s41598-021-92254-4.
19. Risi R, Rossini G, Tozzi R, Pieralice S, Monte L, Masi D, et al. Sex difference in the safety and efficacy of bariatric procedures: a systematic review and meta-analysis. *Surg Obes Relat Dis.* 2022;18(7):983-96. doi: 10.1016/j.soard.2022.03.022.

https://doi.org/10.5455/annalsmedres.2025.08.239

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Baseline detrusor pressure and renal reserve as determinants of botulinum toxin a response in pediatric neurogenic bladder

Derya Yayla

Izmir City Hospital, Clinic of Pediatric Urology, Izmir, Türkiye

**Corresponding author: dryayla@yahoo.com (Derya Yayla)*

MAIN POINTS

- Intradetrusor botulinum toxin A (BTX-A) is a minimally invasive option for pediatric patients with medically refractory neurogenic bladder and can reduce storage-phase detrusor pressure.
- Achieving safe storage pressures (30–35 cmH_2O) is essential for upper urinary tract preservation, yet the response to BTX-A is heterogeneous.
- At 6 months after intradetrusor BTX-A, the reduction in pressure was modest, and only a minority of patients reached safe storage thresholds.
- Higher baseline p_det and preserved renal function (eGFR) were associated with a clinically meaningful response, defined as a decrease of $\geq$10 cmH_2O.
- Baseline p_det demonstrated fair discrimination in predicting response (AUC 0.72), which may support pressure-guided candidate selection (external validation is required).

Cite this article as: Yayla D. Baseline detrusor pressure and renal reserve as determinants of botulinum toxin a response in pediatric neurogenic bladder. *Ann Med Res*. 2026;33(5):222--229. doi: 10.5455/annalsmedres.2025.11.336.

ABSTRACT

Aim: Pediatric neurogenic bladder (NB) represents a challenging condition that jeopardizes both lower and upper urinary tract integrity. This study aimed to evaluate the pressure-lowering efficacy of intradetrusor botulinum toxin A (BTX-A) in children with spinal-origin NB and to identify clinical and urodynamic predictors of treatment response.

Materials and Methods: In this retrospective cohort study, 38 pediatric patients with spinal-origin NB who received 100 IU of BTX-A between October 2023 and December 2024 were analyzed. BTX-A was injected cystoscopically at 20 detrusor sites, sparing the trigone. The primary outcome was the 6-month change in storage-phase detrusor pressure (Δp_det), calculated as p_det = p_ves − p_abd per ICCS standards. Secondary outcomes included an exploratory responder endpoint ($\geq$10 cmH_2O reduction in storage pressure) and attainment of guideline-based safe storage thresholds (<30 and <35 cmH_2O). Associations with demographic, functional, and renal parameters were examined using correlation, multivariable regression, and ROC analyses.

Results: Mean baseline storage-phase detrusor pressure was 59.9 ± 21.2 cmH_2O, decreasing to 52.5 ± 19.7 cmH_2O at six months (Δp_det = −7.4 ± 12.5 cmH_2O; $P < 0.05$). A $\geq$10 cmH_2O reduction in pressure was observed in 26.3% (10/38) of patients, and 10.5% and 13.2% achieved storage pressures <30 cmH_2O and <35 cmH_2O, respectively. In exploratory logistic regression, higher baseline p_det (OR = 1.05; P = 0.039) and preserved renal function (eGFR; OR = 1.04; P = 0.042) were associated with a $\geq$10 cmH_2O reduction, whereas etiology and age showed non-significant trends.

Conclusion: Intradetrusor BTX-A produced a modest but statistically significant reduction in storage-phase detrusor pressure, with only a minority of patients reaching guideline-defined safe storage thresholds. These findings support BTX-A as a clinically relevant pressure-lowering option within a multimodal, pressure-guided strategy --- particularly in children with high baseline pressures and preserved renal reserve --- rather than as a uniformly transformative stand-alone therapy.

Keywords: Pediatric neurogenic bladder, Botulinum toxin A, Detrusor pressure, Renal function, Spinal dysraphism

Received: Nov 13, 2025 **Accepted:** Dec 22, 2025 **Available Online:** May 22, 2026

INTRODUCTION

Neurogenic bladder (NB) in children is a chronic disorder that simultaneously threatens lower urinary tract function and upper urinary tract integrity, often leading to long-term morbidity when inadequately controlled. Contemporary guidelines highlight the central role of urodynamic assessment for diagnosis, risk stratification, and monitoring, recommending standardized terminology and methodology established by the *International Children's Continence Society (ICCS)* [1–3]. The principal therapeutic objective is to maintain safe storage pressures (<30–35 cmH_2O) and adequate bladder compliance to preserve renal function [1,4]. In this study, the term "spinal-origin neurogenic bladder" refers to neurogenic bladder dysfunction secondary to surgically treated myelomeningocele or other spinal dysraphism–related or spinal neurogenic conditions (including tethered cord, di-

https://doi.org/10.5455/annalsmedres.2025.11.336

astematomyelia, and traumatic spinal lesions).

Conservative management—including clean intermittent catheterization (CIC), antimuscarinic agents, and behavioral therapy—remains first-line treatment. Yet approximately one-third of patients remain refractory, exhibiting persistently high storage pressures or poor compliance despite optimal conservative measures [1]. For these cases, intradetrusor botulinum toxin A (BTX-A) has emerged as a minimally invasive, reversible, and repeatable therapeutic option. By inhibiting acetylcholine release at the neuromuscular junction, BTX-A decreases detrusor overactivity and storage pressure, and improves bladder compliance and capacity [5,6].

Clinical studies and meta-analyses consistently demonstrate significant reductions in Δp_det and improvements in bladder compliance and capacity within 3–6 months after BTX-A, with excellent safety and tolerability [5–8]. Repeated injections have been reported to maintain efficacy without cumulative adverse effects [9,10].

Despite these encouraging outcomes, heterogeneity persists in injection protocols, dosing regimens, concomitant therapies, and follow-up durations. Recent systematic reviews reaffirm overall efficacy but emphasize the importance of identifying predictors of response, particularly the effects of elevated baseline storage-phase pressure and preserved renal function on treatment success [7,11,12].

Accordingly, the present study aimed to evaluate the six-month effects of intradetrusor BTX-A on Δp_det in children with NB secondary to spinal pathology, to determine clinically meaningful response rates (≥10 cmH_2O reduction and/or attainment of safe storage thresholds), and to explore baseline urodynamic predictors of outcome.

MATERIALS AND METHODS

Study design and setting

This retrospective cohort study included consecutive pediatric patients diagnosed with neurogenic bladder (NB) who underwent intradetrusor botulinum toxin A (BTX-A) injections at a tertiary pediatric urology center between October 2023 and December 2024. A total of 38 children were included (mean age: 7.8 ± 4.0 years); 20 were female (52.6%) and 18 were male (47.4%). The etiology of NB was operated myelomeningocele (MMC) in 25 patients (65.8%) and other spinal pathologies (diastematomyelia, tethered cord syndrome, or traumatic spinal lesions) in 13 patients (34.2%). Functionally, 17 children (44.7%) were ambulatory, and 21 (55.3%) were non-ambulatory. The mean estimated glomerular filtration rate (eGFR) was 118.0 ± 33.2 mL/min/1.73 m^2, and the mean age at initiation of clean intermittent catheterization (CIC) was 5.1 ± 4.0 years (see Table 1 for demographic characteristics).

Inclusion and Exclusion criteria

Inclusion criteria

- NB secondary to operated MMC or other spinal pathologies (diastematomyelia, tethered cord syndrome, or spinal cord injury) with CIC for ≥ 6 months;
- Availability of complete pre-procedure and 6-month follow-up clinical and urodynamic data.

Exclusion criteria

History of major urological reconstructive surgery (e.g., bladder augmentation or ureteral reimplantation) and/or missing baseline data.

Intervention (BTX-A protocol)

All procedures were performed under general anesthesia by the same pediatric urology team. A total of 100 IU of onabotulinumtoxinA (Botox®, Allergan Inc., Irvine, CA, USA) diluted in 10 mL of normal saline was injected intradetrusorally under cystoscopic guidance at 20 evenly distributed sites over the bladder dome, anterior, and lateral walls, sparing the trigone. Injection depth was standardized to 2–3 mm using a 4-mm injection needle. No intraoperative or postoperative complications (e.g., hematuria, fever, or urinary retention) were observed. All patients continued their baseline CIC and antimuscarinic regimens as clinically indicated.

All patients received a fixed dose of 100 IU of onabotulinumtoxinA, diluted in 10 mL of normal saline, and injected into 20 sites. This fixed-dose regimen reflects our institutional protocol, which is largely shaped by national reimbursement constraints: botulinum toxin A for the treatment of pediatric neurogenic bladder is not reimbursed by the social security system; therefore, families must cover the full cost. To limit the financial burden and to allow single-vial use across a broad weight range (10–35 kg), individual dose titration based on weight or capacity was not routinely feasible.

Variables and definitions

Urodynamic variables: Δp_det (cmH_2O) measured before and 6 months after BTX-A injection, cystometric bladder capacity, and bladder volume at 40 cmH_2O.

Detrusor pressure (p_det): Intravesical pressure (p_ves) and abdominal pressure (p_abd) were recorded simultaneously during urodynamic testing. Storage-phase detrusor pressure (p_det) was calculated, in accordance with ICCS standards, as the difference between the two measures (p_det = p_ves − p_abd). Throughout the study, the parameter previously referred to as "storage-phase detrusor pressure" represents storage-phase detrusor pressure (p_det). In pediatric NB, persistent storage-phase p_det > 40 cmH_2O is associated with vesicoureteral reflux, renal scarring, and progressive loss of renal function; accordingly, storage-phase p_det was selected as the primary urodynamic endpoint for evaluating bladder dynamics and treatment outcomes.

https://doi.org/10.5455/annalsmedres.2025.11.336

Clinical/cohort variables: age, sex, ambulation status, etiology (operated MMC vs. other spinal pathology), and renal function (eGFR).

Renal function: eGFR was calculated using the updated Schwartz formula, adjusted for age, sex, and height, expressed in mL/min/1.73 m^2. Values below age-specific reference ranges were considered indicative of renal impairment.

Treatment-related variables: use and type of antimuscarinic medication, duration of CIC, and age at initiation of CIC.

Outcomes

Primary outcome: the change in storage-phase detrusor pressure (Δp_det = p_det,6-month − p_det,baseline).

Secondary outcomes: Secondary outcomes included (i) a reduction in storage pressure of ≥10 cmH_2O and (ii) attainment of guideline-based safe storage thresholds (<30 and <35 cmH_2O). The ≥10 cmH_2O cut-off is not a universally standardized value in the literature; rather, it was chosen a priori as a pragmatic, exploratory threshold to facilitate categorical interpretation of treatment response alongside the more clinically relevant pressure.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics version 29.0 (IBM Corp., Armonk, NY, USA). The normality of continuous variables was assessed using the Shapiro–Wilk test. Paired comparisons between pre- and post-treatment values were performed using the paired t-test or Wilcoxon signed-rank test, as appropriate. Continuous variables were reported as means ± standard deviations (SDs), 95% confidence intervals (CIs), and effect sizes (Cohen's d_z). Associations between Δp_det and continuous covariates were evaluated with Pearson or Spearman correlation coefficients. Group comparisons were conducted using the Mann–Whitney U test for two-group comparisons and the Kruskal–Wallis test for multi-group analyses. Multivariable linear regression was used to identify independent predictors of Δp_det, whereas binary logistic regression was applied to determine predictors of a clinically significant reduction in pressure (≥10 cmH_2O). Receiver Operating Characteristic (ROC) curve analysis was used to determine the optimal baseline pressure cutoff for predicting a favorable response. A two-sided $P < 0.05$ was considered statistically significant. Because the continuous outcome (Δp_det) and the binary responder outcome (≥10 cmH_2O reduction) capture different aspects of treatment response, predictors may show divergent associations across models. Given the limited number of events (10 responders), logistic regression was prespecified as an exploratory analysis. The events-per-variable (EPV) ratio was below the recommended threshold; therefore, the logistic model may be unstable and prone to overfitting. These limitations were acknowledged when interpreting the results.

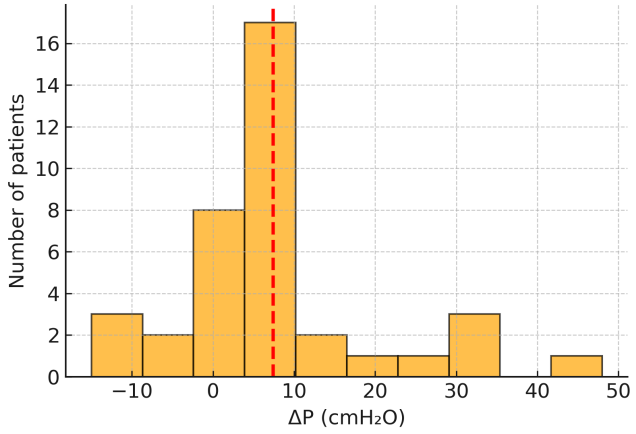

Figure 1. Distribution of storage-phase detrusor pressure change (Δp_det) following BTX-A injection. Most patients experienced a reduction in pressure, while a minority showed paradoxical increases.

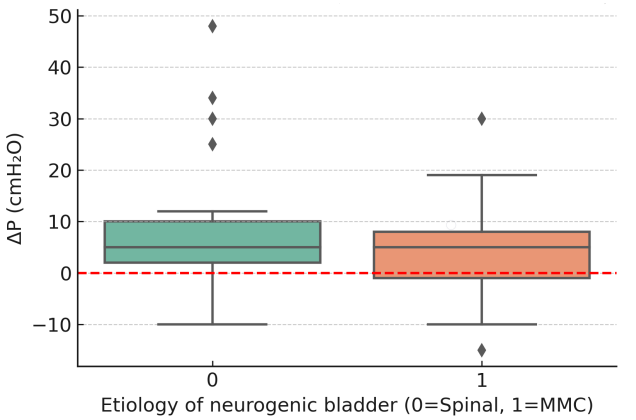

Figure 2. Detrusor pressure change (Δp_det) according to neurogenic bladder etiology (spinal vs. MMC). A greater reduction was observed in the spinal group, but the difference was not statistically significant (P > .05).

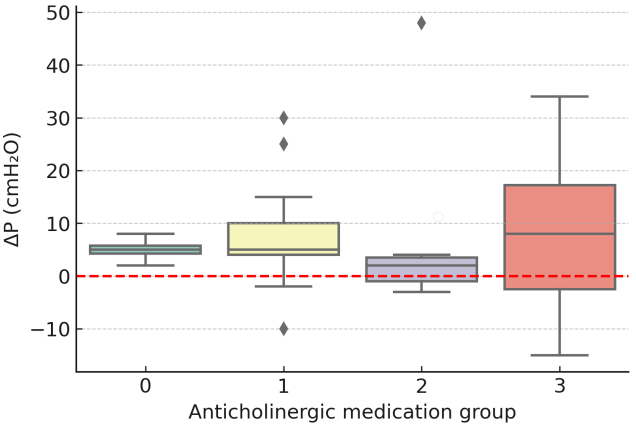

Figure 3. Detrusor pressure change (Δp_det) stratified by anticholinergic medication (0 = None, 1 = Oxybutynin, 2 = Tolterodine, 3 = Propiverine). No significant differences were detected between the groups.

https://doi.org/10.5455/annalsmedres.2025.11.336

Table 1. Patient demographics and clinical characteristics.

Parameter	Value
Age (years)	7.8 ± 4.0
Sex	Female: 20 (52.6%), Male: 18 (47.4%)
Etiology of neurogenic bladder	Operated MMC: 25 (65.8%), Other spinal pathology: 13 (34.2%)
Ambulatory status	Ambulatory: 17 (44.7%), Non-ambulatory: 21 (55.3%)
eGFR (mL/min/1.73m²)	118.0 ± 33.2
Age at initiation of CIC (years)	5.1 ± 4.0

Values are presented as mean ± standard deviation (SD) or number (percentage). Abbreviations: MMC, myelomeningocele; eGFR, estimated glomerular filtration rate; CIC, clean intermittent catheterization.

Table 2. Urodynamic findings.

Parameter	Value
Baseline intravesical pressure (cmH_2O)	59.9 ± 21.2
Post-BTX-A intravesical pressure (cmH_2O)	52.5 ± 19.7
Δ Pressure (cmH_2O)	−7.4 ± 12.5
Patients with ≥10 cmH_2O reduction (%)	10/38 (26.3%)
Patients achieving storage pressure <30 cmH_2O (%)	4/38 (10.5%)
Patients achieving storage pressure <35 cmH_2O (%)	5/38 (13.2%)

Values are presented as mean ± standard deviation (SD) or number (percentage). Abbreviations: BTX-A, botulinum toxin A; cmH_2O, centimeters of water.

Table 3. Multivariable linear regression analysis (predictors of ΔP).

Variable	β coefficient	95% CI	P value
Baseline pressure	−0.43	−0.69 → −0.17	.002
Bladder volume at 40 cmH_2O	+5.59	−0.80 → +11.98	.084
Other covariates	NS	--	>0.1

β = standardized regression coefficient; CI = confidence interval; NS = not significant. Abbreviations: ΔP, change in intravesical pressure; cmH_2O, centimeters of water.

Table 4. Univariable correlation analyses.

Variable	Pearson r	P value
Baseline pressure	−0.41	.011
CIC duration	−0.03	.84
Age at CIC initiation	−0.18	.28
Bladder capacity	−0.08	.63
Bladder volume at 40 cmH_2O	−0.06	.72
eGFR	+0.05	.75
Age	+0.03	.86

Pearson's correlation coefficients between intravesical pressure change (ΔP) and clinical variables. Abbreviations: CIC, clean intermittent catheterization; eGFR, estimated glomerular filtration rate; cmH_2O, centimeters of water.

Table 5. Logistic regression analysis (predictors of ≥10 cmH_2O reduction).

Variable	OR	95% CI	
Baseline pressure	1.05	1.00–1.10	.039
Etiology (MMC)	0.29	0.04–2.16	.227
Antimuscarinic use	1.28	0.49–3.33	.617
Age	1.34	0.99–1.81	.055
eGFR	1.04	1.00–1.09	.042

OR = odds ratio; CI = confidence interval; P < .05 was considered statistically significant. Abbreviations: MMC, myelomeningocele; eGFR, estimated glomerular filtration rate.

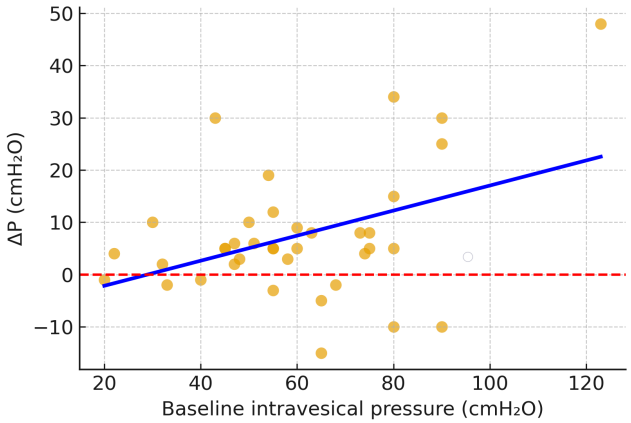

Figure 4. Correlation between baseline storage-phase detrusor pressure and pressure change (Δp_det) after BTX-A injection. Higher baseline pressures were associated with greater reductions (r = --0.41, P = .011).

RESULTS

Primary outcome

The mean baseline p_det was 59.9 ± 21.2 cmH_2O, decreasing to 52.5 ± 19.7 cmH_2O at 6 months, corresponding to an average change of Δp_det = −7.4 ± 12.5 cmH_2O (median −5; range −48 to +15 cmH_2O) (Tables 1 and 2). Most patients exhibited a post-treatment pressure reduction, although a small subset demonstrated minimal or paradoxical increases (Figure 1).

Secondary outcomes

A clinically significant response, defined as a ≥10 cmH_2O reduction in storage pressure, was observed in 26.3% (10/38) of patients, and only 10.5% and 13.2% achieved storage pressures <30 cmH_2O and <35 cmH_2O, respectively. Taken together,

https://doi.org/10.5455/annalsmedres.2025.11.336

these findings indicate a modest overall pressure-lowering effect of BTX-A in this cohort, with a relatively small proportion of patients reaching guideline-recommended safe storage thresholds.

Subgroup and correlation analyses

Etiology: The mean pressure reduction was greater in patients with non-MMC spinal pathologies (−8.5 cmH_2O) than in those with myelomeningocele (MMC) (−5.3 cmH_2O); however, this difference was not statistically significant (P > .05) (Figure 2).

Under antimuscarinic therapy, Δp_det did not differ by drug type or treatment duration (P > .05; Figure 3). Although variable responses were observed within the propiverine subgroup, no clear superiority over other agents was demonstrated.

Functional status and CIC: Ambulatory children showed a larger mean reduction (−9.9 vs.−5.4 cmH_2O), which did not reach statistical significance (P = .19). Neither CIC duration nor age at initiation of CIC correlated with Δp_det (both P > .05).

Filling parameters: Δp_det was not significantly correlated with cystometric capacity (r = −0.08, P = .63) or bladder volume at 40 cmH_2O (r = −0.06, P = .72).

Renal function: No significant association was observed between baseline eGFR and Δp_det in univariable analysis (P > 0.05). Comprehensive univariable correlation results are summarized in Table 4.

Multivariable models

Linear regression (Δp_det as continuous outcome): Baseline storage-phase detrusor pressure emerged as an independent inverse predictor of Δp_det (β = −0.43, P = .002).

Bladder volume at 40 cmH_2O showed a borderline association (P = .084), while other covariates were not significant.

The model explained 36% of the variance (R^2 = 0.36; Table 3).

Logistic regression (≥ 10 cmH_2O reduction): Higher baseline storage-phase pressure (OR = 1.05; 95% CI 1.00–1.10; P = .039) and higher eGFR (OR = 1.04; 95% CI 1.00–1.09; P = .042) were independent predictors of a clinically significant response. MMC etiology and age showed nonsignificant trends (Table 5).

ROC analysis

To predict a ≥10 cmH_2O pressure reduction, baseline p_det yielded an AUC of 0.72, with an optimal cutoff of 58 cmH_2O (sensitivity 70%, specificity 65%) (Figure 4).

DISCUSSION

Overall efficacy

Intradetrusor botulinum toxin A (BTX-A) therapy produced a mean reduction of 7.4 cmH_2O in storage-phase detrusor pressure in children with neurogenic bladder (NB). This finding is consistent with previous systematic reviews and meta-analyses demonstrating significant short- and mid-term decreases in detrusor pressure after BTX-A injection [5–9, 12]. However, the magnitude of reduction in our cohort was smaller than the ~20–25 cmH_2O decreases reported in larger pediatric series. This difference may be attributed to our limited sample size, a high proportion of myelomeningocele (MMC) cases, and relatively low baseline pressures. Additional contributors likely include variations in injection technique, dose distribution, and concomitant antimuscarinic regimens [6, 9, 10].

Another factor that may have contributed to the modest and heterogeneous pressure response in our cohort is the use of a fixed 100-IU dose across a relatively wide weight range (10–35 kg). Due to national reimbursement restrictions and requirements for out-of-pocket payments, individualized weight- or capacity-based dose adjustments were not feasible in our setting. Prospective studies comparing fixed and weight-adjusted dosing protocols are warranted to better define the optimal BTX-A regimen in pediatric neurogenic bladder.

Although BTX-A significantly reduced storage-phase detrusor pressure at the group level, the magnitude of benefit in our series was modest, and only a minority of patients achieved either the predefined responder criterion (≥10 cmH_2O reduction) or the guideline-based safe storage threshold (<30–35 cmH_2O). In this context, our data support BTX-A as a clinically relevant but limited pressure-lowering option, rather than a universally transformative intervention for pediatric neurogenic bladder. The relatively low responder rate, compared with larger pediatric series, may reflect our small sample size, the predominance of myelomeningocele cases, relatively low baseline pressures, and the use of a fixed 100 IU dose across a broad weight range. Overall, these observations suggest that BTX-A provides a limited yet meaningful reduction in storage pressure in this high-risk cohort, and its role should be viewed as part of a multimodal, pressure-guided management strategy rather than as a definitive, stand-alone solution.

Response by etiology

Histopathological studies have shown that fibrotic remodeling of detrusor smooth muscle may attenuate BTX-A efficacy in MMC-related NB [11,12]. Although histological data were unavailable in our series, this explanation accords with prior evidence. In contrast, greater pressure reductions have been documented in patients with non-MMC spinal dysraphism or with traumatic etiologies [13]. Our cohort demonstrated a similar trend toward a greater reduction among non-MMC cases, although statistical significance was not achieved due to limited power. Multicenter studies have likewise indicated that underlying etiology influences BTX-A responsiveness [14,15].

A small subset of patients exhibited paradoxical increases in

https://doi.org/10.5455/annalsmedres.2025.11.336

storage pressure after BTX-A injection. Several mechanisms may account for this phenomenon. First, detrusor–sphincter dyssynergia may increase outlet resistance during bladder filling, thereby elevating storage-phase detrusor pressure despite detrusor chemodenervation. Second, transient inflammatory changes or mucosal edema following injection may temporarily augment storage pressures. Third, in patients with spinal dysraphism, autonomic dysreflexia–related hypertensive bladder responses may contribute to abrupt physiologic pressure surges. Finally, insufficient intradetrusor diffusion of BTX-A in severely fibrotic or poorly compliant bladders may limit therapeutic effect and even provoke compensatory overactivity. These mechanisms may collectively explain the paradoxical findings observed in this cohort and should be considered when interpreting individualized treatment response.

Clinically meaningful outcomes and pressure targets

Maintaining safe storage pressures (<30–35 cmH_2O) remains the principal determinant of upper urinary tract preservation [1,4]. In our series, the rate of clinically meaningful response—defined as a ≥10 cmH_2O reduction—was 26.3%. This is slightly lower than the 30–50% response rates reported in prior meta-analyses [4–9, 12], likely reflecting the small sample size, MMC predominance, and relatively low baseline pressures. Differences in injection techniques and adjunctive medications could also play a role.

Multivariable analysis identified higher baseline p_det as an independent predictor of greater pressure reduction, thereby supporting earlier BTX-A intervention in high-risk patients [10,11,14].

Our ROC analysis confirmed that a baseline pressure ≥58 cmH_2O predicted a clinically meaningful (≥10 cmH_2O) response with an AUC of 0.72, a sensitivity of 70%, and a specificity of 65%. This threshold may serve as a pragmatic marker for selecting optimal BTX-A candidates within the pediatric NB population with elevated baseline pressures.

Although Δp_det was analyzed as a continuous variable, we also reported a ≥10 cmH_2O reduction as a clinically interpretable categorical endpoint. Nevertheless, aligned with EAU/ESPU and ICCS guidelines, the primary therapeutic goal remains to achieve safe storage pressures (<30–35 cmH_2O) (1–4). Attaining this threshold has been associated with reduced renal scarring and better preservation of glomerular filtration rate (eGFR) [15]. In our cohort, no renal deterioration was observed among children who achieved safe storage levels, further underscoring the clinical importance of pressure-targeted management.

Renal function

Although univariable analysis revealed no direct correlation between eGFR and Δp_det, multivariable logistic regression identified preserved renal function as an independent predictor of a clinically meaningful response. eGFR was consistently calculated using the Schwartz formula, thereby ensuring methodological reliability. These findings suggest that preserved renal reserve may enhance responsiveness to BTX-A. Consistent with this, Gillon et al. demonstrated that BTX-A helped maintain renal function in multicenter pediatric cohorts [15]. Hence, renal function may represent not only an outcome but also a determinant of treatment success.

Although baseline eGFR did not correlate with Δp_det in univariable analysis, it emerged as a significant predictor in the exploratory logistic regression model. This discrepancy likely reflects the analytical difference between modeling a continuous change (Δp_det) versus a binary responder outcome (≥10 cmH_2O reduction). Additionally, the logistic model was constrained by a low events-per-variable ratio (10 events per variable), increasing the risk of instability and overfitting. Therefore, the association between preserved renal function and categorical response should be interpreted cautiously and considered hypothesis-generating rather than confirmatory.

Antimuscarinic therapy and combination regimens

We found no significant association between the type or duration of antimuscarinic therapy and Δp_det, consistent with previous pediatric data showing no superiority among agents [16]. Current EAU/ESPU guidelines emphasize tolerability and side-effect profiles rather than efficacy differences when selecting antimuscarinics [4]. Thus, BTX-A responsiveness appears largely independent of prior or concurrent antimuscarinic use.

Comparison with existing literature

Large observational series [4–6, 9–13] and meta-analyses [7–9, 12,14] provide robust evidence supporting the efficacy of BTX-A in pediatric NB. However, long-term renal outcomes and repeat-injection protocols remain underinvestigated. Zhou et al. [14] reported that repeated BTX-A injections can sustain a therapeutic benefit, albeit with diminishing magnitude over time. Other multicenter studies have shown that early BTX-A use in high-risk patients may aid upper tract preservation [15]. Collectively, these findings emphasize the importance of etiology, baseline pressure, and intervention timing in determining both short- and long-term outcomes.

General interpretation

Taken together, our results confirm that BTX-A significantly and safely reduces storage-phase detrusor pressure in pediatric NB. The magnitude of benefit varies according to baseline pressure, renal reserve, and underlying etiology, with patients with non-MMC spinal dysraphism showing a trend toward a greater response. These observations underscore the relevance of risk-stratified, pressure-targeted, and individualized management strategies, which constitute the cornerstone of contemporary pediatric neuro-urological care [1,10,15].

https://doi.org/10.5455/annalsmedres.2025.11.336

Clinical implications

BTX-A is a safe and effective option for reducing storage-phase detrusor pressure in pediatric patients with neurogenic bladder.

Children presenting with higher baseline storage pressures and preserved renal function are more likely to achieve clinically meaningful improvement, highlighting the importance of these parameters for treatment timing and patient selection [10–15].

The observed trend toward greater responsiveness in non-MMC spinal dysraphism than in MMC suggests that etiology-specific management strategies may optimize therapeutic outcomes [11–14].

Incorporating BTX-A into the pediatric neurogenic bladder treatment algorithm can help restore safe storage pressures, preserve upper urinary tract integrity, and delay or even prevent the need for reconstructive surgery [1,4,15].

These findings reinforce the importance of a pressure-targeted, risk-stratified, and individualized approach in accordance with ICCS and EAU/ESPU guidelines and emphasize the role of standardized urodynamic assessment and early identification of high-pressure bladders to improve long-term renal and functional outcomes [1–4].

In the present study, we additionally reported the proportion of patients achieving a reduction of $\geq$10 cmH_2O as an exploratory categorical endpoint. This cut-off was not derived from formal consensus or guideline recommendations, but was selected as a pragmatic threshold to reflect a clearly perceptible change in storage pressure and to allow comparison with previous pediatric BTX-A series that reported similar magnitudes of decrease. Nevertheless, in line with ICCS and EAU/ESPU guidelines, our primary clinical interpretation remains focused on achieving safe storage pressures (<30–35 cmH_2O), which are more directly linked to upper urinary tract protection.

Limitations

This study has several limitations that warrant consideration. The retrospective, single-center design introduces inherent risks of selection and information bias, limiting causal inference. The sample size (n = 38) restricts the statistical power of subgroup analyses, particularly regarding etiology and renal outcomes. The follow-up duration (6 months) captures short-term efficacy but may not reflect the sustained effects observed in longer-term studies [4–9, 14]. Histopathological data were unavailable to clarify biological mechanisms underlying differential responses between MMC and non-MMC etiologies [11–13]. Lastly, although all procedures followed standardized ICCS criteria, variability in injection distribution and concurrent medical regimens could have introduced confounding effects. Future multicenter prospective studies with extended follow-up, larger cohorts, and additional urodynamic parameters (compliance, capacity, electromyography, lesion level) are needed to refine predictive models and establish evidence-based algorithms for BTX-A use in pediatric NB [3,7,10,14].

CONCLUSION

Intradetrusor BTX-A significantly reduced the storage-phase detrusor pressure over six months in children with neurogenic bladder ($\Delta p\_det = -7.4 \pm 12.5$ cmH_2O). Treatment response was independently associated with baseline p_det and renal reserve; etiology exhibited a nonsignificant trend favoring spinal dysraphism over MMC. None of the following correlated with outcomes: a history of antimuscarinic use, clean intermittent catheterization parameters, or cystometric capacity. These findings reinforce the role of BTX-A as a cornerstone in pressure-guided, individualized, and risk-adapted management of pediatric NB [3–5, 7–10, 14–16]. By targeting high-pressure bladders before irreversible detrusor fibrosis and renal injury occur, BTX-A offers a minimally invasive yet effective strategy to protect upper urinary tract integrity and improve long-term urological prognosis.

Acknowledgments: The author would like to thank the pediatric urology nursing staff and research fellows for their valuable assistance in patient care and data collection throughout this study. Their contributions to maintaining consistent urodynamic data and clinical follow-up were essential to the completion of this research.

Ethics Committee Approval: This study was approved by the Ethics Committee of Izmir City Hospital (Decision No: 338/2025, Date: August 13, 2025).

Informed Consent: Due to the retrospective nature of the study, informed consent was waived. No animal or invasive experimental procedures were performed.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author declares that there are no financial or personal relationships that could be perceived as conflicts of interest influencing the research, authorship, or publication of this article.

Financial Disclosure: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The study was conducted using institutional resources without external financial support.

Artificial Intelligence Disclosure: During manuscript preparation, an artificial intelligence–based tool ChatGPT (OpenAI), specifically the GPT-5.1 was used exclusively for language editing and improvement of clarity and readability. The tool was not used for study design, data collection, statistical analysis, data interpretation, literature selection, reference generation, or development of scientific conclusions. All scientific content, analyses, and interpretations were independently developed by the author. The manuscript was thoroughly reviewed, revised, and approved in its entirety by the author, who assumes full responsibility for its accuracy, originality, and academic integrity. The use of AI assistance was limited strictly to linguistic refinement to enhance the quality of academic English expression.

https://doi.org/10.5455/annalsmedres.2025.11.336

REFERENCES

1. Stein R, Bogaert G, Doğan HS, et al. EAU/ESPU guidelines on the management of neurogenic bladder in children and adolescents—Part I: Diagnostics and conservative treatment. *Neurourol Urodyn.* 2020;39(1):45–57. doi: 10.1002/nau.24211.
2. Bauer SB, Nijman RJ, Drzewiecki BA, Sillén U, Hoebeke P. International Children's Continence Society standardization report on urodynamic studies of the lower urinary tract in children. *Neurourol Urodyn.* 2015;34(7):640–647. doi: 10.1002/nau.22783.
3. Austin PF, Bauer SB, Bower W, et al. The standardization of terminology of lower urinary tract function in children and adolescents: Update report from the ICCS. *Neurourol Urodyn.* 2016;35(4):471–481. doi: 10.1002/nau.22751.
4. EAU Guidelines on Paediatric Urology. Arnhem (NL): European Association of Urology Guidelines Office. 2023. ISBN: 978-94-92671-19-1.
5. Wu SY, Chang SJ, Yang SS, Hsu CK. Botulinum Toxin Injection for Medically Refractory Neurogenic Bladder in Children: A Systematic Review. *Toxins (Basel).* 2021;13(7):447. doi: 10.3390/toxins13070447.
6. Badawi JK. Botulinum toxin therapy in children with neurogenic detrusor overactivity. *Turk J Urol.* 2020;46(1):2–12. doi: 10.5152/tud.2019.19070.
7. Panunzio A, Orlando R, Mazzucato G, et al. Response to Treatment with Botulinum Neurotoxin A (BoNT-A) in Children and Adolescents with Neurogenic Lower Urinary Tract Dysfunction and Idiopathic Overactive Bladder: A Systematic Review and Meta-Analysis. *Toxins (Basel).* 2024;16(10):443. doi: 10.3390/toxins16100443.
8. Zulli A, Paglia G, Raffaele A, Mosiello G. Intravesical Botulin Toxin-A injections for neurogenic bladder dysfunction in children: A decade update. *Toxins (Basel).* 2024;16(8):339. doi: 10.3390/toxins16080339.
9. Gamé X, Mouracade P, Chartier-Kastler E, et al. Botulinum toxin-A (Botox) intradetrusor injections in children with neurogenic detrusor overactivity/neurogenic overactive bladder: a systematic literature review. *J Pediatr Urol.* 2009;5(3):156-64. doi: 10.1016/j.jpurol.2009.01.005.
10. Schulte-Baukloh H, Knispel HH, Stolze T, Weiss C, Michael T, Miller K. Repeated botulinum-A toxin injections in the treatment of children with neurogenic detrusor overactivity. *Urology.* 2005;66(4):865–870. doi: 10.1016/j.urology.2005.05.054.
11. Hascoet J, Peyronnet B, Forin V, et al. Intradetrusor Injections of Botulinum Toxin Type A in Children With Spina Bifida: A Multicenter Study. *Urology.* 2018;116:161-167. doi: 10.1016/10.1016/j.urology.2018.02.033
12. Giannantoni A, Mearini E, Del Zingaro M, Porena M. Six-year follow-up of botulinum toxin A intradetrusorial injections in patients with refractory neurogenic detrusor overactivity: Clinical and urodynamic results. *Eur Urol.* 2009;55(3):705–711. doi: 10.1016/j.eururo.2008.08.048.
13. Schulte-Baukloh H, Michael T, Stürzebecher B, Knispel HH. Botulinum-a toxin detrusor injection as a novel approach in the treatment of bladder spasticity in children with neurogenic bladder. *Eur Urol.* 2003;44(1):139-43. doi: 10.1016/s0302-2838(03)00136-2.
14. Franco I, Hoebeke PB, Dobremez E, et al. Long-term safety and tolerability of repeated treatments with onabotulinumtoxinA in children with neurogenic detrusor overactivity. *J Urol.* 2023;209(4):774–784. doi: 10.1097/JU.0000000000003157.
15. He G, Shen N, Feng S. Botulinum toxin A for the treatment of neurogenic bladder in children: A systematic review and meta-analysis. *ANZ J Surg.* 2025;95(6):1115–1121. doi: 10.1111/ans.19309.
16. Madersbacher H, Mürtz G, Alloussi S, et al. Propiverine vs oxybutynin for treating neurogenic detrusor overactivity in children and adolescents: results of a multicentre observational cohort study. *BJU Int.* 2009;103(6):776-81. doi: 10.1111/j.1464-410X.2008.08093.x.

https://doi.org/10.5455/annalsmedres.2025.11.336

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Transcatheter arterial embolization for refractory non-variceal gastrointestinal bleeding: Outcomes and prognostic indicators

Sinan Karatoprak [a,*], Mustafa Bilgili [a], Oguz Aslan [a], Gamze Turk [a], Nur Betul Karatoprak [a], Ferhat Omurca [b]

[a]Kayseri City Training and Research Hospital, Department of Radiology, Kayseri, Türkiye
[b]Kayseri City Training and Research Hospital, Department of Gastroenterology, Kayseri, Türkiye
*Corresponding author: snnkrtprk@hotmail.com (Sinan Karatoprak)

MAIN POINTS

- Transcatheter arterial embolization (TAE) achieved a high technical success rate (97.9%) and a low complication rate (6%) in patients with non-variceal gastrointestinal bleeding refractory to endoscopic therapy.
- Rebleeding occurred in 30% of patients and was significantly associated with comorbidity status, multi-vessel embolization, and negative endoscopic findings.
- In-hospital mortality was associated with higher Rockall scores and lower platelet counts, whereas overall mortality during follow-up was associated with elevated INR and higher Rockall scores.
- Length of hospital stay was independently associated with comorbidity burden, transfusion requirements, delayed embolization, and lower baseline hemoglobin.
- These findings underscore the prognostic importance of comorbidity and the role of preprocedural optimization in improving TAE outcomes for refractory gastrointestinal bleeding.

Cite this article as: Karatoprak S, Bilgili M, Aslan O, Turk G, Karatoprak NB, Omurca F. Transcatheter arterial embolization for refractory non-variceal gastrointestinal bleeding: Outcomes and prognostic indicators. *Ann Med Res*. 2026;33(5):230–237. doi: 10.5455/annalsmedres.2025.10.306.

ABSTRACT

Aim: Transcatheter arterial embolization (TAE) is increasingly used as second-line therapy for non-variceal gastrointestinal bleeding (NVGIB) when endoscopic hemostasis fails. This study investigated factors influencing rebleeding, mortality, and length of hospital stay (LOS) following TAE.

Materials and Methods: We retrospectively analyzed 50 patients (mean age 64.8 years; 72% male) who underwent TAE for acute NVGIB between April 2023 and April 2025. Demographic variables, comorbidities, laboratory values, endoscopic/angiographic findings, embolic materials, and outcomes were reviewed. Primary outcomes were rebleeding and in-hospital mortality; secondary outcomes were LOS and overall mortality.

Results: Technical success was 97.9%. Rebleeding occurred in 30% (15/50) of cases, predominantly within the first week. Risk was higher with malignancy (45.5%), systemic disease or bleeding disorders (29.4%), multivessel embolization (66.7% vs. 21.1%), and negative endoscopic findings (75% vs. 25.6%) (all $p < 0.05$). In-hospital mortality was 32% (16/50) and was associated with higher Rockall scores ($p = 0.001$) and lower platelet counts ($p = 0.037$). Overall mortality reached 52% (26/50) at a median follow-up of 16 months and was significantly associated with elevated INR ($p = 0.005$) and Rockall score ($p = 0.032$). Among survivors, mean LOS was 15.5 days and was correlated with comorbidity burden, transfusion requirements, and delayed embolization; lower baseline hemoglobin predicted longer stays. Complications occurred in 6% of patients: two minor access-site hematomas and one ischemic event.

Conclusion: TAE is a safe and effective treatment for NVGIB that is refractory to endoscopic therapy and is associated with high technical success and low complication rates. Mortality and rebleeding are primarily influenced by comorbidities, disease severity, and procedural complexity rather than by embolic technique, emphasizing the importance of individualized risk assessment and multidisciplinary management.

Keywords: Transcatheter arterial embolization, Gastrointestinal bleeding, Non-variceal hemorrhage, Rebleeding, Mortality, Prognostic factors

Received: Oct 15, 2025 **Accepted:** Dec 24, 2025 **Available Online:** May 22, 2026

INTRODUCTION

Gastrointestinal (GI) bleeding is a potentially life-threatening clinical emergency with significant morbidity, mortality, and healthcare costs worldwide [1]. Non-variceal upper gastrointestinal bleeding (NVUGIB) is most commonly caused by peptic ulcer disease, erosive gastritis, Mallory–Weiss tears, and malignancies [2]. Lower gastrointestinal bleeding (LGIB), on the other hand, is frequently caused by diverticulosis, angiodysplasia, ischemic colitis, colorectal neoplasia, or hemorrhoids [3,4].

Endoscopy and colonoscopy remain the first-line diagnostic and therapeutic approaches, enabling hemostasis in most pa-

https://doi.org/10.5455/annalsmedres.2025.10.306

tients with NVUGIB or LGIB [1,5]. Despite advances in endoscopic technologies, a subset of patients still experiences refractory or recurrent bleeding for which endoscopic management either fails or is technically infeasible [5].

Historically, surgery was the next line of therapy. Yet, surgical intervention in acute GI bleeding is associated with substantial morbidity and mortality, particularly among elderly patients and those with comorbidities [2,6]. Over the past few decades, transcatheter endovascular therapy, especially transcatheter arterial embolization (TAE), has emerged as a widely adopted salvage treatment, with its safety and efficacy for NVUGIB well documented in systematic reviews and clinical series [7–9].

In this context, we present our single-center experience with the endovascular management of patients with NVUGIB and LGIB refractory to endoscopic or colonoscopic therapy.

MATERIALS AND METHODS

This study was designed as a retrospective, single-center study and was approved by the Non-Interventional Clinical Research Ethics Committee of Kayseri City Hospital (Approval No: 548).

Patient selection

This retrospective study included 50 patients with acute nonvariceal upper or lower gastrointestinal bleeding who underwent transcatheter arterial embolization (TAE) at our institution between April 2023 and April 2025. Patients were eligible if bleeding could not be controlled endoscopically or if endoscopy was not feasible (e.g., gastrectomy). In such cases, TAE was performed prior to surgery when massive transfusion was required (more than four units of packed red blood cells [pRBCs] within a 24-hour period) or when hemodynamic instability was present, defined as hemorrhagic shock, systolic blood pressure of < 100 mmHg, or heart rate exceeding 100 beats per minute.

For each patient, demographic variables, presenting symptoms, medication histories, comorbidities, and previous gastrointestinal bleeding episodes were documented. Pre-embolization laboratory values (hemoglobin, INR, platelet count, lactate), pre-embolization transfusion requirements, prior imaging findings, Rockall scores, Forrest classification (if applicable), bleeding location, time from symptom onset to embolization, angiographic findings, type of embolic material used, and number and identity of occluded vessels were also recorded.

Embolization procedure

All angiographic procedures were performed by interventional radiologists with at least five years of experience. Vascular access was obtained via the common femoral artery in all cases. Procedures were typically performed under local anesthesia, with additional conscious sedation or general anesthesia administered when clinically required. A 6-French vascular sheath was introduced, and selective catheterization of the celiac trunk and the superior mesenteric artery was routinely performed using various 5-French diagnostic catheters (Taha Medical, Turkey; Cordis Medical, USA); the inferior mesenteric artery was catheterized when indicated. Iodinated contrast medium was used for all angiographic injections.

A coaxial microcatheter system (Carnelian, Tokai Medical, Japan) with a diameter of 1.9F–2.7F, selected according to the target vessel caliber was then advanced through the diagnostic catheter to perform selective angiography of the suspected bleeding site. When necessary, cone-beam computed tomography (CBCT) was performed either through the diagnostic catheter or selectively via the microcatheter to better localize the bleeding source. All procedures were performed with a digital subtraction angiography system (Artis zee, Siemens Healthcare, Forchheim, Germany) under standard fluoroscopic guidance.

Angiographic findings were categorized as direct or indirect signs of bleeding. Direct findings included contrast extravasation and pseudoaneurysm formation. Indirect findings included vascular irregularity, vasospasm, and abnormal mucosal or tumoral blush. Embolization was performed when there was direct evidence of active bleeding or indirect signs suggestive of ongoing hemorrhage. In addition, when no angiographic signs of bleeding were identified, empiric embolization was performed if supported by complementary findings from endoscopy, colonoscopy, computed tomography angiography (CTA), or red blood cell scintigraphy.

Embolic agents used in this study included detachable coils (Prestige Plus, Balt, France), calibrated microspheres (Embosphere, Merit Medical, USA), and a mixture of n-butyl-2-cyanoacrylate (NBCA; Histoacryl, B. Braun, Germany) and Lipiodol (Lipiodol Ultra Fluid, Guerbet, France). Coils were selected according to the vessel anatomy and diameter. Microspheres were generally preferred in patients with tumoral blush to achieve distal embolization, whereas NBCA was chosen in cases of vasospasm or massive hemorrhage requiring rapid occlusion. NBCA was mixed with lipiodol in ratios ranging from 1:4 to 1:8, determined by the vessel diameter and whether proximal or distal embolization was intended.

Technical success was defined as the successful embolization of the intended target vessel or vascular territory, with achievement of the planned angiographic endpoint; this definition included empiric embolization of the arterial supply to the suspected bleeding site when supported by preprocedural localization. Clinical success was defined as immediate hemostasis and the absence of rebleeding during follow-up.

Definitions of clinical outcomes

The primary outcomes of this study were rebleeding and in-hospital mortality. Rebleeding was defined as the recurrence of hematemesis, melena, or hematochezia after initial stabilization, associated with either a decrease in hemoglobin of

https://doi.org/10.5455/annalsmedres.2025.10.306

≥2 g/dL or the need for an additional transfusion or urgent intervention within 30 days. Secondary outcomes included length of hospital stay and overall mortality during follow-up. Procedure-related complications were also assessed and classified according to the Society of Interventional Radiology (SIR) guidelines as minor or major adverse events.

Statistical analysis

Statistical analysis was performed using SPSS Statistics software version 24.0 (IBM Corp., Armonk, NY, USA). Continuous variables were summarized as mean ± standard deviation, median, and minimum–maximum values, whereas categorical variables were presented as counts and percentages. The Shapiro–Wilk test was used to assess normality and the Levene test to evaluate homogeneity of variances. For comparisons between groups, an independent-samples t-test was used for normally distributed variables, whereas the Mann–Whitney U test was used for non-normally distributed variables. Associations between categorical variables were analyzed using Pearson's chi-square test, and effect sizes were quantified using the Phi coefficient for 2×2 contingency tables or Cramér's V for larger contingency tables. Correlations between continuous variables were assessed with Pearson's or Spearman's correlation coefficients, depending on distributional assumptions. A p-value <0.05 was considered statistically significant for all tests.

Because this retrospective cohort included all consecutive patients undergoing TAE within the study period, no a priori sample size calculation was performed. Post hoc power analyses were conducted. These analyses were based on observed effect sizes obtained from the dataset and were performed using two-sided tests with an α level of 0.05. Power values were calculated for categorical comparisons (Chi-square tests) and for continuous variables (correlation analyses) to assess whether the major associations investigated had adequate statistical power.

RESULTS

Patient characteristics

The study cohort consisted of 14 women and 36 men, with a mean age of 64.8 ± 16.7 years (median, 64.5 years; range, 15–92 years). Seven patients (14%) presented with lower gastrointestinal bleeding, while 43 patients (86%) had upper gastrointestinal bleeding (Table 1).

The most frequent symptom was melena, observed alone (n=23, 46%), in combination with hematemesis (n=9, 18%), or with hematochezia (n=7, 14%). Isolated hematemesis occurred in 6 patients (12%) and isolated hematochezia in 3 patients (6%). In addition, bleeding was detected in the surgical drains of two patients (4%).

Eleven patients (22%) had no comorbid conditions. Malignancy was present in 22 patients (44%); 12 had gastrointestinal malignancies and 10 had non-gastrointestinal malignancies. Fourteen patients (28%) had systemic diseases, including diabetes mellitus, hypertension, coronary artery disease, chronic kidney disease, liver cirrhosis, and cerebrovascular disease. In addition, three patients (6%) had bleeding disorders (immune thrombocytopenic purpura or von Willebrand factor deficiency). Nineteen patients had one comorbidity; nine had two comorbidities; nine had three comorbidities; one had four comorbidities; and one had five comorbidities.

Regarding medication history, 12 patients (24%) were receiving antiplatelet or anticoagulant therapy, and 9 (18%) were undergoing chemotherapy. 25 patients (50%) had no regular medications, while 4 patients (8%) were receiving other medications, including steroids, anti-HIV therapy, and antiparkinsonian medications.

Pre-procedural laboratory values and transfusion

Mean pre-procedural hemoglobin, platelet count, and INR were 7.2 ± 1.3 g/dL (range: 4.2–10.5), 166 ± 90 $\times 10^9$/L (range: 29–428), and 1.28 ± 0.18 (range: 1.05–1.81), respectively. The median lactate level was 1.1 mmol/L (range: 0.4–16.0). The median number of units of pRBCs transfused prior to the procedure was 6.5 (range: 1–26); the mean was 8.3 ± 5.6 (Table 2). The mean interval between symptom onset and TAE was 4.7 ± 3.7 days (median: 4.0 days, range: 0–15 days).

Endoscopic evaluation and risk stratification

Except for the two patients who presented with bleeding from surgical drains and one patient with a prior gastrectomy, all pa-

Table 1. Baseline demographic and clinical characteristics

Variable	Value
Age (years)	64.8 ± 16.7 (median 64.5; range 15-92)
Sex	
- Male	36 (72%)
- Female	14 (28%)
Bleeding site	
- Upper GI	43 (86%)
- Lower GI	7 (14%)
History of GI bleeding	
- No	35 (70%)
- Yes	15 (30%)
Comorbidity count	
- 0	11 (22%)
- 1	19 (38%)
- 2	9 (18%)
- 3	11 (22%)
Medication	
- None	25 (50%)
- Antiplatelet/anticoagulant	12 (24%)
- Chemotherapy	9 (18%)
- Other (steroid, anti-HIV, anti-parkinsonian)	4 (8%)

GI: Gastrointestinal.

https://doi.org/10.5455/annalsmedres.2025.10.306

Table 2. Pre-procedural laboratory values and diagnostic work-up.

Variable	Value
Pre-procedural Hb (g/dL)	7.2 ± 1.3 (range 4.2-10.5)
Pre-procedural Plt (×10³/μL)	166 ± 90 (range 29-428)
Pre-procedural INR	1.28 ± 0.18 (range 1.05-1.81)
Pre-procedural lactate (mmol/L)	1.8 ± 2.4 (range 0.4-16.0)
Blood transfusion requirement (units of pRBCs)	8.3 ± 5.6 (median 6.5; range 1-26)
Endoscopy/colonoscopy	
-Not performed	2 (4%)
-Positive findings	44 (88%)
-Negative findings	4 (8%)
Rockall score	5.0 ± 1.8 (range 2-9)
Forrest classification	37
-1a	8
-1b	16
-2	10
-3	3
CTA/red blood cell scintigraphy	
-Not performed	25 (50%)
-Positive findings	16 (32%)
-Negative findings	9 (18%)

Hb: Hemoglobin; Plt: Platelet count; INR: International Normalized Ratio; pRBCs: packed Red Blood; Cells; CTA: Computed Tomography Angiography.

Table 3. Procedural findings, embolization details and outcomes.

Variable	Value
Time from symptom onset to DSA (days)	4.7 ± 3.7 (median 4; range 0-15)
DSA/CBCT	
- Positive findings	
• extravasation	18 (36%)
• blush	13 (26%)
• vasospasm	4 (8%)
• pseudoaneurysm	2 (4%)
• vascular irregularity	1 (2%)
- Negative findings	12 (24%)
Embolic agents	
-None	3 (6%)
-Coils	31 (62%)
-Microspheres	9 (18%)
-NBCA	7 (14%)
Number of embolized vessels	
-None	3 (6%)
-1	38 (76%)
-≥2	9 (18%)
Periprocedural complications	
-Minor	2 (4%)
-Major	1 (2%)

DSA: Digital Subtraction Angiography; CBCT: Cone-beam Computed Tomography; NBCA: N-butyl-2-cyanoacrylate.

tients underwent either an upper endoscopy or a colonoscopy prior to angiography.

Endoscopic findings were negative in 4 patients (8%). Among the patients with ulcer bleeding (n=37), Forrest classification revealed that the majority were classified as Ib (43.2%), followed by Ia (21.6%); the remainder were distributed across IIa, IIb, IIc, and III categories. The mean Rockall score for the entire cohort was 5.0 ± 1.8 (median, 5; range, 2–9) (Table 2).

Procedural findings

Angiographic evaluation revealed contrast extravasation (Figure 1) in 17 patients (34%) and tumoral blush (Figure 2) in 12 patients (24%); these were the most frequent angiographic findings. No angiographic abnormality was detected in 14 patients (28%); however, cone-beam CT demonstrated mucosal blush or extravasation in two of these patients.

Embolization was not performed in 3 patients (6%): in 2 patients (4%) because there was no bleeding and therefore no

https://doi.org/10.5455/annalsmedres.2025.10.306

Table 4. Outcomes.

Outcome		
Rebleeding	15 (30%)	Comorbidity status: p=0.035 -Malignancy (45.5%) -Systemic or bleeding disorders (29.4%) -None (0%) Multivessel embolization: p=0.006 Negative endoscopic findings: p=0.035
Management of rebleeding -Conservative -Repeat embolization -Surgery	 6 (12%) 6 (12%) 4 (8%)	
In-hospital mortality	16 (32%)	Rockall score: p=0.001 Plt count(inversely correlated): p=0.037
Overall mortality	26 (52%)	Rockall score: p=0.032 INR: p=0.005
Length of hospital stay (days) *	15.5 ± 8.1 (median 14; range 1-43)	Number of comorbidities: p=0.019 Transfusion requirement: p<0.001 Bleeding to TAE time: p<0.001 Hb (inversely correlated): p=0.042

Plt: Platelet count; INR: International Normalized Ratio; Hb: Hemoglobin. * The length of hospital stay was calculated after excluding patients who died during the index hospitalization.

indication for empiric embolization, and in 1 patient (2.1%) because catheterization of the left gastric artery failed (technical failure). Accordingly, the overall technical success rate was 97.9% (47/48). Among the 47 patients who underwent embolization, 38 (81%) underwent single-vessel embolization, while 9 (19%) underwent embolization of multiple vessels (≥2). Coils were the most frequently used embolic material, applied in 31 patients (66%). Microspheres were used in 9 patients (19%) and NBCA-lipiodol mixture was used in 7 patients (15%).

Three patients (6%) experienced periprocedural complications. Minor complications consisted of two femoral access-site hematomas, both of which were managed conservatively. One major complication was non-target embolization that led to ischemia in a patient who subsequently required surgery because hemostasis could not be achieved.

Rebleeding

Rebleeding occurred in 15 patients (30%) within 30 days after the index procedure. The majority of rebleeding episodes (12 of 15, 80%) occurred during the first week.

When patients were stratified by comorbidity status, no rebleeding was observed in those without comorbidities (0/11, 0%). Rebleeding occurred in 10 of 22 patients (45.5%) with malignancy, the majority of whom (9 patients) had underlying gastrointestinal cancers. Among 17 patients with systemic diseases or bleeding disorders, rebleeding was observed in 5 (29.4%). This association was statistically significant ($\chi^2(2) = 6.705$, $p = 0.035$), with a moderate effect size (φ/Cramér's $V = 0.370$).

Rebleeding was also significantly associated with the number of embolized vessels. Rebleeding occurred in 8 of 38 patients (21.1%) who underwent single-vessel embolization, compared with 6 of 9 patients (66.7%) in the multiple-vessel group. This difference was statistically significant (Fisher's exact test, $p = 0.006$) and had a moderate-to-strong effect size (Phi = 0.444).

Furthermore, patients with negative endoscopic or colonoscopic findings prior to embolization (n = 4) had a markedly higher rate of rebleeding (3/4, 75%) compared with those with positive endoscopic findings (11/43, 25.6%). This association was statistically significant ($\chi^2(1)=4.437$, $p=0.035$), with a moderate effect size (Phi=0.304).

No significant associations were found between rebleeding and pre-procedural laboratory values (hemoglobin, INR, platelet count, lactate), age, sex, time from symptom onset to procedure, Forrest classification, Rockall score, or the type of embolic agent used (all $p > 0.05$).

Among the 15 patients who developed rebleeding, 6 were managed conservatively (1 in-hospital death), 5 underwent repeat TAE (1 in-hospital death), and 4 required surgery (3 in-hospital deaths).

Mortality

In-hospital mortality occurred in 16 patients (32%). A significant association was observed between the Rockall score and in-hospital mortality ($p = 0.001$). In addition, pre-procedural platelet levels were significantly lower in patients who died in the hospital (128 ± 73 vs. $184 \pm 92 \times 10^3/\mu L$, $p = 0.037$). No significant associations were found with other laboratory parameters (hemoglobin, INR, lactate).

When medication groups (chemotherapy, antiplatelet/anticoagulant therapy, and no regular medication) were analyzed, no

https://doi.org/10.5455/annalsmedres.2025.10.306

statistically significant association between medication group and in-hospital mortality was found (p = 0.209).

In-hospital mortality was higher in patients with comorbidities than in those without (36% vs. 18%; odds ratio approximately 2.5), although the difference was not statistically significant (p = 0.266). When stratified by comorbidity burden, mortality rates were 18.2% for no comorbidities, 36.8% for one comorbidity, 22.2% for two comorbidities, and 45.5% for three or more comorbidities. While these findings suggest a trend toward increased mortality with higher comorbidity burden, the association did not reach statistical significance (p = 0.479).

Overall mortality was 52% (26/50) during follow-up, which ranged from 4 to 28 months (median 16 months). Both pre-procedural INR (p = 0.005) and Rockall score (p = 0.032) were significantly associated with overall mortality, whereas other baseline laboratory and demographic parameters were not significantly associated (all p>0.05). Mortality was 27.3% in patients without comorbidities, compared with 54.5% in those with malignancies and 64.7% in those with systemic diseases or bleeding disorders. Similarly, overall mortality varied by comorbidity burden: 63.2% in patients with one comorbidity, 55.0% in those with two or more, and 27.3% in those without comorbidities. Although these findings suggest a clinically relevant trend toward higher overall mortality in patients with comorbidities, the differences did not reach statistical significance (p = 0.146 for type; p = 0.156 for count).

Length of stay

Length of hospital stay (LOS) was analyzed for the 34 patients discharged alive; patients with in-hospital mortality were excluded. The mean LOS was 15.5 ± 8.1 days, with a median of 14 days (range: 1–43 days).

A significant positive correlation was also found between the number of comorbidities and LOS (r=0.399, p=0.019). Similarly, LOS was positively correlated with units of pRBCs transfused (r = 0.593, p<0.001) and with the time interval between symptom onset and embolization (r = 0.636, p < 0.001). Pre-procedural hemoglobin levels were inversely correlated with LOS (r = -0.350, p = 0.042), indicating that patients with lower baseline hemoglobin had longer hospital stays.

No significant associations were found between LOS and age, sex, platelet count, INR, Rockall score, Forrest classification, angiographic findings, or the type of embolic agent used (all p>0.05).

Procedural findings and outcomes are summarized in Tables 3 and 4.

Post hoc power analyses demonstrated that the association between multivessel embolization and rebleeding (χ^2 = 9.08, w ≈ 0.44, N = 47) had statistical power of approximately 0.85. The correlation between length of stay and transfusion requirement (r = 0.59, n = 34) showed a power of approximately 0.98. For moderate correlations, such as the relationship between length of stay and comorbidity count (r≈0.40; n=34), the achieved power was approximately 0.70. Smaller effect sizes yielded correspondingly lower power levels.

DISCUSSION

In our study, TAE demonstrated a high technical success rate and acceptable clinical outcomes, consistent with previously published reports. While the efficacy and safety of TAE have been well documented in both upper and lower GI bleeding, considerable variability persists across studies with regard to rebleeding, mortality, and complication rates. Therefore, placing our results in the context of current evidence provides valuable insights into patient selection, procedural factors, and outcome predictors. Moreover, current international guidelines recommend TAE as the preferred second-line therapy after failed endoscopy, reserving surgery for refractory cases [10,11].

Our rebleeding rate of 30% falls within the broad range of 9–56% reported in prior studies [12–15]. Importantly, our findings underscore specific risk factors for recurrence. Rebleeding was significantly more common in patients with malignancy (particularly gastrointestinal cancers), in those undergoing multi-vessel embolization, and in those in whom no bleeding source was identified endoscopically prior to TAE. These factors have also been highlighted in previous reports. Dablan et al., focusing exclusively on non-variceal upper GI bleeding, identified multi-vessel embolization and comorbidity as predictors of recurrence [16], while Yu et al., in a meta-analysis of lower GI bleeding, found malignancy and coagulopathy to be significant risk factors [17]. In addition, Lee et al. reported a rebleeding rate of 37.3% in a mixed cohort of upper and lower GI bleeding, with coagulopathy and massive transfusion independently associated with recurrence [15]. Extrat et al. similarly reported an early rebleeding rate of 25% and identified the units of pRBCs as the main predictor of recurrence [18]. By contrast, we observed no significant association between rebleeding and demographic variables, baseline laboratory parameters, or embolic material. Taken together, our results suggest that rebleeding after TAE is determined more strongly by underlying disease burden, procedural complexity, and the adequacy of bleeding source localization than by demographic characteristics or the choice of embolic agent.

In-hospital mortality was 32, and overall mortality during follow-up was 52%. These rates are relatively high compared with those in the most recent series, in which 30-day mortality has been the primary outcome. Ozen et al. [13] reported a 30-day mortality rate of 22%; Lee et al. [15] reported a rate of 18.6%; and Loffroy's meta-analysis reported a wide range from 4% to 46% [12]. Dablan et al. [16] reported a 30-day mortality rate of 48.2%, which is broadly comparable to that observed in our population. Extrat et al. observed a 30-day mortality rate of 22.1%, with hyperlactatemia (≥2 mmol/L) emerging as the strongest predictor of early mortality [18].

https://doi.org/10.5455/annalsmedres.2025.10.306

We identified higher Rockall scores and lower platelet counts as predictors of in-hospital mortality, while elevated INR and Rockall scores were associated with overall mortality. The finding that INR was not predictive of short-term mortality but was predictive of long-term outcomes is particularly noteworthy because it may reflect the influence of subclinical coagulation abnormalities, frailty, and the cumulative burden of comorbidities beyond the acute hemodynamic phase of bleeding.

Rockall score, by contrast, consistently correlated with both in-hospital and overall mortality, as expected given its established prognostic value. Compared with prior studies, our results partially diverge; Yu et al. [17] highlighted coagulopathy, malignancy, and massive transfusion as predictors of mortality, while Lee et al. [15] identified coagulopathy as a significant factor.

In our series, malignancy and transfusion burden were not statistically significant, though clinically relevant trends were observed. Another important observation is that overall mortality rates substantially exceeded in-hospital mortality rates, underscoring the long-term vulnerability of this patient group, particularly those with malignancy and systemic disease. This emphasizes the need for structured long-term follow-up and multidisciplinary management, including oncology, internal medicine, and supportive care teams [19].

The mean length of stay (LOS) was 15.5 days, calculated only for patients discharged alive to avoid spuriously shortened LOS values due to early in-hospital deaths. This survivor-only approach provides a more accurate reflection of the hospitalization burden, yet it makes direct comparisons with prior studies difficult, as LOS has rarely been systematically reported in TAE cohorts. Among the few that did report, Nykänen et al. [20] documented a median LOS of 7 days (range, 1–91) after TAE for lower GI bleeding, which is considerably shorter than that observed in our series. Extrat et al. [18] similarly reported a median LOS of around 12 days following TAE, while Twene et al. [21], in a comparative analysis, demonstrated that patients treated with TAE had shorter hospitalizations than those undergoing surgery. Our finding of a longer LOS likely reflects, in our cohort, the higher comorbidity burden, greater transfusion requirements, delayed embolization, and lower baseline hemoglobin levels. These associations are consistent with prior observations that disease severity, massive transfusion, and low hemoglobin are more influential than procedural variables in determining resource utilization after TAE. Taken together, our results suggest that prolonged LOS in this setting primarily reflects underlying disease complexity and delayed presentation rather than the embolization procedure itself, underscoring the importance of early referral and optimization of preprocedural status.

Procedure-related complications occurred in three patients (6%), consisting of two minor femoral access-site hematomas and one major event that was identified intraoperatively in a patient who underwent surgery for persistent bleeding. This rate is at the lower end of the spectrum reported in the literature, where complication rates vary widely from 0% to 26% [12]. Extrat et al. observed peri- and post-procedural complications in nearly 30% of cases, with ischemia being the most frequent adverse event, and Nykänen et al. reported a complication rate of 36% in lower GI bleeding [18,20]. Compared with these findings, our series demonstrates a relatively low complication burden, which may reflect the use of superselective techniques. Importantly, consistent with previous reports, we found no significant difference between empiric and selective embolization with respect to complications or clinical outcomes [22,23]. This supports current evidence that empiric embolization, when guided by reliable pre-procedural findings, is a safe and effective option for carefully selected patients. Overall, TAE offers an acceptable safety profile and remains a less invasive and safer option compared with surgery for refractory gastrointestinal bleeding.

Limitations

This study has several limitations that should be acknowledged. First, its retrospective and single-center design may introduce selection bias and limit the generalizability of our findings. Second, the sample size was relatively small, which may have reduced the statistical power to detect associations for some variables and may partly explain the discrepancies with previous reports. Third, although we analyzed both upper and lower gastrointestinal bleeding together to reflect real-world practice, this heterogeneity may have influenced the outcome measures, because prognostic factors can differ between these populations. Fourth, variations in embolic materials and empirical versus selective embolization techniques were determined by clinical context rather than by random allocation, which may have introduced procedural bias. And finally, non-significant associations with small-to-moderate effect sizes should be interpreted with caution, as post-hoc power analyses indicated limited statistical power for detecting weaker effects.

CONCLUSION

Transcatheter arterial embolization (TAE) is a safe and effective option for non-variceal gastrointestinal bleeding refractory to endoscopic therapy, offering high technical success rates and low complication rates. Rebleeding was primarily linked to malignancy, multi-vessel embolization, and negative endoscopic findings, while mortality correlated with Rockall score, platelet count, and INR. These results emphasize that outcomes are mainly determined by comorbidity burden and disease severity rather than procedural technique. Overall, TAE remains a less invasive and reliable alternative to surgery.

Ethics Committee Approval: Approved by the Ethics Committee of Kayseri City Hospital, Non-Interventional Clinical Research (Approval No:548).

https://doi.org/10.5455/annalsmedres.2025.10.306

Informed Consent: This was not deemed necessary because the study was retrospective.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declared no competing interests.

Author Contributions: S.K. and M.B. contributed to the conception and design of the study. S.K. wrote the paper. S.K., O.A., and F.O. performed data collection and image analysis. N.B.K. and G.T. contributed to statistical analysis and interpretation of results. All authors participated in drafting and critically revising the manuscript, approved the final version, and agree to be accountable for the integrity of the work.

Financial Disclosure: No financial support was obtained for this study.

Artificial Intelligence Disclosure: The authors declare that an artificial intelligence–based tool (ChatGPT, OpenAI, version 5.0) was used solely for language editing and improving the clarity and readability of the manuscript. The AI tool did not contribute to the study design, data analysis, interpretation of results, or scientific conclusions. All content was critically reviewed and approved by the authors, who take full responsibility for the manuscript.

REFERENCES

1. Barkun AN, Almadi M, Kuipers EJ, et al. Management of Nonvariceal Upper Gastrointestinal Bleeding: Guideline Recommendations From the International Consensus Group. *Ann Intern Med.* 2019;171(11):805-822. doi: 10.7326/M19-1795.
2. Laine L, Jensen DM. Management of Patients With Ulcer Bleeding. *Am J Gastroenterol.* 2012;107(3):345-60. doi: 10.1038/ajg.2011.480.
3. Oakland K. Changing epidemiology and etiology of upper and lower gastrointestinal bleeding. *Best Pract Res Clin Gastroenterol.* 2019;42-43:101610. doi: 10.1016/j.bpg.2019.04.003.
4. Strate LL, Gralnek IM. ACG Clinical Guideline: Management of Patients With Acute Lower Gastrointestinal Bleeding. *Am J Gastroenterol.* 2016;111(4):459-74. doi: 10.1038/ajg.2016.41.
5. Kamboj AK, Hoversten P, Leggett CL. Upper Gastrointestinal Bleeding: Etiologies and Management. *Mayo Clin Proc.* 2019;94(4):697-703. doi: 10.1016/j.mayocp.2019.01.022.
6. Ghassemi KA, Jensen DM. Lower GI bleeding: epidemiology and management. *Curr Gastroenterol Rep.* 2013;15(7):333. doi: 10.1007/s11894-013-0333-5.
7. Loffroy R, Favelier S, Pottecher P, et al. Transjugular intrahepatic portosystemic shunt for acute variceal gastrointestinal bleeding: Indications, techniques and outcomes. *Diagn Interv Imaging.* 2015;96(7-8):745-55. doi: 10.1016/j.diii.2015.05.005.
8. Ini' C, Distefano G, Sanfilippo F, et al. Embolization for acute nonvariceal bleeding of upper and lower gastrointestinal tract: a systematic review. *CVIR Endovasc.* 2023;6(1):18. doi: 10.1186/s42155-023-00360-3.
9. Chevallier O, Comby PO, Guillen K, et al. Efficacy, safety and outcomes of transcatheter arterial embolization with N-butyl cyanoacrylate glue for non-variceal gastrointestinal bleeding: A systematic review and meta-analysis. *Diagn Interv Imaging.* 2021;102(7-8):479-487. doi: 10.1016/j.diii.2021.03.004.
10. Gralnek IM, Dumonceau JM, Kuipers EJ, et al. Diagnosis and management of nonvariceal upper gastrointestinal hemorrhage: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy.* 2015;47(10):a1–a46. doi: 10.1055/s-0034-1393172.
11. Laine L, Barkun AN, Saltzman JR, Martel M, Leontiadis GI. ACG Clinical Guideline: Upper Gastrointestinal and Ulcer Bleeding. *Am J Gastroenterol.* 2021;116(5):899-917. doi: 10.14309/ajg.0000000000001245.
12. Loffroy R, Rao P, Ota S, De Lin M, Kwak BK, Geschwind JF. Embolization of acute nonvariceal upper gastrointestinal hemorrhage resistant to endoscopic treatment: results and predictors of recurrent bleeding. *Cardiovasc Intervent Radiol.* 2010;33(6):1088-1100. doi: 10.1007/s00270-010-9829-7.
13. Ozen C, Al-Hashimi M, Tornby Stender M, Thorlacius-Ussing O, Larsen AC. Transarterial embolization of gastroduodenal peptic ulcer bleeding: a single-center study of safety and efficacy. *Langenbecks Arch Surg.* 2025;410(1):117. doi: 10.1007/s00423-025-03695-8.
14. Matsumoto T, Yoshimatsu R, Shibata J, et al. Transcatheter arterial embolization of nonvariceal gastrointestinal bleeding with n-butyl cyanoacrylate or coils: a systematic review and meta-analysis. *Sci Rep.* 2024;14(1):27377. doi: 10.1038/s41598-024-79133-4.
15. Lee S, Kim T, Han SC, Pak H, Jeon HH. Transcatheter arterial embolization for gastrointestinal bleeding: Clinical outcomes and prognostic factors predicting mortality. *Medicine (Baltimore).* 2022;101(31):e29342. doi: 10.1097/MD.0000000000029342.
16. Dablan A, Uluman YB, Güzelbey T, Cingöz M, Kınacı E, Kılıçkesmez Ö. Identifying Predictors of Early Rebleeding Following Transcatheter Embolization in Acute Non-Variceal Upper Gastrointestinal Bleeding: A Retrospective Analysis. *Osmangazi Journal of Med.* 2025;47(4):556-566. doi: 10.20515/otd.1578761.
17. Yu Q, Funaki B, Ahmed O. Twenty years of embolization for acute lower gastrointestinal bleeding: a meta-analysis of rebleeding and ischaemia rates. *Br J Radiol.* 2024;97(1157):920-932. doi: 10.1093/bjr/tqae037.
18. Extrat C, Grange S, Mayaud A, et al. Transarterial Embolization for Active Gastrointestinal Bleeding: Predictors of Early Mortality and Early Rebleeding. *J Pers Med.* 2022;12(11):1856. doi: 10.3390/jpm12111856.
19. Oakland K, Chadwick G, East JE, Guy R, Humphries A, Jairath V. Diagnosis and management of acute lower gastrointestinal bleeding: guidelines from the British Society of Gastroenterology. *Gut.* 2019;68(5):776–789. doi: 10.1136/gutjnl-2018-317807.
20. Nykänen T, Peltola E, Kylänpää L, Udd M. Transcatheter Arterial Embolization in Lower Gastrointestinal Bleeding: Ischemia Remains a Concern Even with a Superselective Approach. *J Gastrointest Surg.* 2018;22(8):1394-1403. doi: 10.1007/s11605-018-3728-7.
21. Twene P, Ogunti R, Asemota J, et al. In-Hospital Outcomes of Transcatheter Arterial Embolization vs Surgical Control of Refractory Bleeding Peptic Ulcer Disease: A Propensity Score-Matched Analysis. *Am J Gastroenterol.* 2020;115(Suppl):S296. doi: 10.14309/01.ajg.0000704416.26752.a6.
22. Yu Q, Funaki B, Navuluri R, et al. Empiric Transcatheter Embolization for Acute Arterial Upper Gastrointestinal Bleeding: A Meta-Analysis. *AJR Am J Roentgenol.* 2021;216(4):880-893. doi: 10.2214/AJR.20.23151.
23. Hermie L, Dhondt E, Vanlangenhove P, De Waele J, Degroote H, Defreyne L. Empiric cone-beam CT-guided embolization in acute lower gastrointestinal bleeding. *Eur Radiol.* 2021;31(4):2161-2172. doi: 10.1007/s00330-020-07232-7.

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Pediatric atypical antipsychotic poisoning: A single-center cohort study

Kivanc Terzi

Ağrı Training and Research Hospital, Department of Pediatrics, Division of Pediatric Intensive Care, Ağrı, Türkiye

**Corresponding author: kvnctrz@gmail.com (Kivanc Terzi)*

MAIN POINTS

- In our cohort, atypical antipsychotic poisonings were predominantly mild to moderate in severity, and no fatalities occurred.
- Risperidone exposure was the most common, and risperidone-associated extrapyramidal symptoms were resolved with biperiden.
- The prognosis was favorable with symptom-directed supportive care, and patients achieved complete recovery.

Cite this article as: Terzi K. Pediatric atypical antipsychotic poisoning: A single-center cohort study. *Ann Med Res*. 2026;33(5):238–242. doi: 10.5455/annalsmedres.2025.10.304.

ABSTRACT

Aim: This study aimed to describe the clinical features, exposure characteristics, management, and outcomes of pediatric atypical antipsychotic poisonings treated in a single tertiary pediatric intensive care unit (PICU) and to contextualize findings using the Poisoning Severity Score (PSS).

Materials and Methods: We conducted a retrospective, descriptive review of children and adolescents (1 month–18 years) managed for single-agent atypical antipsychotic ingestion at the PICU of Ağrı Training and Research Hospital between January 2018 and January 2023. De-identified electronic records were abstracted for demographics, implicated agent (amisulpride, aripiprazole, quetiapine, olanzapine, risperidone, ziprasidone), dose (mg and mg/kg), time from ingestion to presentation, decontamination (gastric lavage, activated charcoal), treatments (including biperiden), clinical/laboratory findings, and outcomes. Severity was classified according to PSS (0–4). Continuous data are reported as median (IQR); categorical data are reported as n (%). A two-sided $p<0.05$ was considered significant.

Results: Forty-one pediatric cases were included. The most frequent agent was risperidone (46.3%). Decontamination included gastric lavage in 43.9% and activated charcoal in 46.3%. Tachycardia, hypokalemia, confusion, extrapyramidal symptoms (EPS), and drowsiness were common findings. The PSS distribution was none 14.6%, minor 43.9%, moderate 36.6%, and severe 4.9%; no deaths occurred (therefore, no PSS grade 4). In a risperidone subgroup analysis (n = 19), biperiden was administered to 6 patients for acute EPS; time to presentation tended to be longer in those given biperiden (median 9.0 vs 4.0 h; p = 0.08). Ordinal logistic modeling indicated higher PSS categories in the biperiden group (odds ratio [OR], 68.11; 95% confidence interval [CI], 3.59–1291.68; p = 0.005).

Conclusion: In this cohort, pediatric atypical antipsychotic poisonings were predominantly mild to moderate with no fatalities. Risperidone exposures were the most common and frequently associated with EPS, which resolved with biperiden. These data support a generally favorable prognosis with timely supportive care and targeted symptomatic management.

Keywords: Poisoning, Pediatrics, Antipsychotic agents, Risperidone, Biperiden

Received: Oct 13, 2025 **Accepted:** Dec 26, 2025 **Available Online:** May 22, 2026

INTRODUCTION

Atypical (second-generation) antipsychotics are increasingly prescribed not only in adults but also in pediatric populations for conditions such as autism spectrum disorders, bipolar disorder, and schizophrenia. As a result, these medications have become more prevalent in households, increasing the risk of pediatric exposures through accidental ingestion or intentional overdose [1]. Atypical antipsychotics include risperidone, quetiapine, olanzapine, aripiprazole, ziprasidone, and amisulpride. Each of these agents carries a risk of toxicity in children, with clinical manifestations related to their pharmacologic profiles. In overdose, central nervous system (CNS) depression and cardiovascular effects are prominent across this class [2]. In contrast, risperidone and ziprasidone, which have potent dopamine D_2 receptor antagonism and less anticholinergic activity, tend to produce less sedation but can precipitate extrapyramidal symptoms (EPS) such as acute dystonia, especially in pediatric patients [3].

This study aimed to characterize the presentation, management (including biperiden use), and outcomes (PSS-based severity) of pediatric atypical antipsychotic poisonings and compare these patterns with recent pediatric reports.

MATERIALS AND METHODS

The study was designed as a retrospective, descriptive review of cases of atypical antipsychotic poisoning in the Pediatric Intensive Care Unit of Ağrı Training and Research Hospital between January 2018 and January 2023. The study was approved by the Ağrı İbrahim Çeçen University Scientific Research Ethics Committee (approval date: May 25, 2023; approval number: 2023/132). Cases were identified through daily PICU admission logs and electronic health record queries using drug intoxication diagnosis codes. Institutional electronic records were deidentified and processed under the study identifiers.

The inclusion criteria were:

- Pediatric age (1 month–18 years)
- Single-drug exposure
- Exposure to an atypical antipsychotic at a known ingested dose.

The exposure was confirmed by the caregiver's history, corroborated by pill counts/packaging or medication lists, and the formulation was recorded (immediate vs. extended-release; drops vs. tablets; route of exposure). Demographic variables (age in months, sex, body weight); the implicated agent (amisulpride, aripiprazole, quetiapine, olanzapine, risperidone, ziprasidone); total dose (mg); circumstance of exposure (unintentional/therapeutic error/intentional/unknown); time from exposure to presentation (hours); decontamination procedures (activated charcoal, gastric lavage); patients' vital signs; and laboratory results (glucose, serum electrolytes, renal function tests, liver function tests, creatine kinase) were abstracted retrospectively from patient charts, along with administered treatments (biperiden). Time to presentation was defined as the interval from reported ingestion to first medical contact. Organ-system domains coded clinical effects, and overall poisoning severity was determined according to the Poisoning Severity Score (PSS) developed by the International Program on Chemical Safety and the European Association of Poison Center and Clinical Toxicologists (IPCS/EAPCCT). The severity grades were as follows: none (0), minor (1), moderate (2), severe (3), and fatal (4).

Statistical analysis

Analyses were performed in IBM SPSS Statistics, version 27.0 (Armonk, NY: IBM Corp.). Continuous variables are summarized as median (IQR) owing to non-normal distributions; when approximately normal, mean ± SD is also presented. Categorical variables are expressed as n (%). Normality was assessed using the Shapiro–Wilk test. Between-group comparisons were performed using the Mann–Whitney U test for 2 groups. Categorical comparisons were performed using the chi-square test or Fisher's exact test when the expected counts were <5. For severity graded by the Poisoning Severity Score (0–3), proportional-odds (ordinal logistic) models were additionally fitted to estimate common odds ratios (ORs) with 95% confidence intervals for prespecified contrasts (e.g., biperiden use vs no use), checking the proportional-odds assumption (test of parallel lines). All tests were two-sided, and $p<0.05$ was considered statistically significant. No a priori sample size calculation was conducted; using a two-sided $\alpha=0.05$ and the observed effect size (common OR 68.11; 95% CI 3.59–1291.68), the post-hoc power achieved was approximately 0.80 (80%).

RESULTS

The analysis included 41 pediatric cases of atypical antipsychotic intoxication. The median age was 163 months (IQR 51–198), and 56.1% of the participants were female (Table 1). The implicated agents were risperidone (46.3%), quetiapine (22.0%), olanzapine (19.5%), aripiprazole, ziprasidone,

Table 1. Patients' characteristics.

Age (month), median, (25-75p)	163 (51-95)	
	n	%
Sex		
Female	23	56.1
Male	18	43.9
Drugs		
Risperidone	19	46.3
Quetiapine	9	22
Olanzapine	8	19.5
Aripiprazole	2	4.9
Ziprasidone	2	4.9
Amisulpride	1	2.4
Decontamination		
Gastric lavage	18	43.9
Activated charcoal	19	46.3

Table 2. Symptoms and signs.

	n	%
Hypokalemia	10	24.4
Tachycardia	9	22
Confusion	7	17.1
Extrapyramidal symptoms	7	17.1
Drowsiness	4	9.8
Metabolic acidosis	4	9.8
Hyperthermia	3	7.3
Agitation	2	4.9
Dizziness	2	4.9
Vomiting	2	4.9
Hallucinations	2	4.9
Hypoglycemia	1	2.4
Transaminase elevation	1	2.4
Rhabdomyolysis	1	2.4
AV block (I)	1	2.4
Nausea	1	2.4
Hypotension	1	2.4
Coma	1	2.4

https://doi.org/10.5455/annalsmedres.2025.10.304

Table 3. Poisoning severity and ingested dose.

	n (%)	Dose range (mg/kg), median IQR (25–75p)
Total		
• None	6 (14.6)	
• Minor	18 (43.9)	
• Moderate	15 (36.6)	
• Severe	2 (4.9)	
Risperidone		
• None	2 (10.5)	0.10–0.16, 0.13 (0.12–0.15)
• Minor	10 (52.6)	0.07–0.45, 0.28 (0.15–0.32)
• Moderate	6 (31.6)	0.12–0.35, 0.25 (0.20–0.27)
• Severe	1 (5.3)	0.33
Quetiapine		
• None	2 (22.2)	0.7-3.5, 2.1 (1.4-2.8)
• Minor	4 (44.5)	2.5-33.3, 7.5 (2.65-17.55)
• Moderate	3 (33.3)	2.2-28.5, 5.7 (3.95-17.10)
Olanzapine		
• None	1 (12.5)	0.5
• Minor	2 (25)	1.2-2.6, 1.90 (1.55–2.25)
• Moderate	5 (62.5)	1-15.8, 1.7 (1.5-2.9)
Aripiprazole		
• None	1 (50)	1
• Minor	1 (50)	1.25
Ziprasidone		
• Minor	1 (50)	1
• Severe	1 (50)	8
Amisulpride		
• Moderate	1 (100)	25

Table 4. Clinical characteristics by Biperiden use.

	Biperiden not administered	Biperiden administered	p
n (%)	13 (68.4)	6 (31.6)	
Age (month)	180 (46-213)	57 (25-182)	0.059
Sex			
• Female	9 (69.2)	4 (66.7)	1.00
• Male	4 (30.8)	2 (33.3)	
Time to presentation	4 (2-6)	9 (5.8-10)	0.08
Decontamination			
• Gastric lavage	6 (46.2)	-	0.11
• Activated charcoal	8 (61.5)	1 (16.7)	0.14
Dose (mg)	10 (5-17)	5 (4.5-6.75)	0.21
Dose (mg/kg)	0.26 (0.15-0.32)	0.25 (0.2-0.27)	0.93

and amisulpride. Decontamination consisted of gastric lavage of 43.9% of the samples and activated charcoal in 46.3%.

Clinical findings (Table 2) were as follows: hypokalemia, 24.4%; tachycardia, 22.0%; confusion, 17.1%; extrapyramidal symptoms, 17.1%; drowsiness, 9.8%; metabolic acidosis, 9.8%; hyperthermia, 7.3%; agitation, 4.9%; dizziness, 4.9%; vomiting, 4.9%; hallucinations, 4.9%; and single instances (2.4% each) of hypoglycemia, transaminase elevation, rhabdomyolysis, first-degree AV block, nausea, hypotension, and coma.

The PSS distribution was as follows: none 14.6%, minor 43.9%, moderate 36.6%, and severe 4.9%. No deaths were recorded; therefore, no fatal (grade 4) intoxication occurred. Dose ranges by drug and severity are presented in Table 3.

Biperiden (Table 4) was more frequently used among younger patients, although the difference was not statistically signif-

https://doi.org/10.5455/annalsmedres.2025.10.304

icant (p = 0.059). The time to presentation tended to be longer in the biperiden group (9 vs. 4 h; p = 0.08). All other comparisons—sex distribution, gastric lavage, activated charcoal, and dose metrics (mg and mg/kg)—showed no between-group differences ($p \geq 0.05$). Biperiden administration was associated with a higher common odds of being in a more severe category (OR 68.11, 95% CI 3.59–1291.68; p = 0.005) in an ordinal logistic model for PSS 0–3.

DISCUSSION

In this 5-year retrospective series of pediatric atypical antipsychotic poisonings, we identified 41 cases managed in the PICU, with no fatalities and universally good outcomes. The children in our study ranged from infancy to late adolescence (approximately 1 year to 17 years old), reflecting two primary exposure scenarios: accidental ingestion in toddlers/young children and intentional overdose in adolescents. This age distribution is similar to patterns reported in the literature, where unintentional poisonings predominate in children and adolescent cases often involve deliberate self-harm or misuse [4]. Risperidone was the single most commonly used agent. This reflects the widespread use of risperidone in pediatric psychiatry. Our finding agrees with prior reports; for example, data from the U.S. poison center (2005–2013) showed that risperidone was responsible for thousands of pediatric exposures [1].

Despite being drawn from an intensive care population, the clinical severity of poisonings in our cohort was generally low to moderate. Using the PSS, over half of the patients (≈59%) had either no symptoms or only mild effects (PSS 0–1), and about one-third (37%) experienced moderate effects (PSS 2). Only two patients (4.9%) were classified as having severe poisoning (PSS 3), and none had fatal outcomes (PSS 4). These findings reinforce the favorable prognosis of single-agent atypical antipsychotic ingestion in children. They closely parallel the experience reported in other pediatric studies: for instance, Meli et al. observed no deaths and only two severe cases out of 106 young children (<6 years) exposed to risperidone, olanzapine, quetiapine, or clozapine [5]. Similarly, Stassinos and Klein-Schwartz, analyzing >16,000 U.S. cases in children <6, found that central toxicity was under 1% and there were no deaths [1]. All patients in our series made a full recovery, consistent with extensive pediatric studies showing that—absent polypharmacy—atypical antipsychotic poisonings rarely cause lasting harm; the lack of mortality in our cohort further indicates that, with appropriate supportive care, even children who develop severe symptoms can be successfully managed [6].

Several patients in our PICU had minimal or no symptoms despite significant reported ingestions. For example, some adolescents taking high doses of quetiapine or risperidone had a PSS of 0 or 1. This aligns with the observations of Lofton et al, who reported that adolescents can sometimes tolerate very high aripiprazole doses with only transient or minor effects [7]. In contrast, younger children in our study tended to manifest more noticeable toxicity from comparatively minor ingestions. The only two severe cases we encountered illustrate this point: a 19-month-old child who ingested ziprasidone (PSS: 3) and a 46-month-old who ingested risperidone (PSS: 3). In both cases, the patients were infants/toddlers, and the severity of their presentations was disproportionate to the doses that an older child or adult would tolerate. This age-related vulnerability has been well documented. Young children have immature physiology and often ingest doses per kilogram that far exceed therapeutic levels. For example, the pediatric case reported by Fasano et al. shows that ingestion of a single ziprasidone tablet in a 30-month-old child resulted in coma and respiratory depression, underscoring that severe toxicity can occur from low doses at a young age [8]. Our findings concur that younger pediatric patients are at higher risk of severe central nervous system (CNS) depression and other effects, reinforcing the need for aggressive management and observation in that age group.

Risperidone was used in nearly half of our cases, and it exhibited a fairly characteristic clinical profile. Most risperidone-ingested children in our cohort developed mild CNS depression and sinus tachycardia, which is consistent with the known effects of risperidone. Acute dystonic reactions have emerged as a salient issue in risperidone poisoning. Intravenous treatment with biperiden, which led to complete resolution of the dystonic symptoms. Acute anticholinergic treatment was is needed in more than a quarter nearly one-third of our risperidone cases. Previous pediatric studies have reported varying incidence rates of EPS. For example, Meli et al observed extrapyramidal symptoms in only 1% of <6-year-old cases [5]. In the study conducted by Vial et al., extrapyramidal findings were detected in 16.8% (n= 17) of the patients, and EPS was reported in only one patient aged 5 years [9]. Our higher rate reflects the inclusion of older children and adolescents (who may be more prone to EPS at higher doses) as well as intentional cases with higher doses ingested. Fortunately, these reactions were benign and rapidly reversible with the appropriate use of anticholinergic agents, though frightening for the patient and caregivers. In our series, no child suffered airway compromise due to dystonia, and none had recurrent EPS after treatment.

CONCLUSION

In conclusion, our discussion highlights that when appropriately managed, pediatric atypical antipsychotic poisonings have a low incidence of severe morbidity. PICU care was crucial for the handful of severe cases, and supportive measures were uniformly successful. With vigilant monitoring and symptomatic treatment, pediatric patients who ingest psychotropic medications can be expected to fully recover.

Ethics Committee Approval: The study was approved by the Ağrı İbrahim Çeçen University Scientific Research Ethics

https://doi.org/10.5455/annalsmedres.2025.10.304

Committee (approval date: May 25, 2023; approval number: 2023/132).

Informed Consent: This was not applicable due to the retrospective design.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author declares no conflicts of interest.

Financial Disclosure: The author declare that this study received no financial support.

Artificial Intelligence Disclosure: The author declare that no artificial intelligence tools were used in the preparation of this article.

REFERENCES

1. Stassinos G, Klein-Schwartz W. Comparison of pediatric atypical antipsychotic exposures reported to U.S. poison centers. *Clin Toxicol (Phila).* 2017;55(1):40-45. doi: 10.1080/15563650.2016.1233342.
2. Stroup TS, Gray N. Management of common adverse effects of antipsychotic medications. *World Psychiatry.* 2018;17(3):341-356. doi: 10.1002/wps.20567.
3. Page CB, Calver LA, Isbister GK. Risperidone overdose causes extrapyramidal effects but not cardiac toxicity. *J Clin Psychopharmacol.* 2010;30(4):387-90. doi: 10.1097/JCP.0b013e3181e5f7a5.
4. Kameg B, Champion C. Atypical antipsychotics: Managing adverse effects. *Perspect Psychiatr Care.* 2022;58(2):691-695. doi: 10.1111/ppc.12837.
5. Meli M, Rauber-Lüthy C, Hoffmann-Walbeck P, et al. Atypical antipsychotic poisoning in children: a multicentre analysis of poisons centers data. *Eur J Pediatr.* 2014;173(6):743-50. doi: 10.1007/s00431-013-2241-y.
6. Hui WF, Hon KL, Leung AKC. An Overview of the Pediatric Toxidromes and Poisoning Management. *Curr Rev Clin Exp Pharmacol.* 2021;16(4):318-329. doi: 10.2174/1574884715666201201090210.
7. Lofton AL, Klein-Schwartz W. Atypical experience: a case series of pediatric aripiprazole exposures. *Clin Toxicol (Phila).* 2005;43(3):151-3. PMID: 15902787.
8. Fasano CJ, O'Malley GF, Lares C, Rowden AK. Pediatric ziprasidone overdose. *Pediatr Emerg Care.* 2009;25(4):258-9. doi: 10.1097/PEC.0b013e31819e3775.
9. Vial T, Patat AM, Paret N, et al. Risperidone medication errors in children: an analysis of French poison centres data. *Clin Toxicol (Phila).* 2019;57(5):362-367. doi: 10.1080/15563650.2018.1523424.

https://doi.org/10.5455/annalsmedres.2025.10.304

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Extensive involvement of multiple cranial nerves in relapsed multiple myeloma

Mehmet Fatih Erbay [a,*], Ilhami Berber [b]

[a] İnönü University, Faculty of Medicine, Department of Radiology, Malatya, Türkiye
[b] İnönü University, Faculty of Medicine, Department of Internal Medicine, Division of Hematology, Malatya, Türkiye
*Corresponding author: drfatiherbay@hotmail.com (Mehmet Fatih Erbay)

Cite this article as: Erbay MF, Berber I. Extensive involvement of multiple cranial nerves in relapsed multiple myeloma. *Ann Med Res*. 2026;33(5):243–243. doi: 10.5455/annalsmedres.2026.02.063.

Received: Feb 17, 2026 **Accepted:** Feb 24, 2026 **Available Online:** May 22, 2026

A 36-year-old man with relapsed IgG-kappa multiple myeloma presented with bilateral lower extremity weakness and ptosis. He had previously undergone two autologous hematopoietic stem cell transplantations following multiple chemotherapy regimens. Contrast-enhanced T1-weighted brain MRI demonstrated linear thickening and enhancement of bilateral oculomotor (III), trochlear (IV), trigeminal (V), facial–vestibulocochlear (VII–VIII), glossopharyngeal–vagal (IX–X), and the right abducens (VI) nerve at the level of the brainstem. The enhancement pattern was predominantly linear without nodular mass formation or foraminal widening. No diffusion restriction or abnormal FLAIR signal was observed. Spinal MRI revealed nodular leptomeningeal involvement of the conus medullaris and cauda equina nerve roots (not shown here). Consistent with laboratory findings and systemic relapse, the imaging features were attributed to progression of multiple myeloma in the central nervous system. The linear cranial nerve enhancement, together with nodular involvement of the cauda equina, was radiologically consistent with leptomeningeal dissemination. Cranial nerve involvement in multiple myeloma is rare and typically associated with leptomeningeal spread [1, 2]. Differential diagnoses include leptomeningeal metastasis, infectious neuritis, and inflammatory cranial neuropathies.

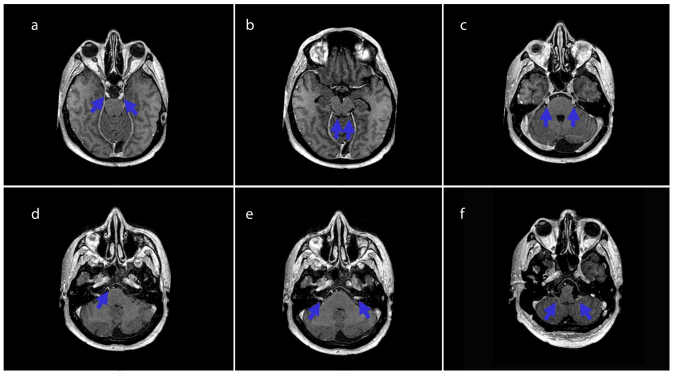

Figure 1. Contrast-enhanced thin-slice (1 mm) axial T1-weighted images demonstrate linear enhancement of bilateral oculomotor (III) (a), trochlear (IV) (b), trigeminal (V) (c), right abducens (VI) (d), bilateral facial--vestibulocochlear (VII--VIII) (e), and glossopharyngeal--vagal (IX--X) (f) nerves (arrows).

Informed Consent: The manuscript contains only fully anonymized radiological images and non-identifiable clinical information; therefore, written informed consent was not required.

Conflict of Interest: The authors declare that they have no conflict of interest.

Author Contributions: Conception: M.F.E; Design: M.F.E; Supervision: İ.B; Materials: İ.B; Analysis and/or Interpretation: M.F.E; Literature Review: M.F.E; Writing: M.F.E; Critical Review: İ.B.

Financial Disclosure: No financial support was received for this work.

REFERENCES

1. Pak N, Shakki Katouli F, Radmard AR, Shaki Katuli MH, Rezwanifar MM, Sepehri Boroujeni N. Multiple Cranial Nerve Palsy Concomitant with Leptomeningeal Involvement in Multiple Myeloma: A Case Report and Review of Literature. *Int J Hematol Oncol Stem Cell Res.* 2018;12(1):8-13. PMID: 29951172.
2. Thiruvengadam SS, Prayson RA. Multiple myeloma presenting with unilateral abducens and trigeminal nerve palsies. *J Clin Neurosci.* 2016;26:143-4. doi: 10.1016/j.jocn.2015.10.015.

Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

ERRATUM

In the article by *Keser S, et al.,* entitled *"An examination of the ocular surface parameters of patients with non-exophthalmic Graves disease"* published in Ann Med Res. 2023;30(5):607--610, the authors informed us that the year "2022" had been inadvertently included in the description of the study period in the Abstract and Materials and Methods sections.

Following editorial evaluation, it was deemed necessary to amend the study period to 2023 to ensure clarity and consistency throughout the manuscript.

This correction does not affect the results, statistical analyses, or conclusions of the study. However, we believe that this minor revision improves the overall clarity and consistency of the manuscript.

The relevant sections have been corrected accordingly. You can access the updated article

PDF via the following link:

https://annalsmedres.org/index.php/aomr/article/view/4438

https://doi.org/10.5455/annalsmedres.1970.01.05