

BJAN

Brazilian Journal of
Anesthesiology

Contact us

✉ editor.bjan@sbahq.org 🌐 www.bjan-sba.org 📞 +55 21 979 770 024

BJAN



BJAN
Brazilian Journal of
Anesthesiology

Brazilian Journal of Anesthesiology

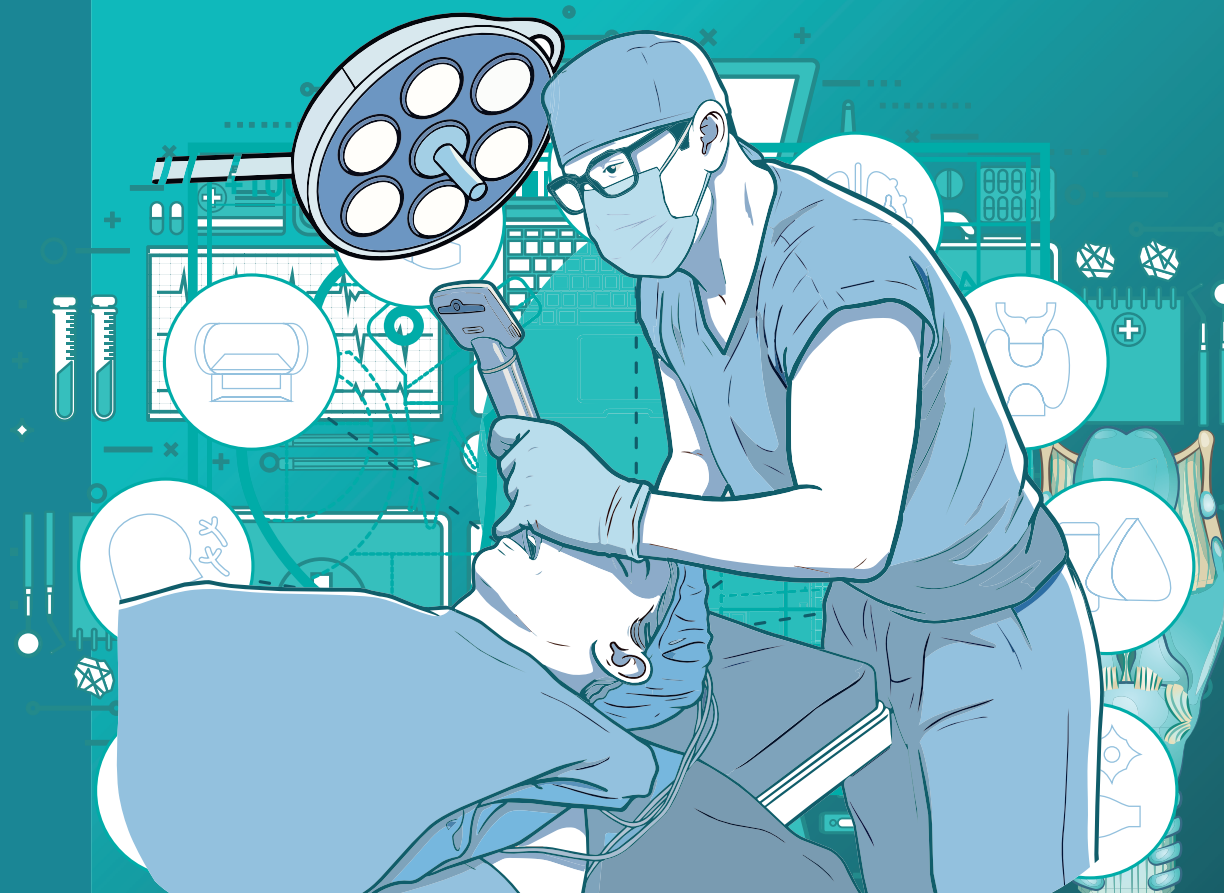
Vol. 76 N. 4 | July/August 2026

Vol. 76, N. 4 | Jul-Ago 2026 | ISSN 0104-0014

bjan-sba.org

Brazilian Journal of ANESTHESIOLOGY

Airway Management:
why competence remains
the key technology



SciELO

SBA
Sociedade
Brasileira de
Anestesiologia

Brazilian Journal of ANESTHESIOLOGY



Sociedade Brasileira
de Anestesiologia

Editor-in-Chief

Liana Maria Torres de Araújo Azi - Departamento de Anestesiologia e Cirurgia da Faculdade de Medicina da Bahia - Universidade Federal da Bahia - Salvador, BA, Brazil

Co-Editor

Rodrigo Leal Alves - Hospital São Rafael, Salvador, BA, Brazil

Associate Editors

Ana Maria Menezes Caetano - Universidade Federal de Pernambuco, Recife, PE, Brazil
André Prato Schmidt - Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
Célio Gomes de Amorim - Universidade Federal de Uberlândia, Uberlândia, MG, Brazil
Claudia Marquez Simões - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Durval Campos Kraychete - Universidade Federal da Bahia, Salvador, BA, Brazil
Edmundo Pereira de Souza Neto - Clinique du Pont de Chaume - ELSAN, Montauban, Tarn-et-Garonne, France
Eric Benedet Lineburger - Hospital São José, Criciúma, SC, Brazil
Florentino Fernandes Mendes - Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brazil
Gabriel Magalhães Nunes Guimarães - Universidade de Brasília, Brasília, DF, Brazil
João Manoel da Silva Júnior - Hospital do Servidor Público, São Paulo, SP, Brazil
João Helena Navarro e Lima - University of Manitoba, Manitoba, Winnipeg, Canada
Ligia Andrade da S. Telles Mathias - Irmandade da Santa Casa de Misericórdia de São Paulo, São Paulo, SP, Brazil
Lorena Ibiapina Mendes de Carvalho - Hospital Getúlio Vargas, Teresina, PI, Brazil
Luiz Vicente Garcia - Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Ribeirão Preto, SP, Brazil
Luiz Guilherme Villares da Costa - Hospital Israelita Albert Einstein, São Paulo, SP, Brazil
Luiz Marcelo Sá Malbouisson - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Luiz Marciano Cangiani - Hospital da Fundação Centro Médico Campinas, Campinas, SP, Brazil
Marcello Fonseca Salgado-Filho - Universidade Federal Fluminense, Niterói, RJ, Brazil
Maria José Carvalho Carmona - Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Mariana Fontes Lima Neville - Universidade Federal de São Paulo, São Paulo, SP, Brazil
Neuber Martins Fonseca - Faculdade de Medicina da Universidade Federal de Uberlândia, Uberlândia, MG, Brazil
Norma Sueli Pinheiro Módolo - Faculdade de Medicina de Botucatu da Universidade Estadual Paulista, Botucatu, SP, Brazil
Oscar César Pires - Universidade de Taubaté, Taubaté, SP, Brazil
Paulo Correia da Silva Neto - Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
Paulo do Nascimento Junior - Faculdade de Medicina de Botucatu da Universidade Estadual Paulista, Botucatu, SP, Brazil
Priscilla Ferreira Neto Cardoso - Centre Hospitalier du Haut Anjou (CHHA), Pays de Loire, France
Ricardo Vieira Carlos - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Rodrigo Moreira e Lima - University of Manitoba, Manitoba, Winnipeg, Canada
Vanessa Henriques Carvalho - Universidade Estadual de Campinas, Campinas, SP, Brazil
Vinicius Caldeira Quintão - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil

Editorial Committee

Achal Dhir - London Health Sciences Centre, Western University, London, ON, Canada
Adrian Alvarez - Hospital Italiano de Buenos Aires, Buenos Aires, BA, Argentina
Adrian Gelb - University of California, San Francisco, CA, USA
Alexandra Rezende Assad - Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
Alexis F. Turgeon - Université Laval, Québec, QC, Canada
Antônio Carlos Aguiar Brandão - Universidade do Vale do Sapucaí, Pouso Alegre, MG, Brazil
Ary Serpa Neto - Monash University, Melbourne, Australia
Bernd W. Böttiger - University Hospital of Cologne, Klinikum Köln, NW, Germany
Bobbie Jean Sweitzer - Northwestern Medicine, Chicago, IL, USA
Carlos Galhardo Júnior - Instituto Nacional de Cardiologia (INC/MS), Rio de Janeiro, RJ, Brazil
Carlos Manuel Correia Rodrigues de Almeida - Hospital CUF Viseu, Viseu, Beira Alta, Portugal
Catia Sousa Goveia - Universidade de Brasília, Brasília, DF, Brazil
Clarita Bandeira Margarido - Sunnybrook Health Sciences Centre, Toronto, ON, Canada
Claudia Regina Fernandes - Universidade Federal do Ceará, Fortaleza, CE, Brazil
Clovis Tadeu Bevilacqua Filho - Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil
Clyde Matava - The Hospital for Sick Children, Toronto, ON, Canada
Cyril David Mazer - St. Michael's Hospital, Toronto, ON, Canada
dade de São Paulo, São Paulo, SP, Brazil
Daniel Cordovani - McMaster University, Hamilton, Canada
David Perez - Universidade Federal de São Paulo, São Paulo, SP, Brazil
Deborah Culley - Harvard University, Boston, MA, USA
Deepak K. Tempe - GB Pant Institute of Postgraduate Medical Education and Research, New Delhi, India
Eduardo Quarenghi - Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, MI, Italy
Eduarda Schütz Martinelli - Toronto General Hospital, Toronto, ON, Canada
Eduardo Giroud Joaquim - Universidade Federal de São Paulo, São Paulo, SP, Brazil
Eliane Cristina de Souza Soares - Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
Emery Brown - Massachusetts Institute of Technology, Cambridge, MA, USA
Enrique A. Goytizolo - Hospital for Special Surgery (HSS), New York, NY, USA
Fabiana A. Penachi Bosco Ferreira - Universidade Federal de Goiás, Goiânia, GO, Brazil
Fábio Papa - University of Toronto, Toronto, ON, Canada
Federico Bilotta - Sapienza Università Di Roma, Roma, RM, Italy
Felipe Chiodini - Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Fernanda Bono Fukushima - Universidade Estadual Paulista, Botucatu, SP, Brazil
Fernando Abelha - Hospital de São João, Porto, Portugal
Filipe Nadir Caparica Santos - University of Toronto, Toronto, ON, Canada
Francisco A. Lobo - Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates
Frederic Michard - MiCo, Consulting and Research, Denens, Switzerland

Getúlio Rodrigues de Oliveira Filho - Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil
Gildásio de Oliveira Júnior - Alpert Medical School, Brown University, Providence, RI, USA
Giovanni Landoni - Vita-Salute San Raffaele University, Milano, LOM, Italy
Glauber Gouvea - Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
Glenio Bitencourt Mizubuti - Queen's University, Kingston, Canada
Gregory Hare - University of Toronto, Toronto, ON, Canada
Guilherme Antonio Moreira de Barros - Faculdade de Medicina de Botucatu da Universidade Estadual Paulista, Botucatu, SP, Brazil
Hazem Adel Ashmawi - Universidade de São Paulo, São Paulo, SP, Brazil
Ismar Lima Cavalcanti - Universidade Federal Fluminense, Niterói, RJ, Brazil
Jean-Jacques Rouby - Pierre and Marie Curie University, Paris, France
Jean-Louis Teboul - Paris-Sud University, Paris, France
Jean-Louis Vincent - Université Libre De Bruxelles, Bruxelles, Belgium
Jean-Michel Constantin - La Pitié Salpêtrière Hospital, University Paris-Sorbonne, Paris, France
Joana Berger-Estilita - University of Bern, Bern, Switzerland
João Batista Santos Garcia - Universidade Federal do Maranhão, São Luís, MA, Brazil
João Paulo Jordão Pontes - Universidade Federal de Uberlândia, Uberlândia, MG, Brazil
José Carlos Rodrigues Nascimento - Hospital Geral de Fortaleza, Fortaleza, Ceará, Brazil
José Otávio Costa Auler Junior - Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Judymara Lauzi Gozzani - Universidade Federal de São Paulo, São Paulo, SP, Brazil
Kurt Ruetzler - Cleveland Clinic, Cleveland, OH, USA
Laszlo Vutskits - Geneva University Hospitals, Geneva, GE, Switzerland
Leandro Gobbo Braz - Faculdade de Medicina de Botucatu da Universidade Estadual Paulista, Botucatu, SP, Brazil
Leonardo Henrique Cunha Ferraro - Universidade Federal de São Paulo, São Paulo, SP, Brazil
Leopoldo Muniz da Silva - Faculdade de Medicina de Botucatu da Universidade Estadual Paulista, Botucatu, SP, Brazil
Luciana Paula Cadore Stefani - Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
Luciano Gattinoni - University of Göttingen, Göttingen, Germany
Luís Antonio dos Santos Diego - Universidade Federal Fluminense, Niterói, RJ, Brazil
Luiz Fernando dos Reis Falcão - Universidade Federal de São Paulo, São Paulo, SP, Brazil
Marcelo Gama de Abreu - Cleveland Clinic, Cleveland, OH, USA
Márcio Matsumoto - Hospital Sirio Libanes, São Paulo, SP, Brazil
Marcos Antônio Costa de Albuquerque - Universidade Federal de Sergipe, São Cristóvão, SE, Brazil
Marcos Francisco Vidal Melo - Columbia University, New York, NY, USA
Maria Angela Tardelli - Universidade Federal de São Paulo, São Paulo, SP, Brazil
Mário J. da Conceição - Fundação Universidade Regional de Blumenau, Blumenau, SC, Brazil
Massimiliano Sorbello - AOU Policlinico Vittorio Emanuele, Catania, SIC, Italy
Matheus Fachini Vane - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Mônica Maria Siallys - Hospital e Maternidade Santa Joana, São Paulo, SP, Brazil
Nádia Maria da Conceição Duarte - Universidade Federal de Pernambuco, Recife, PE, Brazil
Nicola Disma - Istituto Giannina Gaslini, Genova, GE, Italy
Oscar César Pires - Universidade de Taubaté, Taubaté, SP, Brazil
Paulo Adilson Herrera - Hospital Evangélico de Londrina, Londrina, PR, Brazil
Paulo Alípio - Universidade Federal Fluminense, Niterói, RJ, Brazil
Pedro Amorim - Centro Hospitalar e Universitário do Porto, Porto, Portugal
Pedro Francisco Brandão - Universidade Federal do Espírito Santo, Vitória, ES, Brazil
Peter D. Slinger - University of Toronto, Toronto, ON, Canada
Philip Peng - University of Toronto, Toronto, ON, Canada
Plínio da Cunha Leal - Hospital São Domingos, São Luís, MA, Brazil
Raffael Pereira Cezar Zamper - London Health Science Center, London, UK
Rajinder K. Mirakhur - Royal Hospital, Belfast, Northern Ireland, UK
Ricardo Antônio Guimarães Barbosa - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil
Ricardo Esper Tremil - Department of Anesthesiology, Perioperative and Pain Medicine, Stanford Health Care, CA, USA
Roberto Fumagalli - Università degli studi Milano Bicocca, Milano, MI, Italy
Rodrigo Daly Guris - Children's Hospital of Philadelphia, Philadelphia, PA, USA
Rogean Rodrigues Nunes - Centro de Ensino e Treinamento/SBA (Residência Médica) em Anestesiologia do Hospital Geral de Fortaleza, Fortaleza, CE, Brazil
Sara Lucia Ferreira Cavalcante - Hospital Geral do Inamps de Fortaleza, Fortaleza, CE, Brazil
Suzana Margareth Lobo - Faculdade de Medicina de São José do Rio Preto (FAMERP), São José do Rio Preto, SP, Brazil
Thais Cancado - Serviço de Anestesiologia de Campo Grande, Campo Grande, MS, Brazil
Thomas Engelhardt - Montreal Children's Hospital, McGill University, Montreal, Canada
Tom G. Hansen - Akershus University Hospital, Oslo, Norway
Vicente Faraon Fonseca - Unisinos, Porto Alegre, RS, Brazil
Waynice Paula-Garcia - Universidade de São Paulo, São Paulo, SP, Brazil
Wesley Pfeiffer - University of Toronto, Toronto, ON, Canada
Wolnei Caumo - Universidade do Rio Grande do Sul, Porto Alegre, RS, Brazil

Previous Editors-in-Chief

Oscar Vasconcellos Ribeiro (1951-1957)
Zairo Eira Garcia Vieira (1958-1964)
Bento Mário Villamil Gonçalves (1965-1979)
Masami Katayama (1980-1988)
Antônio Leite Oliva Filho (1989-1994)
Luiz Marciano Cangiani (1995-2003)
Judymara Lauzi Gozzani (2004-2009)
Mario José da Conceição (2010-2015)
Maria Angela Tardelli (2016-2018)
Maria José Carvalho Carmona (2019-2021)
André Prato Schmidt (2022-2024)

Editorial Office

Editorial Coordinator: Gregory Neres
Marketing Analyst: Thadeu de Moraes Carvalho
Librarian: Marcelly Ramos

The Brazilian Journal of Anesthesiology (BJAN) is the official Journal of the Sociedade Brasileira de Anestesiologia (SBA). The BJAN only accepts original articles for publication submitted and published in English. Before submitting a manuscript, authors must read carefully the Instructions to Authors. It can be found at: <<https://bjan-sba.org/instructions>>. Manuscripts must be submitted electronically via the Journal's online submission system <<http://www.editorialmanager.com/bjan>>.

The BJAN publishes original work in all areas of anesthesia, surgical critical care, perioperative medicine and pain medicine, including basic, translational and clinical research, as well as education and technological innovation. In addition, the Journal publishes review articles, relevant case reports, pictorial essays or contextualized images, special articles, correspondence, and letters to the editor. Special articles such as guidelines and historical manuscripts are published upon invitation only, and authors should seek subject approval by the Editorial Office before submission.

The BJAN accepts only original articles that are not under consideration by any other journal and that have not been published before, except as academic theses or abstracts presented at conferences or meetings. A cloud-based intuitive platform is used to compare submitted manuscripts to previous publications, and submissions must not contain any instances of plagiarism. Authors must obtain and send the Editorial Office all required permissions for any overlapping material and properly identify them in the manuscript to avoid plagiarism.

All articles submitted for publication are assessed by two or more members of the Editorial Board or external peer reviewers, assigned at the discretion of the Editor-in-chief or the Associate Editors. Published articles are a property of the Brazilian Society of Anesthesiology (SBA), and their total or partial reproduction can be made with previous authorization. The BJAN assumes no responsibility for the opinions expressed in the signed works.

Edited by | Editada por

Sociedade Brasileira de Anestesiologia (SBA)
Rua Prof. Alfredo Gomes, 36, Botafogo, Rio de Janeiro/RJ, Brazil - CEP 22251-080
Telephone: +55 21 3528-1050
E-mail: contato@sbahq.org
www.sbahq.org

Published by | Publicada por
Elsevier Editora Ltda.
Telephone RJ: +55 21 3970-9300
Telephone SP: +55 11 5105-8555
www.elsevier.com
ISSN: 0104-0014 © 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier Editora Ltda. All rights reserved.

Editorial

- 844781** Airway Management: why training, judgement, and competence remain the most important technologies

Daniel Perin, and Rodrigo Leal Alves

Original Investigations

- 844757** Comparative analysis of a 3D-printed videolaryngoscope versus the Macintosh model in adult airway management: a randomized clinical trial

Florentino Fernandes Mendes, Liége Caroline Immich, Thiago Fagundes Gross, Luis Fernando Marcelino Braga, Andreia Gomes Aires, Elizandra Braganhol, Ramon Magalhães Mendonça Vilela, Francine Ullrich Carrazzoni dos Reis, Gustavo Zordan Piva, Victor Hugo Reinoso Gomez, Vinicius Hofstätter Rodrigues, Afonso Reguly, Leandro de Freitas Spinelli, Simone Schneider Amaral, and Gisele Orlandi Introini

- 844760** Use, perceptions, and barriers to supraglottic airway devices in Brazil: a national cross-sectional survey with global implications

Jayme M. dos Santos Neto, Flávia J. de Barros, Victor M. Lemos, and Clístenes Cristian de Carvalho

- 844761** Clinical impact of extubation in the operating room after cardiac surgery: a retrospective analysis of a prospective registry

Patricia Silva de Marco, Pedro Henrique Romani Lauand, Luiz Carlos Volpi Júnior, Raphael Fernandes Cardoso Vianna Alves, Maria Paula Belini, Lilia Nigro Maia, Maurício Nassau Machado, and Marcelo Arruda Nakazone

- 844772** Effects of propofol versus sevoflurane induction on echocardiographic parameters in patients with mitral stenosis: a randomized clinical trial

Mohammed Jaffer Sherif, Saravana Babu MS, Prasanta Kumar Dash, Nagarjuna Panidappu, Subin Sukesan, and Thomas Koshy

- 844769** Effects of shoulder rotation and table tilt maneuvers on the lumbar and thoracic paramedian acoustic window in older adults: a quasi-experimental study

Marcos V. Nunes de Souza, Ana M.M. Caetano, Nádia M.C. Duarte, Camila T. Nunes de Souza, Artur J.L. Bezerra, and Emmanuelle T.A.G.B.B. Silva

- 844770** Continuous erector spinae plane block for robotic-assisted thoracic surgery in lung cancer: a dual-center retrospective propensity-score weighted cohort study

Matheus Arraes, Sara Amaral, Saullo Queiroz Silveira, Leopoldo Muniz da Silva, Rafael Sousa Fava Nersessian, Francisco Jose Lucena Bezerra, Manoel de Souza Neto, Caroline Machado Nunes, Helidea de Oliveira Lima, Fernando Nardy Bellicieri, Glenio B. Mizubuti, and Rafael Lombardi

-
- 844765** **Coronary artery bypass grafting in Jehovah's Witness patients: a retrospective propensity matched cohort study**
Henrique Ryu Yamanaka Nakano, Guilherme Mota Filho, Bruno Camargo Calabria Araújo, Heloísa Zogheib, Enzo Camargo Calabria Araújo, Leonardo Cardoso Saraiva, Jessica Silva Nicolau, Suely Pereira Zeferino, and José Otávio Costa Auler Junior
-
- 844768** **Evaluation of two large language models for intensive care unit discharge decisions: a prospective observational cohort study**
Engin İhsan Turan, Abdurrahman Engin Baydemir, Ebru Kaya, Zehra Polat Turan, and Ayça Sultan Şahin
-
- 844771** **Prevalence and factors associated with burnout syndrome among anesthesiology residents in Rio Grande do Sul, Brazil: a multicenter cross-sectional study**
Airton Bagatini, Felipe de Sousa Junqueira, Eduarda Druck Magadan, Silvia Ramos Hecktheuer, Regis Borges Aquino, Florentino Fernandes Mendes, Eduardo Francisco Mafassoli Corrêa, Cleiton da Silva Pando, Daniela Benzano Bumaguin, and Cassiana Gil Prates
-
- 844786** **Preoperative angiotensin-converting enzyme inhibitor or angiotensin receptor blocker use and acute kidney injury after high-risk cardiac surgery: a multicenter prospective cohort study**
Iñigo Rubio, Angel Candela, Gemma Echarri, Raquel Callejas, Eduardo Tamayo, David Nagore, Jorge M. Núñez-Córdoba, and Marc Vives
-
- Review Articles**
-
- 844782** **Artificial intelligence in the prediction of intraoperative red blood cell transfusion in cardiac surgery: a systematic review and diagnostic test accuracy meta-analysis**
Filipe Giordano Valério, Amanda Carneiro Rodrigues, Davi Ricardo Soares Gama de Amorim, Roberta Esterque Cantarino, Glauco Martins de Araujo, Jimmy Yusuf, Leticia Vianna Leal Soares Bessa, Millena Mendonça Andrade Paes Leme, Marcella Freire de Campos Euzebio, Flávio Luiz Seixas, Alexandra Rezende Assad, and Luis Antonio dos Santos Diego
-
- 844766** **Programmed intermittent epidural bolus versus continuous epidural infusion combined with patient-controlled epidural analgesia for postoperative analgesia: a meta-analysis of randomized controlled trials**
Gabriel Lemos González, Vinícius Fernandes Sarmento, Bruno Francisco Minetto Wegner, Gustavo Roberto Minetto Wegner, Alesson Marinho Miranda, Iaci Luisa Lopes de Mattos, and Tatiana Souza do Nascimento
-
- Letters to the Editor**
-
- 844767** **Erector spinae plane block vs. paravertebral block for postoperative analgesia in breast surgery: a meta-analysis of randomized trials**
Burhan Dost, Yunus Emre Karapinar, Muzeyyen Beldagli, Esra Turunc, Cengiz Kaya, and Alessandro De Cassai
-
- 844783** **Psychotropic medication use during preanesthetic evaluation at a Brazilian university hospital: a retrospective cross-sectional analysis**
Matheus Salera Gordim, Vitor Yan Bezerra de Araújo, Danilo de Oliveira da Câmara Filho, Luzinelson Muniz da Silva Júnior, Emerson Arcoverde Nunes, Heleno Paiva Oliveira, and Wallace Andrino da Silva
-
- 844763** **Preoperative fasting and GLP-1 receptor agonists: gastric retention without increased aspiration risk**
Filomena Di Vezza, Marco Sanvitti, and Federico Bilotta
-
- 844759** **Erector spinae plane block versus caudal block in children: similar analgesia, different stories beneath the surface**
Tuhin Mistry, and Abhijit Sukumaran Nair
-
- 844784** **Response to Letter to the Editor: "Erector spinae plane block versus caudal block in children: similar analgesia, different stories beneath the surface"**
Barbara Bombassaro Masiero, and Rafael Arsky Lombardi
-

EDITORIAL

Airway Management: why training, judgement, and competence remain the most important technologies



Airway management has undergone a remarkable transformation over the past two decades. Better data on safety and efficacy of different techniques and treatment algorithms, the expansion of video laryngoscopy, the evolution of second-generation supraglottic airway devices (SADs), the emergence of point-of-care ultrasound and the development of low-cost technological innovations have substantially increased the number of tools available to anesthesiologists. Yet, despite these advances, the determinants of successful airway management remain fundamentally unchanged: knowledge, training, judgment, and experience.^{1,2}

In this issue of the Brazilian Journal of Anesthesiology, three complementary studies provide important insights into contemporary airway practice.³⁻⁵ Together, they highlight a central message for the specialty: technological innovation alone is insufficient unless accompanied by structured airway education, clinical judgment, and sustained training.

The randomized trial by Mendes et al.³ represents an important step toward democratizing access to video laryngoscopy. The development of a reproducible, low-cost 3D-printed videolaryngoscope addresses a critical global need. The study was performed in adult patients scheduled for surgical procedures under general anesthesia and involved anesthesiologists. Although the device provided superior glottic visualization compared with conventional Macintosh laryngoscopy, improved visualization did not translate into easier or faster intubation. Despite all limitations of the study, it emphasizes the importance of device-specific training and airway expertise.

This apparent paradox is well recognized in airway management. Even though there is a robust set of evidence that video laryngoscopy offers improved visualization of vocal cords when compared with conventional direct laryngoscopy, obtaining an excellent view of the glottis does not necessarily guarantee successful tube delivery.^{6,7} Device-specific psychomotor skills, hand-eye coordination, stylet manipulation, and familiarity with indirect visualization techniques remain essential competencies. Another layer of

discussion in that matter involves the scarcity of studies comparing 3D-printed models with commercially available products with little evidence favoring the latter.^{8,9}

The notion that SADs have an important role not only in anesthesia but also in emergency and resuscitation scenarios is not new and has been around now for over three decades.^{10,11} Concerning that subject, Santos Neto et al.⁴ conducted a national survey that reinforced the relevance of training and skill. Although Brazilian anesthesiologists overwhelmingly perceived supraglottic airway devices as safe and effective, actual utilization remained modest. Exposure during residency emerged as one of the strongest predictors of current clinical use, while lack of operator experience was identified as the most important barrier to broader adoption.

The importance of comprehensive airway training is further underscored by the retrospective cohort study by Azem and colleagues.⁵ Examining patients undergoing surgical drainage of severe odontogenic infections, the authors demonstrated that Awake Fiberoptic Intubation was preferentially selected for anatomically complex cases characterized by trismus and overweight patients with elevated body mass index. Importantly, the study illustrates that successful airway management in high-risk patients depends not only on access to advanced devices but also on the ability of clinicians to recognize predictors of difficulty, select the most appropriate airway strategy, and execute advanced techniques safely. The finding that postoperative spontaneous ventilation was associated with earlier discharge further highlights how airway management decisions can influence the broader perioperative trajectory.

Together with the studies on video laryngoscopy and supraglottic airway devices, the collected data emphasize that airway excellence is built upon a hierarchy of competencies ranging from routine device use to the management of the most challenging anticipated difficult airways. Recently, international literature has further strengthened this perspective. Contemporary discussions on airway education have argued that modern airway management should be

viewed as a competency continuum rather than a collection of isolated technical skills. The focus is shifting from simply mastering devices toward developing integrated airway expertise, including decision-making, team communication, crisis resource management, and adaptability in rapidly evolving clinical situations.¹

Collectively, the studies presented in this issue illustrate two sides of the same coin. Innovation is expanding access to airway technologies, yet even widely accepted technologies remain underutilized when training opportunities are insufficient. The challenge facing anesthesiology is therefore not merely technological dissemination but educational implementation.

For Brazil and other resource-variable healthcare systems, these findings are particularly relevant. Investments in airway education, simulation programs, residency curricula, and continuing professional development may ultimately have a greater impact on patient outcomes than the acquisition of additional devices alone. The future of airway management will undoubtedly include increasingly sophisticated technologies. However, the most important innovation may not be a new device but rather the development of robust educational ecosystems capable of transforming technology into clinical competence.

As the studies by Mendes et al., Santos Neto et al., and Azem et al. demonstrate airway safety is achieved not when clinicians have access to more devices, but when they are trained to use the right device, at the right time, for the right patient. In airway management, training, judgment, and competence remain the most important technologies.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability statement

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Editor

Liana Azi

References

1. Ahmad I, El-Boghdady K, Iliff H, et al. Difficult Airway Society 2025 guidelines for management of unanticipated difficult tracheal intubation in adults. *Br J Anaesth.* 2026;136:283–307.
2. Apfelbaum JL, Hagberg CA, Connis RT, et al. 2022 American Society of Anesthesiologists Practice Guidelines for Management of the Difficult Airway*. *Anesthesiology.* 2022;136:31–81.
3. Mendes FF, Immich LC, Gross TF, et al. Comparative analysis of a 3D-printed videolaryngoscope versus the Macintosh model in adult airway management: a randomized clinical trial. *Braz J Anesthesiol.* 2026;76:844757.
4. dos Santos Neto JM, de Barros FJ, Lemos VM, et al. Use, perceptions, and barriers to supraglottic airway devices in Brazil: a national cross-sectional survey with global implications. *Braz J Anesthesiol.* 2026;76:844760.
5. Azem K, Kashkush A, Zribi B, et al. Airway management and outcomes in surgical drainage of severe odontogenic infections: a retrospective cohort study. *Braz J Anesthesiol.* 2026;76:844756.
6. Hansel J, Rogers AM, Lewis SR, et al. Videolaryngoscopy versus direct laryngoscopy for adults undergoing tracheal intubation: a Cochrane systematic review and meta-analysis update. *Br J Anaesth.* 2022;129:612–23.
7. Meitzen SE, Benumof JL. Video Laryngoscopy: Positives, Negatives, and Defining the Difficult Intubation. *Anesth Analg.* 2019;128:399–401.
8. Detoni PB, Nascimento JS, Araújo Azi LMT, et al. McGrath MAC Versus Three-Dimensional Printed Video Laryngoscopes: A Randomized, Manikin-Simulated Noninferiority Controlled Study with Medical Students. *Anesth Analg.* 2025;140:334–41.
9. Ataman AK, Altıntaş E. Comparison of a commercial 3D fabricated laryngoscope (Airangel®) with a widely-used video laryngoscope (Glidescope®): Randomized controlled cross-over study. *Trends Anaesth Crit Care.* 2021;40:35–40.
10. Verghese C, Brimacombe JR. Survey of Laryngeal Mask Airway Usage in 11,910 Patients. *Anesth Analg.* 1996;82:129–33.
11. Dickinson M, Curry P. Training for the use of the laryngeal mask in emergency and resuscitation situations. *Resuscitation.* 1994;28:111–3.

Daniel Perin ^a, Rodrigo Leal Alves ^{b,c,d,*}

^a Hospital Israelita Albert Einstein, São Paulo, SP, Brazil

^b Universidade Federal da Bahia (UFBA), Salvador, BA, Brazil

^c Universidade Estadual Paulista (UNESP), São Paulo, SP, Brazil

^d Hospital São Rafael, Salvador, BA, Brazil

* Corresponding author.

E-mail: rlalves@ufba.br (R.L. Alves).

ORIGINAL INVESTIGATION

Comparative analysis of a 3D-printed videolaryngoscope versus the Macintosh model in adult airway management: a randomized clinical trial



Florentino Fernandes Mendes ^{a,b,c,d,*}, Liège Caroline Immich ^{a,b,c},
Thiago Fagundes Gross ^a, Luis Fernando Marcelino Braga ^a, Andreia Gomes Aires ^{a,e},
Elizandra Braganhol ^{a,e}, Ramon Magalhães Mendonça Vilela ^c,
Francine Ullrich Carrazzoni dos Reis ^c, Gustavo Zordan Piva ^c,
Victor Hugo Reinoso Gomez ^c, Vinicius Hofstätter Rodrigues ^c, Afonso Reguly ^f,
Leandro de Freitas Spinelli ^{a,d,e}, Simone Schneider Amaral ^{a,d},
Gisele Orlandi Introíni ^{a,d}

^a Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

^b Programa de Pós-Graduação em Tecnologia da Informação e Gestão em Saúde, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

^c Santa Casa de Porto Alegre, Porto Alegre, RS, Brazil

^d Laboratório de Inovação, Prototipagem, Educação Criativa e Inclusiva (LIPECIN), Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

^e Programa de Pós-Graduação em Ciências da Reabilitação, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

^f Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

Received 26 August 2025; accepted 5 April 2026

Available online 22 April 2026

KEYWORDS

3D Printing;
Additive
Manufacturing;
Airway Management;
Intratracheal;
Intubation

Abstract

Background and objectives: Video Laryngoscopy (VL) provides enhanced glottic visualization in airway management, but its high cost limits adoption. Three-Dimensional (3D) printing offers a cost-effective alternative for producing VL devices. This randomized trial compared a 3D-printed Videolaryngoscope (VL3D) with the standard Macintosh Laryngoscope (MAC) during tracheal intubation in adults.

Methods: This was a randomized, controlled, and assessor-blinded clinical trial. The primary outcome was intubation difficulty, assessed using the Intubation Difficulty Scale (IDS). Secondary outcomes included glottic visualization (Cormack-Lehane grade) and Intubation Time (IT). The final analysis followed a modified intention-to-treat principle.

* Corresponding author.

E-mail: florentinomendes@gmail.com (F.F. Mendes).

Results: This study included 52 patients scheduled to undergo general anesthesia with tracheal intubation. Baseline characteristics were comparable between groups. The analysis included 49 patients (VL3D, $n = 26$; MAC, $n = 23$). Categorical analysis of the primary outcome showed a higher incidence of moderate-to-severe intubation difficulty in the VL3D group (23.1% vs. 0%, $p = 0.032$), although median IDS scores were statistically similar. In relation to secondary outcomes, the VL3D provided significantly superior glottic visualization (Cormack-Lehane grade I: 72.0% in the VL3D group vs. 33.3% in the MAC group, $p = 0.002$). Conversely, the VL3D was associated with a significantly longer intubation time (median 70 seconds vs. 40 seconds, $p = 0.005$).

Conclusions: The VL3D demonstrated superior glottic visualization but correlated with significantly longer intubation duration and greater procedural difficulty. These findings suggest that additional specialized training or design optimization is necessary before widespread clinical implementation.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Evidence from randomized trials and systematic reviews indicates that video-assisted techniques yield better glottic exposure and improved first-pass success rates than direct laryngoscopy.¹⁻³ However, despite these visual benefits, no statistically significant differences have been consistently observed in terms of intubation duration or the incidence of airway-related injuries, including trauma to the lips, gums, teeth, or pharyngeal structures. While videolaryngoscopes optimize management in anatomically complex cases, their implementation is often hindered by high procurement costs and restricted availability in low-resource environments.

A promising alternative to conventional industrial manufacturing is the use of Industry 4.0 technologies, particularly Computer Numerical Control (CNC) systems and additive manufacturing (3D printing). These tools allow for the low-cost, small-scale production of customized medical devices. One study employed 3D printing to fabricate a laryngoscope blade compatible with a borescope, which served as both the light source and video feed, highlighting the potential for low-cost, anatomically tailored blade designs.⁴ Another innovation combined a Universal Serial Bus (USB) endoscopic camera with a laryngoscope, utilizing a standard personal computer for real-time image display.⁵

Consistent with this approach, our research group developed a customized, 3D-printed Videolaryngoscope (VL3D) for adult and pediatric use. The device's functional reliability and usability were previously validated in preclinical settings. Mechanical testing demonstrated adequate structural integrity, while simulation-based evaluations using airway mannequins confirmed ergonomic handling and optimal glottic visualization. Furthermore, the video output was consistently rated as high-quality by independent evaluators.⁶

Building upon these promising preclinical and simulation-based findings, this randomized clinical trial was designed to assess the performance of the VL3D device in comparison with the conventional Macintosh laryngoscope. The primary outcome was the level of difficulty encountered during intubation (measured by the Intubation Difficulty Scale). Secondary outcomes included glottic visualization (Cormack-Lehane grade), intubation time, procedural efficiency, and

the incidence of adverse events, including bleeding, dental trauma, mucosal injury, and lip damage.

Material and methods

Study location

This research was conducted at the Surgical Center of the Santa Casa de Porto Alegre (SCPA), in collaboration with the Outpatient Preoperative Evaluation Clinic (OPEC) and the Santa Casa Anesthesia Service (SCAS).

Ethical considerations

This study included 52 patients scheduled to undergo general anesthesia with tracheal intubation. All procedures were conducted in accordance with Resolution n°466/2012 of the Brazilian National Health Council and the ethical principles of the World Medical Association's Code of Ethics. The study protocol was approved by the Institutional Research Ethics Committee (CAAE: 58735522.8.0000.5335), and the trial was registered at the Brazilian Clinical Trials Registry (ReBEC) under number RBR-9t36r35 before the first patient enrollment. The manuscript adheres to the Consolidated Standards of Reporting Trials (CONSORT) guidelines, and written informed consent was obtained from all participants.

Trial design

This study was a randomized, controlled, and assessor-blinded clinical trial designed to compare the VL3D device with the Macintosh laryngoscope, which is widely regarded as the gold standard in airway management.

The randomization sequence was generated using the randomizeR package in R software by an independent investigator not involved in recruitment or data collection. To ensure allocation concealment, participants were assigned to study groups using sequentially numbered, opaque, sealed envelopes. These envelopes were opened by the attending anesthesiologist only after the participant had been formally enrolled and baseline data were obtained, thereby minimizing selection bias.

Inclusion criteria

Eligible participants were adult patients (≥ 18 years), scheduled for surgical procedures under general anesthesia requiring tracheal intubation. All were classified as ASA physical status I to IV and provided written informed consent.

Exclusion criteria

Patients were excluded from the study if they were scheduled for emergency surgical procedures; undergoing upper airway, head and neck, or orthognathic surgery; or required planned nasotracheal intubation. Additional exclusion criteria included a history of cervical spine surgery, the presence of cervical disc injuries or fractures, or the use of a cervical collar. Furthermore, individuals with a Body Mass Index (BMI) $\geq 35 \text{ kg.m}^{-2}$, those diagnosed with Type 1 Diabetes Mellitus, or patients with syndromic conditions or head and neck trauma that could interfere with standard intubation techniques were also ineligible.

Primary and secondary outcome measures

Primary outcome

The primary outcome of this study was the intubation difficulty, quantified by the Intubation Difficulty Scale (IDS).

Secondary outcomes

Secondary outcomes included glottic visualization, as assessed by the Cormack-Lehane (C-L) grade, Intubation Time (IT), procedural efficiency, and the incidence of adverse events: bleeding, dental trauma, mucosal injury, and lip damage.

Anesthetic management

Intraoperative monitoring included electrocardiography with leads DII and V5, pulse oximetry, capnography, Noninvasive Blood Pressure (NIBP), and neuromuscular blockade assessment using Train-of-Four (TOF) monitoring. Anesthetic management was standardized across participants. Fentanyl was administered at an initial bolus dose of $5 \mu\text{g.kg}^{-1}$, with supplemental doses provided as needed and recorded at the end of the procedure. Propofol was used for induction at an initial dose of 2 mg.kg^{-1} , titrated according to clinical requirements. The administration of inhalational anesthetic (sevoflurane) was documented as volume percent, while oxygen consumption was measured in liters per minute over the total period of use. Rocuronium was employed as the neuromuscular blocking agent at a dose of 0.6 mg.kg^{-1} (equivalent to $2 \times \text{ED}_{95}$), with additional boluses of one-fifth the induction dose administered at the anesthesiologist's discretion. Other adjuvant drugs used during induction and maintenance of anesthesia were also administered based on clinical judgment.

Airway management: Standardization of intubation technique

All patients were positioned on a 10 cm rigid cushion to optimally align the laryngeal and tracheal axes. Preoxygenation

with $\text{FiO}_2 > 80\%$ for ≥ 3 minutes preceded anesthetic induction. Intubation began once TOF = 0 (adductor pollicis), with a 15-second pause between stimuli.

Intubation time was recorded from TOF zero to tracheal sealing by a cuff inflated. Laryngoscopy was performed with a VL3D (model LIPECIN)⁶ or Macintosh laryngoscope (blade sizes 4–5), according to the randomized allocation. In the event of intubation difficulty, the attending anesthesiologist was permitted to switch to the alternative device, guided by the Intubation Difficulty Scale (IDS). The internal diameter of the endotracheal tube was chosen at the discretion of the anesthesiologist, and intubation was performed using a standardized, single-use tube introducer.

Intubator training with the VL3D

Prior to patient enrollment, all anesthesiologists responsible for airway management – each with 3 to 6 years of specialty experience – underwent structured training with the VL3D at the Institutional Simulation Center (UFCSPA). To ensure proficiency, each participant completed a minimum of 20 supervised intubations on simulation mannequins featuring varying airway difficulty levels.

Airway assessment and variable definitions

Airway assessment followed standardized protocols based on validated predictors of difficult intubation. The Modified Mallampati Test (MMT) was conducted with the patient seated, head in a neutral position, and mouth opened to its maximum. Scores ranged from Class I to IV, with Classes III and IV considered predictive of difficult intubation. Thyromental Distance (TMD) was measured from the mentum to the thyroid cartilage with the head fully extended; distances $\leq 6.0 \text{ cm}$ were classified as high risk. Head and neck mobility was evaluated by measuring the angle between the mandible and cervical spine during maximal extension. Dental protrusion was classified visually as absent, moderate ($0\text{--}0.5 \text{ cm}$), or severe ($> 0.5 \text{ cm}$). Interincisor distance was measured with the mouth maximally opened and recorded in centimeters.

Mandibular mobility and dental relationship were further assessed using the Upper Lip Bite Test (ULBT), in which the lower incisors' ability to contact the upper lip vermilion line was categorized as Class I (complete cover), Class II (partial cover), or Class III (unable to reach). To ensure consistency in measurement, all evaluators underwent specific training, and measuring instruments were calibrated. The Airway Assessment Score (AAS) was calculated by assigning numerical values to the results of the above tests, according to predefined criteria. A total score above 6 was considered predictive of a difficult airway (Table 1).⁷

Intubation difficulty assessment

The Intubation Difficulty Scale (IDS) was employed during the procedure to objectively quantify intubation difficulty. This assessment was based on seven parameters: the number of intubation attempts, the amount of operators involved, the use of alternative techniques, the Cormack-Lehane (C-L) grade, the application of increased lifting force, the use of external laryngeal pressure, and the position of the vocal

Table 1 Airway Assessment Score (AAS).

Variables	Points		
	0	1	2
MMP	Class I	Class II	Class III – IV
TMD (cm)	> 6.5	6 – 6.5	< 6.0
Head / Neck mobility (°)	> 90	90	< 90
BMI (kg.m ⁻²)	< 25	≥ 25	–
Protruded teeth	No	Moderate	Severe
Interincisor distance	> 5	4 – 5	< 4
ULBT	Class I	Class II	Class III

MMP, Modified Mallampati classification; TMD, Thyromental Distance; BMI, Body Mass Index; ULBT, Upper Lip Bite Test. Note: Scores above 6 points are considered predictive of difficult intubation.

cords. Scores were recorded immediately after intubation by the performing anesthesiologist. Total IDS scores were categorized as follows: 0 = Easy, 1–5 = Mild difficulty, and > 5 = Moderate to severe difficulty.

The Cormack-Lehane grade was documented during initial laryngoscopy without external laryngeal pressure. For analytical purposes, grades I and II were considered easy, and grades III and IV were classified as difficult. The study analyzed correlations between C-L grade, AAS, and IDS, with continuous and categorical analyses of IDS scores (Table 2).⁸

Table 2 Intubation Difficulty Scale (IDS).

Variables	Points
N1 – Each intubation attempt adds	1
N2 – Each additional operator adds	1
N3 – Each alternative technique adds	1
N4 – Apply Cormack-Lehane (C-L) grade for first attempt	0 – 1
N5 – Increased lifting force required for intubation add 1 point	0 – 1
N6 – External laryngeal pressure to improve C-L?	1
N7 – Position of the vocal cords during intubation	0 – 1

Notes: N3 – One point is added if any alternative technique is required, including patient repositioning, modification of equipment (e.g., blade, endotracheal tube, or the addition of a stylet), alteration of the intubation route (nasotracheal or orotracheal), or the use of devices such as fiberoptic scopes or laryngeal mask airways; N4 – A Grade I view on the first attempt corresponds to 0, while any deviation from this grade adds 1-point; N5 – Normal lifting force corresponds to 0. Increased force during laryngoscopy adds 1-point; N6 – Application of external laryngeal pressure to improve glottic visualization adds 1-point; N7 – Vocal cord position is scored as 0 for abduction and 1 for adduction. An IDS score of 0 indicates easy intubation; IDS 1–5 indicates mild difficulty; and IDS >5 indicates moderate-to-severe difficulty.

Biological safety and processing of devices

The VL3D devices underwent a standardized sterilization protocol. This process included immediate pre-cleaning, manual washing with a mild detergent and non-abrasive materials, drying using disposable cloths and compressed air, followed by inspection and packaging. Packaging utilized a cellulose-free barrier suitable for hydrogen peroxide processing. Sterilization was performed using hydrogen peroxide. Post-sterilization, the devices were stored in a clean, dry environment, shielded from light exposure and excessive handling. The validated shelf life of the sterilization process was one year. Furthermore, all units were subjected to microbiological testing conducted by the institutional laboratory.

Borescope cameras

Borescope cameras were subjected to a separate process: they were manually washed, disinfected with 70% alcohol, dried, and stored in sealed, labeled plastic bags under the same handling conditions as the VL3D units. The disinfection efficacy was considered valid for 30 days. In contrast, conventional Macintosh laryngoscopes were processed and sterilized in accordance with the established institutional protocol.

VL3D production and traceability

The VL3D device is a fully reproducible, low-cost solution (approx. \$12 USD), developed based on a well-established blade geometry, such as the C-MAC configuration. Its construction utilized conventional Fused Deposition Modeling (FDM) technology and incorporated a standard camera module, an LED light source, and a dedicated display unit.

The design initiative is licensed under the Creative Commons Attribution-Noncommercial-No Derivatives (CC BY-NC-ND) license, permitting non-commercial redistribution under the condition that the device is not altered or used for commercial gain. The complete design was previously published and validated by our research group.⁶ All requisite STL files, comprehensive assembly instructions, and the full Bill of Materials (BOM) are publicly accessible in the supplementary materials of the original publication, thereby ensuring high reproducibility across intrahospital makerspaces.

A total of 16 VL3D units were produced in the institutional prototyping laboratory. To ensure traceability and aseptic handling, each device was labeled with a unique identifier, stored in sealed packaging, and handled aseptically. A designated researcher, independent of the clinical phases of the study, was responsible for sterilization, biological safety monitoring, and reuse tracking.

Adverse event monitoring

The following adverse events were monitored and recorded as binary outcomes: bleeding, dental trauma, mucosal injury, and lip damage.

Statistical analysis

Data were entered into Microsoft Excel and analyzed using SPSS version 20.0. The final analysis followed a Modified Intention-To-Treat (mITT) principle. Although 52 patients were initially randomized, three patients in the MAC group were excluded from the final analysis due to post-randomization protocol deviations (failure to meet inclusion criteria discovered after allocation), resulting in a final sample of 49 patients (VL3D, $n = 26$; MAC, $n = 23$), as detailed in the CONSORT flow diagram. Categorical variables were summarized as frequencies and percentages and compared using the Chi-square test with Yates' correction or Fisher's exact test, as appropriate. Categorical analysis of the primary outcome (IDS) was also performed. Pairwise comparisons of proportions were conducted with the z-test when applicable.

Normality of quantitative data was assessed with the Shapiro-Wilk test. Variables that were normally distributed were expressed as mean $\pm$ standard deviation and compared using the Student's *t*-test for independent samples. Non-normally distributed variables were presented as median and interquartile range (25th–75th percentiles) and analyzed with the Mann-Whitney *U*-test.

The Intubation Time (IT) was additionally analyzed using the Kaplan-Meier method to estimate the probability of successful intubation over time, with groups compared via the log-rank test. To supplement this, a Cox proportional hazards regression model was applied to calculate the Hazard Ratio (HR) and its 95% Confidence Interval, evaluating the relative rate of intubation success between the VL3D and Macintosh groups.

Sample size and power calculation

A minimum sample size of 46 patients (23 per group) was initially estimated based on a conservative a priori calculation for the primary outcome (IDS). This was calculated to detect an expected two-point difference in Intubation Difficulty Scale (IDS) scores (an effect considered clinically relevant). The calculation assumed a standard deviation of approximately 2 points, a significance level (α) of 0.05, and 80% power. To compensate for potential losses due to follow-up or protocol deviations, an additional 10% was added, resulting in a total sample of 52 patients to be enrolled. This calculation was performed using the PSS Health online tool.⁹

Results

A total of 52 adult patients undergoing general anesthesia were initially enrolled in the study. The final analysis followed a modified intention-to-treat principle, including 49 participants: 26 in the VL3D group and 23 in the MAC group. Following randomization, three patients from the MAC group were excluded due to protocol deviations (failure to meet inclusion criteria discovered after allocation). These findings are detailed in the CONSORT flow diagram (Fig. 1). No statistically significant differences were observed between the groups in terms of sociodemographic or clinical characteristics (Table 3).

Primary outcome

The primary outcome, assessed by the Intubation Difficulty Scale (IDS), showed that median scores were statistically similar between groups. However, categorical analysis revealed a significantly higher incidence of moderate-to-severe intubation difficulty ($IDS > 5$) in the VL3D group (23.1%) compared to the MAC group (0%) ($p = 0.032$). These results are detailed in Table 4.

Secondary outcomes

No statistically significant differences were observed between the groups in relation to the Airway Assessment Score (AAS), whether analyzed as a continuous variable ($p = 0.479$) or dichotomized into scores below and above 6 points ($p = 0.999$). Regarding glottic visualization, the Cormack-Lehane classification revealed a statistically significant difference between groups ($p = 0.002$). A significantly higher proportion of VL3D patients were classified as Cormack-Lehane grade I (72.0%) compared to MAC patients (33.3%) ($p = 0.002$). Conversely, grade III classifications were more frequent in the MAC group (28.6%) and entirely absent in the VL3D group (0%) (Table 4).

Intubation time and adverse events

The Kaplan-Meier analysis demonstrated a significantly lower probability of rapid successful intubation over time for the VL3D group compared to the MAC group (log-rank test, $p = 0.001$; Fig. 2). The estimated median time for successful intubation was 70 seconds (95% CI 39.1–100.8) for the VL3D group and 40 seconds (95% CI 36.6–43.3) for the MAC group. Cox proportional hazards regression confirmed these findings, indicating that the MAC group achieved successful intubation at a significantly higher rate per unit of time than the VL3D group. No significant differences were found in the incidence of adverse events (bleeding, dental trauma, or mucosal injury) between the VL3D and MAC groups.

Discussion

In this randomized controlled trial, we investigated the performance of the VL3D in comparison with the traditional Macintosh laryngoscope for tracheal intubation in patients undergoing general anesthesia. Our findings demonstrate that the VL3D was associated with higher procedural difficulty and longer intubation times, despite providing improved glottic visualization. The primary outcome (IDS) revealed a higher frequency of moderate-to-severe intubation difficulty in the VL3D group compared to the MAC group in categorical analysis.

Videolaryngoscopes (VLs) offer several advantages over Direct Laryngoscopes (DLs), including enhanced glottic visualization, improved coordination between operator and assistant, a shorter learning curve, ease of use for novice practitioners, a wider field of view, reduced intubation failure rates, and minimal need for head or neck manipulation.¹⁰ These benefits have contributed to the growing adoption of VLs not only in cases of anticipated difficult

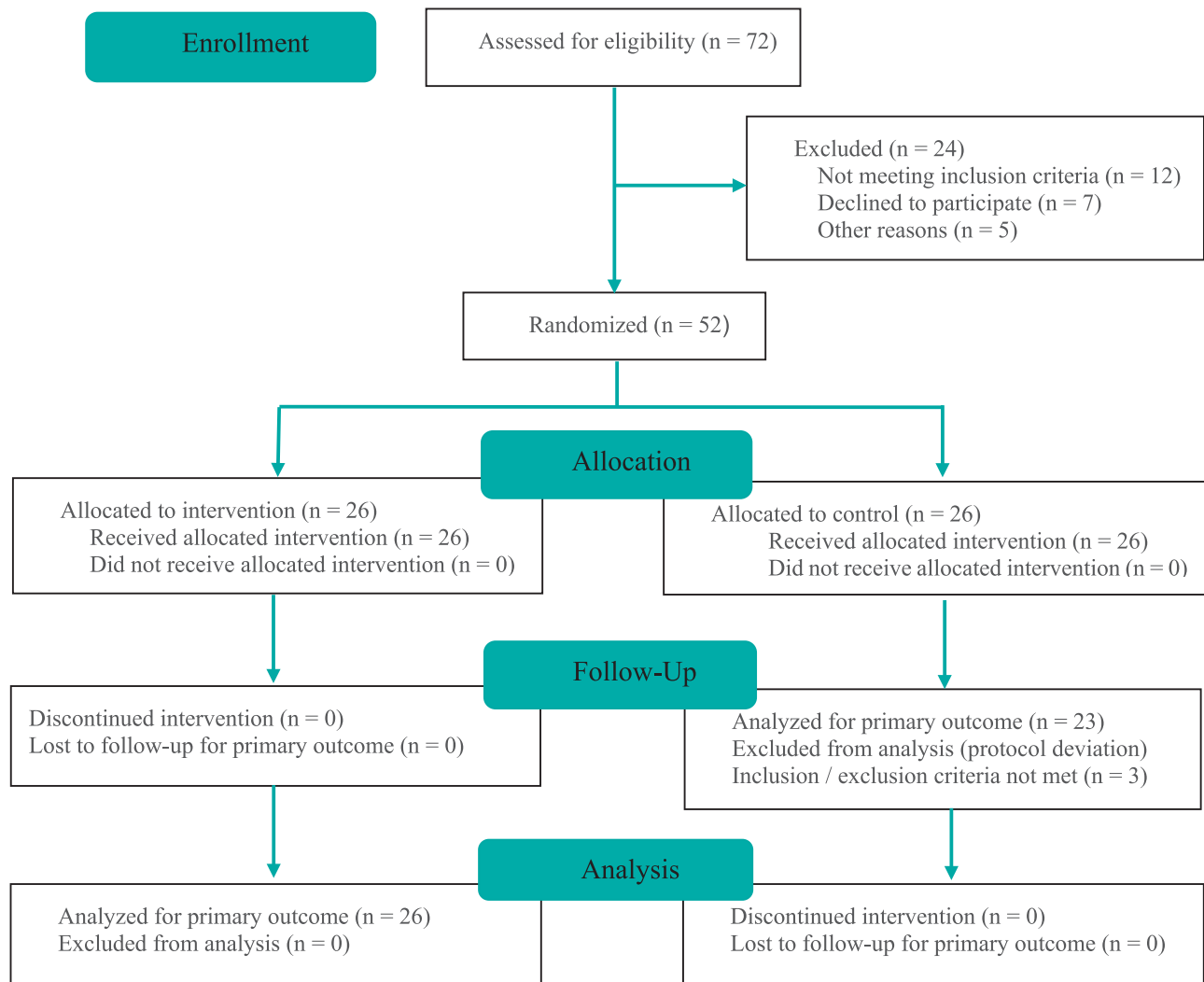


Figure 1 CONSORT flow diagram illustrating the enrollment allocation, follow-up, and analysis of patients randomized to receive intubation with VL3D versus the MAC.

airway management but also as replacements for DLs in routine clinical practice.

Commercially available VLs vary considerably in blade design. Some models retain the general shape of the Macintosh blade, while others feature more pronounced angulation.

The VL3D used in this study featured a curved blade that consistently provided an adequate view of the glottis. Regarding our secondary outcomes, the VL3D consistently provided an adequate view of the glottis. Notably, all patients in the VL3D group were classified as Cormack-Lehane Grade I or II, whereas the MAC group included patients with Grades III and IV classifications. This distinction is clinically relevant, as poor glottic visualization (Grades III or IV) is associated with a twofold increase in airway-related complications, especially in non-operating room environments.¹¹

It is important to recognize that patient safety during airway management is multifactorial. Although VLs represent a significant advancement in airway visualization, their contribution to safety is maximized only when integrated into comprehensive, well-structured airway management protocols.¹²

Although the VL3D provided enhanced glottic visualization, it is crucial to note that this did not translate into an easier intubation process. The implementation of Kaplan-Meier curves and Cox regression analysis provided a robust reinforcement of our findings regarding the intubation time. These methods highlighted a lower probability of rapid intubation success with the VL3D compared to the Macintosh laryngoscope. This suggests that while the VL3D model offers superior visualization, it may currently present additional difficulties in clinical performance, particularly in emergency situations or difficult airway management where time is a critical factor. These results reinforce that the VL3D model requires further design optimization and more extensive clinical testing in real-life scenarios before it can be recommended for time-sensitive airway interventions. This outcome may be partially attributed to disparities in operator experience between the two devices. The anesthesiologists participating in the study had between 3 to 6 years of experience/training in the specialty and had greater familiarity with the Macintosh laryngoscope, which likely contributed to the increased challenges observed with the VL3D.

Table 3 Comparison between different sociodemographic and clinical groups.

Characteristics	VL3D Group		MAC Group		p
	n	Descriptive measures	n	Descriptive measures	
Gender, n (%)	26		23		0.405 ^a
Female			14 (53.8)		16 (69.6)
Male			12 (46.2)		7 (30.4)
Age, mean $\pm$ SD	26	58.2 $\pm$ 13.9	23	61.9 $\pm$ 16.2	0.414 ^b
Surgical complexity n (%)	26		22		0.100 ^a
Low			4 (15.4)		2 (9.1)
Medium			11 (42.3)		16 (72.7)
High			11 (42.3)		4 (18.2)
Weight, mean $\pm$ SD	26	73.2 $\pm$ 15.3	23	75.3 $\pm$ 15.3	0.641 ^b
Height, mean $\pm$ SD	26	1.65 $\pm$ 0.09	23	1.66 $\pm$ 0.10	0.689 ^b
BMI, mean $\pm$ SD	26	26.8 $\pm$ 4.3	23	27.4 $\pm$ 5.2	0.662 ^b
ASA – PS, n (%)	25		22		0.999 ^a
I			2 (8.0)		2 (9.1)
II			17 (68.0)		14 (63.6)
III			6 (24.0)		6 (27.3)

BMI, Body Mass Index; ASA-PS, American Society of Anesthesiologist Physical Status Classification System; SD, Standard Deviation.

^a Chi-Square test with Yates correction or Fisher's exact test.

^b Student's *t*-test for independent samples.

This discrepancy in prior training represents a significant methodological limitation of the study.

It is also plausible that the VL3D prototype used in this study may delay intubation in straightforward cases, although performance may vary across different VL models. Blade design may confer specific advantages depending on clinical context, such as intubation in obese patients, anticipated difficult airways, emergency settings, or when performed by less experienced clinicians.²

Study limitations and applicability

It is worth noting that randomization produced well-balanced groups in terms of demographic and clinical characteristics. Furthermore, no statistically significant differences were found in the Airway Assessment Score (AAS), whether analyzed as a continuous or categorical variable, suggesting a comparable baseline risk of difficult airway between groups. These findings strengthen the internal validity of our results

Table 4 Comparison of outcomes between groups.

	VL3D Group		MAC Group		p
	n	Descriptive measure	n	Descriptive measure	
AAS, mean $\pm$ SD	26	5.04 $\pm$ 2.74	23	4.48 $\pm$ 2.7	0.479*
Categories AAS, n (%)	26		23		
$\leq$ 6 points		19 (73.1)		17 (73.9)	0.999**
$>$ 6 points		7 (26.9)		6 (26.1)	
Cormack-Lehane grade, n (%)	25		21		0.002**
1		18 (72.0) ^b		7 (33.3) ^a	
2		7 (28.0)		6 (28.6)	
3		0 (0) ^b		6 (28.6) ^a	
4		0 (0)		2 (9.5)	
IT (sec), median (IQR)	26	70 (40–105)	23	40 (30–43)	0.005***
IDS, median (IQR)	26	2 (0–5)	23	1 (0–3)	0.240***
Categories IDS, n (%)	26		22		0.032**
0 – Easy		10 (38.5)		8 (36.4)	
0–5 – Mild difficulty		10 (38.5)		14 (63.6)	
$>$ 5 – Moderate to severe difficulty		6 (23.1) ^b		0 (0%) ^a	

AAS, Airway Assessment Score; SD, Standard Deviation; IT, Intubation Time (sec); IQR, Interquartile Range (percentiles 25–75); IDS, Intubation Difficulty Scale.

* Student's *t*-test for independent samples.

** Chi-Square test with Yates correction or Fisher's exact test.

*** Mann-Whitney test.

^{a,b} Different superscript letters indicate significant.

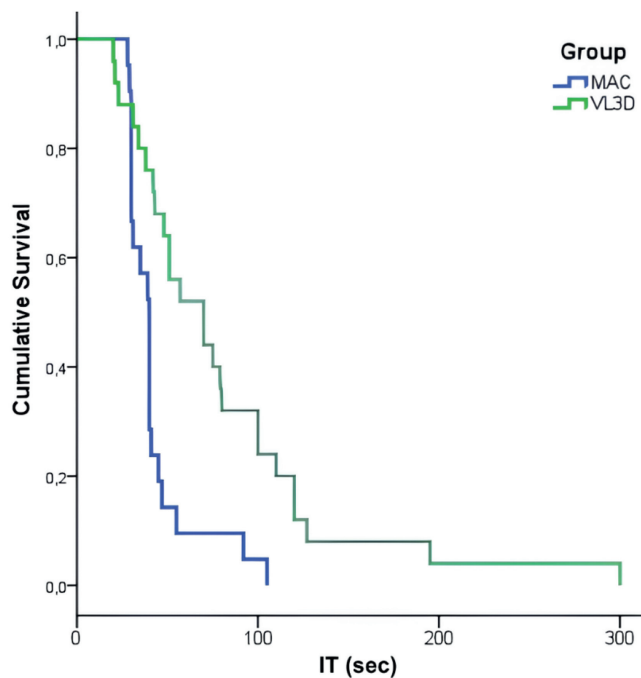


Figure 2 Kaplan-Meier curves for the probability of successful intubation over time. The Macintosh (MAC) group (blue line) demonstrated a significantly higher probability of rapid intubation success compared to the VL3D group (green line) (log-rank test, $p = 0.001$). IT, Intubation Time (sec).

and support the interpretation that increased intubation difficulty in the VL3D group was not due to intrinsic patient-related airway complexity. A significant limitation of this trial is the exclusion of patients with predictors of difficult airway, such as those with a Body Mass Index (BMI) greater than 35 kg.m^{-2} . Given that videolaryngoscopes are particularly recommended for difficult airways, it is highly plausible that, in this specific subgroup of patients (e.g., morbidly obese), the VL3D would demonstrate superior performance compared to the conventional Macintosh blade. Therefore, the findings regarding procedural time and difficulty are primarily applicable to patients with non-difficult airways.

Additional limitations include the lack of blinding for the intubating clinicians, who were aware of group assignments. Although blinding was maintained throughout other phases of the study, the inability to blind the clinician responsible for the intubation may have introduced bias.

Conclusion

In conclusion, the VL3D videolaryngoscope demonstrated superior glottic visualization but was associated with increased intubation difficulty and longer procedural times compared to the Macintosh laryngoscope. These results underscore the importance of structured training and suggest potential for further optimization of the device before clinical implementation. Additional studies are being conducted by our research group to optimize the design and validate the long-term clinical feasibility of 3D-printed laryngoscopes.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare no conflicts of interest.

Clinical trial registration

Brazilian Clinical Trials Registry (ReBEC) under number RBR-9t36r35.

AI assistance disclosure

The authors used Gemini (Google AI) for language polishing and structural refinement of the manuscript to ensure clarity and adherence to the journal's requirements. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This study was supported by the Santa Casa de Porto Alegre, listed as the primary sponsor in the Brazilian Clinical Trials Registry (ReBEC: RBR-9t36r35). No additional external funding was received.

Associate Editor

Lorena Ibiapina Mendes de Carvalho

References

1. Lewis SR, Butler AR, Parker J, Cook TM, Smith AF. Videolaryngoscopy versus direct laryngoscopy for adult patients requiring tracheal intubation. *Cochrane Database Syst Rev.* 2022;4. CD011136.
2. Hansel J, Rogers AM, Lewis SR, Cook TM, Smith AF. Videolaryngoscopy versus direct laryngoscopy for adults undergoing tracheal intubation: A Cochrane systematic review and meta-analysis update. *Br J Anaesth.* 2022;129:612–23.
3. Prekker ME, Driver BE, Trent SA, Resnick-Aultet D. DEVICE Investigators and the Pragmatic Critical Care Research Group. Video versus direct laryngoscopy for tracheal intubation of critically ill adults. *N Engl J Med.* 2023;389:418–29.
4. Fonternel T, van Rooyen H, Joubert G, Turton E. Evaluating the usability of a 3D-printed video laryngoscope for tracheal intubation of a manikin. *Med Devices.* 2023;16:157–65.
5. Karippacheril JG, Cong ML. Videolaryngoscopy using an Android smartphone: A direct digital technique. *Indian J Anaesth.* 2016;60:143–5.

6. Mendes FF, Spinelli LF, Dutra PAS, de Braga ES, et al. Three-dimensional printed laryngoscopes as allies against COVID-19. *3D Print Addit Manuf.* 2023;10:930–40.
7. Seo SH, Lee JG, Yu SB, Kim DS, et al. Predictors of difficult intubation defined by the Intubation Difficulty Scale (IDS): predictive value of 7 airway assessment factors. *Korean J Anesthesiol.* 2012;63:491–7.
8. Adnet F, Borron SW, Racine SX, Clemessy JL, et al. The Intubation Difficulty Scale (IDS): proposal and evaluation of a new score characterizing the complexity of endotracheal intubation. *Anesthesiology.* 1997;87:1290–7.
9. Borges RB, Mancuso ACB, Camey S, Leotti VD, et al. Power and Sample Size for Health Researchers: uma ferramenta para cálculo de tamanho amostral e poder do teste voltado a pesquisadores da área da saúde. *Clin Biomed Res.* 2021;40.
10. Sakles JC, Rodgers R, Keim SM. Optical and video laryngoscopes for emergency airway management. *Intern Emerg Med.* 2008;3:139–43.
11. Martin LD, Mhyre JM, Shanks AM, Tremper KK, Kheterpal S. 3,423 emergency tracheal intubations at a university hospital: airway outcomes and complications. *Anesthesiology.* 2011;114:42–8.
12. Paolini JB, Donati F, Drolet P. Review article: video-laryngoscopy: another tool for difficult intubation or a new paradigm in airway management? *Can J Anaesth.* 2013;60:184–91.

ORIGINAL INVESTIGATION

Use, perceptions, and barriers to supraglottic airway devices in Brazil: a national cross-sectional survey with global implications



Jayme M. dos Santos Neto ^{a,*}, Flávia J. de Barros ^a, Victor M. Lemos ^a,
Clístenes Cristian de Carvalho ^{b,c}

^a Hospital das Clínicas da Universidade Federal de Pernambuco (HC-UFPE), Recife, PE, Brazil

^b Universidade Federal de Campina Grande (UFCG), Campina Grande, PB, Brazil

^c Guy's and St. Thomas' NHS Foundation Trust, Department of Anaesthesia and Perioperative Medicine, London, UK

Received 11 September 2025; accepted 9 April 2026

Available online 27 April 2026

KEYWORDS

Airway management;
Anesthesiology;
Continuing medical
education;
Laryngeal masks;
Patient safety;
Surveys and
questionnaires

ABSTRACT

Background: Supraglottic Airway Devices (SADs) are essential in airway management. However, little is known about their use, perceptions, and barriers in large, resource-variable healthcare systems such as Brazil. We aimed to assess current practice patterns, training exposure, and barriers to SADs use among Brazilian anesthesiologists and trainees.

Methods: A nationwide, web-based survey that addressed demographics, training exposure, device availability, frequency of use, perceived safety and efficacy, barriers to use, and knowledge of SADs generations. Descriptive and inferential statistics were used to analyze regional and experiential patterns.

Results: A total of 432 completed responses were analyzed. Most respondents were fully trained anesthesiologists (75.5%) from all five Brazilian regions. Although SADs were widely perceived as safe (76.9%) and capable of providing adequate ventilation (74.3%), only 35% reported using them at least once per week, and fewer than 20% felt confident distinguishing between first and second-generation devices. Binomial logistic regression revealed that greater training exposure during residency was strongly associated with more frequent current use (OR = 7.3; 95% CI 3.8–14.3, $p < 0.001$). The most common barrier was lack of operator experience (81.5%), followed by limited availability of devices (28.9%). Use varied across regions ($p < 0.001$), with 82% of anesthesiologists in the South reporting weekly use compared to only 15.4% in the North. Respondents were more likely to use SADs in limb surgeries, while use is limited in gastrointestinal and thoracic procedures.

* Corresponding author.

E-mail: jaymemed@yahoo.com.br (J.M. dos Santos Neto).

Conclusions: Despite favorable perceptions of safety and efficacy, SADs remain underused in Brazil, with marked regional variation and limited familiarity with device generations. Training exposure emerged as a strong predictor of current use, underscoring the importance of structured education and hands-on experience. Addressing gaps in training and improving device availability may help standardize practice and enhance the use of SADs across the country.

Institutional Research Board approval n° 7.214.701; CAAE: 84060624.6.0000.8807.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Airway management is a cornerstone of anesthetic practice and a critical determinant of patient safety.¹ Over recent decades, Supraglottic Airway Devices (SADs) have become established as reliable, versatile tools in both routine and emergency settings.² Their ease of insertion, favorable safety profile, and applicability in both spontaneous and controlled ventilation have made them integral to modern anesthesia practice. Second-generation SADs are increasingly favored due to improved oropharyngeal sealing, integrated gastric channels, and reduced risk of aspiration.²⁻⁵

A growing body of evidence from large observational studies and randomized trials supports the efficacy and safety of second-generation SADs across diverse clinical contexts, including complex surgeries and high-risk populations.^{4,6,7} For instance, a prospective international cohort study reported a 98.6% success rate with the i-gel Plus, with minimal complications and high ease-of-use ratings.⁶ Additionally, real-world studies have shown that transitioning from first to second-generation SADs improves airway safety in routine care.⁷ A recent systematic review and meta-analysis further demonstrated that, in adults undergoing abdominopelvic surgery, second-generation SADs reduced major airway complications and significantly improved postoperative recovery outcomes compared to tracheal intubation.⁴

Despite this compelling evidence, significant variability in SADs usage persists across countries and institutions.¹ Disparities in training, device access, and risk perception continue to shape practice patterns. Brazil, home to one of the largest and most diverse healthcare systems globally, represents a critical yet underexplored setting in this context.

As a country of continental dimensions, Brazil exhibits substantial regional disparities in healthcare infrastructure, resource allocation, and anesthesiology training. While regions such as the South and Southeast mirror high-income systems in terms of practice and access, others, particularly in the North, face challenges more typical of low-resource environments.⁸ Thus, national data from Brazil can serve as a proxy for the broader challenges encountered in resource-variable global contexts. Yet, little is known about how Brazilian anesthesiologists currently use SADs, perceive their safety and efficacy, or experience barriers to their adoption.

Accordingly, this study aimed to evaluate current practice patterns, perceptions of safety and efficacy, and barriers to the use of supraglottic airway devices among anesthesiologists in Brazil.

Methods

This cross-sectional study was based on an anonymous electronic questionnaire distributed to eligible participants. The

research adhered to the ethical principles outlined in the Declaration of Helsinki and complied with Resolution 466/12 of the Brazilian National Health Council. Ethical approval was granted by the study's hospital Research Ethics Committee (approval #7.214.701; CAAE: 84060624.6.0000.8807). This manuscript conforms to established reporting standards and adheres to the Checklist for Reporting of Survey Studies (CROSS).⁹⁻¹¹

A link to the survey, accompanied by a brief cover letter outlining the study's objectives, was created for national distribution and subsequently deactivated after the study period. Before accessing the questionnaire, participants viewed an online information sheet describing the study objectives, procedures, voluntary nature of participation, right to withdraw at any time, and contact details for the investigators. Electronic informed consent was obtained by selecting an "I agree to participate" option. Respondents who did not provide consent could not proceed to the survey.

The final version of the questionnaire comprised 26 questions across 11 sections: General Information; Experience in Anesthesiology; Professional Practice; Experience with Supraglottic Devices; Use of Supraglottic Devices in Elective General Anesthesia; Perceived Safety of Supraglottic Airway Devices; Adequacy of Ventilation with Supraglottic Airway Devices; Barriers to Supraglottic Airway Devices; Knowledge of Device Generation; Scenarios for Use; and Continuing Education (Appendix S1). The survey was designed to assess clinical use, perceived safety and efficacy, confidence and training, availability, and barriers to the adoption of supraglottic airway devices across a range of clinical scenarios. The primary outcomes were defined and established *a priori* at initiation of the study design.

Initially consisting of 33 items, the questionnaire was refined over four iterative rounds of review by the research team. A pretest was conducted with four anesthesiologists to evaluate clarity and technical functionality. Responses from the pretest phase were excluded from the final analysis. After minor adjustments, the finalized version of the survey was locked and not modified thereafter.

The link was distributed twice by the Brazilian Society of Anesthesiology (SBA) via institutional email to all registered members and was additionally shared by the research team through social media platforms (WhatsApp® and Instagram®) to enhance dissemination. Eligible participants included board-certified anesthesiologists and anesthesiology residents with a minimum of three months of specialty training. The survey remained open from November 27, 2024, to January 2, 2025.

Study data were collected and managed using REDCap electronic data capture tools hosted at the study hospital's university.^{12,13} REDCap (Research Electronic Data Capture),

developed by Vanderbilt University (Nashville, TN, USA), is a secure, web-based software platform designed to support data capture for research studies, providing: 1) An intuitive interface for validated data capture; 2) Audit trails for tracking data manipulation and export procedures; 3) Automated export procedures for seamless data downloads to common statistical packages; and 4) Procedures for data integration and interoperability with external sources. The data were downloaded as a comma-separated values (.csv) file, and only fully completed questionnaires were included in the analysis.

Sample size estimation was performed using the 'samplesize4surveys'¹⁴ package in R software (R Foundation for Statistical Computing, Vienna, Austria), adopting a conservative approach. Considering the total membership of the SBA as the population ($n = 13,265$), we assumed a 95% confidence level, a 3% margin of error, and a conservative estimated prevalence of 10% of anesthesiologists using a SAD at least once per week. Under these assumptions, the required sample size was calculated to be 374. Anticipating a low response rate of approximately 5% via email and social media distribution, we estimated that outreach to at least 7480 members would be necessary. To maximize participation and enhance representativeness, the survey was sent to all 13,265 SBA members.

Statistical analysis

All statistical analyses were performed using JAMOV (version 2.0, MacOS X) and R (version 4.4.2). Descriptive statistics were used to summarize respondent characteristics and survey responses, with results presented as means $\pm$ standard deviation for continuous variables, and as absolute frequencies and percentages for categorical variables. For ordinal and non-normally distributed variables, nonparametric tests were applied, including the Kruskal-Wallis rank sum test for regional comparisons and Spearman's rank correlation to assess the association between training exposure and current use of SADs. A two-sided p -value of < 0.05 was considered statistically significant.

A binomial logistic regression was conducted considering the responses of fully trained anesthesiologists only. The independent variables of interest were frequency of use during anesthesiology specialization, self-reported experience with SADs, knowledge of device generation, age ($45 < \text{vs.} \geq 45$) and professional workplace (public or private hospital). Bivariate analyses (Chi-Square or Fisher's exact test) were employed to determine the association between each independent variable of interest and the outcome (use of the device). Bivariate models with $p < 0.30$ and theoretically relevant variables in the context of the study were selected for the multiple model. The model fit was checked using deviance and AIC (Akaike Information Criterion) with results presented as odds ratio values, 95% confidence intervals, and statistical significance. For all analyses, a p -value < 0.05 was considered.

Results

A total of 587 responses were collected through REDCap[®] following survey distribution to all 13,265 members of the

Brazilian Society of Anesthesiology, resulting in an overall response rate of approximately 4.4%. After excluding incomplete submissions and entries without signed consent, 432 fully completed questionnaires were included in the final analysis, yielding a usable response rate of 3.3%. Respondents who identified as students were automatically excluded via survey logic.

Participants were predominantly fully trained anesthesiologists (75.5%) from all five Brazilian regions, though the majority were based in the Northeast (53.9%). Most respondents worked in general hospitals (91.9%) and had substantial clinical exposure, with over 80% involved in more than 150 general anesthesia procedures annually (Table 1).

Most participants reported at least moderate experience with SADs: 39.6% indicated "moderate experience", 33.8% "extensive experience", and 15.5% considered themselves experts (Table 2). Frequent clinical use was reported by 35.4% (at least once a week) and 25.0% (at least once a month). SADs were reportedly always available in 70.6% of workplaces and partially available in 28.9%.

Usage patterns varied substantially by region. In the South, 82% of respondents reported using SADs at least once per week, while in the North – where overall use was markedly lower – only 15.4% reported weekly use and 30.8% reported use once a year or less (Fig. 1). A statistically significant difference across regions was observed (Kruskal-Wallis test, $\chi^2(4) = 107.49$, $p < 0.001$).

Figure 2 shows the association between training exposure and current use: clinicians who used SADs regularly during residency were more likely to report frequent use in current practice. This was supported by Spearman correlation ($\rho = 0.60$, $p < 0.001$), indicating a moderate-to-strong positive relationship.

SADs were generally perceived as safe: 76.9% of respondents rated them as "almost always" or "always" safe, and 74.3% considered ventilation via SADs "almost always" or "always" adequate (Table 3). The lack of operator experience was the most commonly cited barrier, with 81.5% agreeing it limited broader use. Only 18.3% felt fully confident distinguishing between first- and second-generation devices.

SADs were most commonly selected for upper (85.2%) and lower limb surgeries (84.5%), followed by ophthalmology and thoracoabdominal wall procedures. Their use was less favored for spinal, thoracic, and cardiac surgeries. The primary hypothetical concerns limiting their use were inadequate ventilation (72.5%), regurgitation (68.8%), and aspiration of gastric contents (68.6%).

A large majority (91.5%) of respondents agreed that greater access to training and information would likely encourage broader implementation of SADs in anesthetic practice.

Binomial logistic regression analysis revealed that self-reported experience with SADs (OR = 14.8; 95% CI 5.5–38.0, $p < 0.001$), the frequency of use during anesthesiology specialization (OR = 7.3; 95% CI 3.8–14.3, $p < 0.001$) and age ≥ 45 (OR = 2.0; 95% CI 1.0–4.0, $p = 0.035$) were associated with a greater chance of using the devices in clinical practice. Knowledge of device generation and professional workplace (public or private hospital) were not associated with its use ($p > 0.05$).

Table 1 Demographic and professional characteristics of survey respondents (n = 432).

Variable	n/Average	%/DP
Age (years)	39.6	10.9
Gender		
Female	181	41.9
Male	249	57.6
Not declared	2	0.5
Level of training		
Resident/Specializing in Anesthesiology	106	24.5
Anesthesiologist	326	75.5
Region of operation		
North	12	2.8
Northeast	233	53.9
Midwest	25	5.8
Southeast	112	25.9
South	50	11.6
Workplace		
General Hospital	398	91.9
Specialized hospital	155	35.8
Day hospital	117	27.0
Profile of the work institution		
Public	253	58.4
Private	238	55.0
Mixed	138	31.9
Number of surgical procedures with general anesthesia per year		
0–50 (up to 1-per week)	24	5.6
51–150 (1–3 per week)	58	13.4
> 150 (more than 3-per week)	350	81.0

SD, Standard Deviation.

Discussion

This national survey likely provides the first comprehensive snapshot of SADs use, perceptions, and barriers among Brazilian anesthesiologists. While SADs are widely recognized as safe and effective, our findings reveal considerable variability

in their use across regions. This variation appears to be shaped by multiple factors, including differences in training, clinical confidence, institutional norms, and access to equipment. These patterns carry implications not only for Brazil, but also for other middle-income countries and health systems with internal disparities.

Table 2 Experience with, use and availability of Supraglottic Airway Devices (SADs).

Variable	n	%
Experience with SADs		
No experience	4	0.9
Minimal experience	44	10.2
Moderate experience	171	39.6
Extensive experience	146	33.8
Expert level	67	15.5
Use of SADs in elective procedures		
Once a year or less	57	13.2
Two to three times a year	51	11.8
Four to six times a year	63	14.6
Once a month or more	108	25.0
Once a week or more	153	35.4
Availability of SADs in the performance services		
Never available	0	0
Partially available	125	28.9
Only available for emergencies	2	0.5
Always available	305	70.6

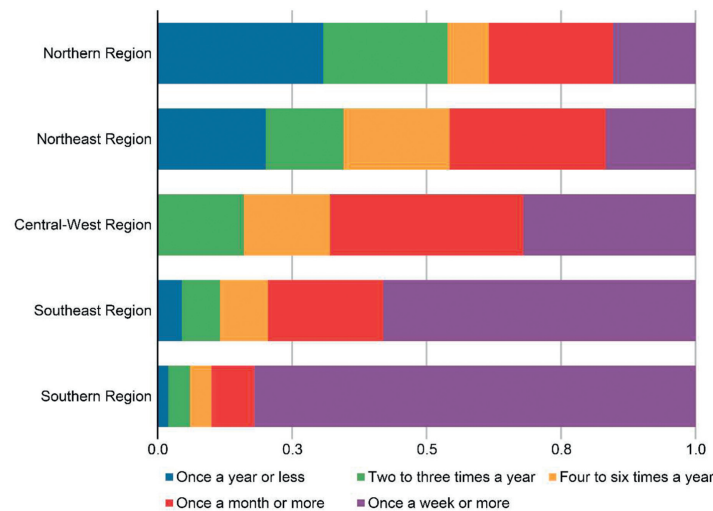


Figure 1 Regional variability in the frequency of Supraglottic Airway Devices (SADs) use in Brazil. This horizontal stacked bar chart displays the distribution of reported frequency of SADs use in elective general anesthesia across Brazil's five geographic regions. The Y-axis lists the regions, while the X-axis shows the percentage of respondents within each region. Colored segments represent frequency categories, ranging from "Once a year or less" to "Once a week or more", as indicated in the legend. Respondents from the South most commonly reported weekly use (82%), whereas in the North, only 15.4% reported using SADs weekly and 30.8% reported use once a year or less, highlighting pronounced regional disparities in clinical practice.

Participants mostly viewed SADs as safe and effective for ventilation, in line with robust international evidence supporting their use in both elective and emergency settings.^{4,5,15-18} Large-scale studies, randomized trials, and meta-analyses have shown that SADs, particularly second-generation ones, reduce perioperative complications and enhance recovery compared to tracheal intubation.^{4,15,18} However, real-world usage remains limited: only one-third of respondents reported weekly use, and fewer than one in five felt confident distinguishing between first- and second-generation devices.

These findings highlight a persistent gap in training and continuing education. Accordingly, early and structured exposure during residency was associated with current use,

reinforcing the critical role of specialist training in shaping long-term practice. Similarly, lack of operator experience emerged as the most frequently cited barrier, underscoring the need to strengthen education through simulation, clinical mentorship, and structured airway curricula – particularly in regions where SADs remain underused.

The educational gap is especially relevant in light of recent advances such as video-guided SADs, which aim to improve device placement, airway seal, and anatomical visualization.^{2,19} While promising, the successful adoption of such innovations depends on a foundational understanding of existing SADs technologies. Our findings suggest that even this baseline familiarity remains limited in many

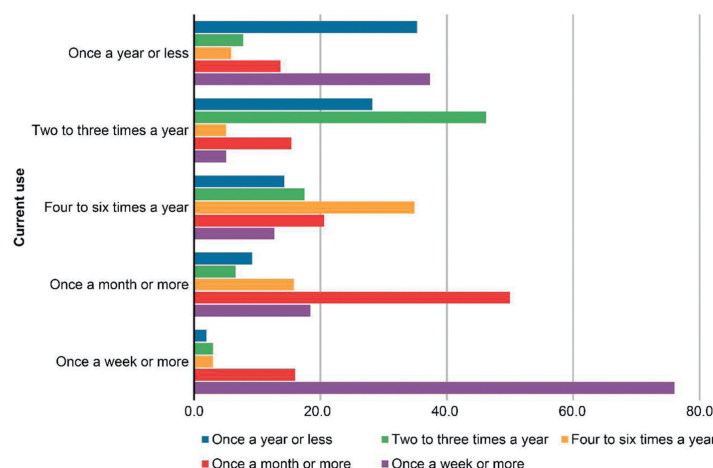


Figure 2 Association between training exposure during residency and current use of Supraglottic Airway Devices (SADs). This bar chart illustrates the relationship between the frequency of SADs use during anesthesiology residency (represented by colored bars) and current use in elective general anesthesia (Y-axis). Each colored segment reflects a level of training exposure during anesthesiology specialization, ranging from "Once a year or less" to "Once a week or more", as indicated in the legend. The X-axis shows the percentage of respondents in each current use category. The figure demonstrates a clear pattern: more frequent exposure to SADs during residency is associated with greater frequency of use in current practice.

Table 3 Perceptions, barriers, and training exposure related to Supraglottic Airway Devices (SADs) use.

Variable	n	%
Degree of safety		
Not safe at all	2	0.5
Rarely safe	8	1.9
Sometimes safe	90	20.8
Usually, safe	235	54.4
Always safe	97	22.5
Adequate ventilation		
Not effective at all	3	0.7
Rarely effective	7	1.6
Sometimes effective	101	23.4
Usually effective	292	67.6
Always effective	29	6.7
Supposed barriers to use		
Cost		
Strongly disagree	56	13.0
Disagree	131	30.3
Neither agree nor disagree	83	19.2
Agree	139	32.2
Totally agree	23	5.3
Availability		
Strongly disagree	38	8.8
Disagree	95	22.0
Neither agree nor disagree	59	13.7
Agree	186	43.1
Totally agree	54	12.5
Experience of professionals		
Strongly disagree	7	1.6
Disagree	40	9.3
Neither agree nor disagree	33	7.6
Agree	255	59.0
Totally agree	97	22.5
Knowledge of the differences between 1 st and 2 nd generation SADs		
No knowledge at all	10	2.3
Basic understanding	111	25.7
Moderate understanding	123	28.5
Good understanding	109	25.2
Excellent understanding	79	18.3
Elective surgeries in which you would use a SAD		
Videolaparoscopic gastrointestinal surgery	29	6.7
Videolaparoscopic urogenital surgery	94	21.7
Open gastrointestinal tract	62	14.3
Open urogenital	233	53.8
Upper limbs	369	85.2
Lower limbs	366	84.5
Head and neck	35	8.1
Thoracoabdominal wall	247	57.0
Ophthalmology	262	60.5
ENT (except oropharynx and larynx)	101	23.3
Spine	7	1.6
Cardiac	3	0.7
Thoracic	6	1.4
Caesarean section	29	6.7
I wouldn't use it in any situation	12	2.8
Potential limitations to use		
Regurgitation	298	68.8
Aspiration of gastric contents	297	68.6
Failed 1st attempt	29	6.7
Insertion failure	195	45.0
Low sealing pressure	156	36.0
Inadequate ventilation	314	72.5
Sore throat	51	11.8
Hoarseness	23	5.3
Post-operative nausea and vomiting	12	2.8

Table 3 (Continued)

Variable	n	%
Coughing on waking	7	1.6
Dysphagia	39	9.0
Tissue damage	61	14.1
Insertion time	122	28.2
No limitations	10	2.3
Access to information would increase the use of SADs		
Strongly disagree	4	0.9
Disagree	7	1.6
Neither agree nor disagree	26	6.0
Agree	164	38.0
Totally agree	231	53.5

SAD, Supraglottic Airway Device.

settings, particularly those with fewer resources. To avoid widening practice disparities, global stakeholders must prioritize basic training and knowledge dissemination alongside technological advancement.

In addition to training, limited access to SADs remains a key constraint.²⁰ Nearly 30% of respondents reported only partial availability in their workplaces. This disconnect between perceived clinical value and real-world access highlights ongoing logistical and procurement barriers that must be addressed to enable broader and more equitable use in routine anesthesia care.

Regional differences further illustrate these challenges. In the South – a more well-resourced region – over 80% of respondents reported using SADs weekly. In contrast, in the North – where healthcare infrastructure is often more limited – only 15% reported weekly use, while nearly one-third used them just once per year or less. These findings reflect broader patterns of inequity and call for targeted interventions to improve access and training in underserved regions.²⁰⁻²³

Such disparities are not unique to Brazil. Given its continental scale and internal diversity, Brazil represents a microcosm of global anesthetic practice – with regions resembling high-income settings and others facing constraints more typical of low-resource environments. These findings therefore offer transferable insights into the barriers that affect SADs implementation in a variety of global contexts.²⁰⁻²³

Our data also suggest a cautious approach to the types of surgical procedures in which SADs are used. While upper and lower limb surgeries were commonly cited as appropriate, SAD use in gastrointestinal, laparoscopic, and thoracic procedures remained limited. The most frequent concerns were inadequate ventilation, regurgitation, and aspiration – despite strong evidence supporting the safety of second-generation devices, even in high-risk populations.^{3,4,6,7} Furthermore, the perception of an “appropriate surgical duration” for the use of supraglottic airway devices may represent a potential barrier to wider adoption. Although this aspect was not specifically explored in this study, it is an important dimension of clinical decision-making that may warrant further investigation in future research. This reinforces the need for targeted educational efforts to update clinical perceptions and expand the indications for SADs use.

Similar challenges have been reported internationally. In China – a continental country with regional disparities

similar to Brazil – a national survey found wide variation in airway management strategies for difficult airways and emphasized training and education as priorities for standardization.²² In India, a national survey among anesthesiologists also revealed regional variability in both equipment access and airway management practices, further underscoring the need to improve training and ensure equitable availability of airway tools.²⁰ Notably, even in high-income settings such as the United Kingdom, despite widespread access to airway tools, a survey revealed considerable variation in SADs use during in-hospital cardiac arrests, reinforcing the influence of institutional culture, education, and confidence on practice.²³ An international survey across 54 countries likewise reported broad inconsistencies in laryngeal mask airway use, with variability in perceived contraindications, procedure types, and patient characteristics. The authors called for broader dissemination of evidence-based guidelines to harmonize practice and improve safety.²¹

This study has several limitations. First, the use of convenience sampling and voluntary participation through social media channels may introduce selection bias and limit generalizability. Second, the survey relied on self-reported data, which may be subject to recall bias or social desirability effects. Of note, while a cross-sectional survey can rapidly provide a contemporary snapshot of practice and help describe a complex scenario, it does not permit firm causal inferences and is prone to biases and unmeasured confounding. Third, the questionnaire was not formally validated as a psychometric instrument. Fourth, we did not independently verify the institutional availability of SADs. Fifth, we did not delve into whether the perception of the device quality correlates with the generation available (first versus second generation SADs) or with the frequency of use. These factors may influence the notions and preferences reported by physicians and could help contextualize some of the findings. Exploring these associations in subsequent studies may provide additional insights into practice patterns. Sixth, although the collected sample met the a priori estimated size and allowed for precise analyses, the overall response rate was modest (4.4% total, 3.3% included), raising the potential for nonresponse bias and the possibility that respondents may differ meaningfully from nonrespondents. Nevertheless, these rates are comparable to those reported in other large-scale, web-based surveys of healthcare professionals and reflect the well-

documented challenges of engaging clinicians through electronic platforms.^{20,21} Finally, although the study captured a wide national distribution, representation from certain regions was limited. This may affect the extent to which the findings reflect regional practice variability. Even so, the inclusion of over 400 fully completed questionnaires provides valuable insight into national practice patterns and offers lessons that may be relevant to other settings with similar healthcare challenges.

Conclusion

This survey reveals that although SADs are broadly recognized as safe and effective, their adoption remains constrained by disparities in training, education, and access. Persistent concerns about certain clinical scenarios may also contribute to underuse. These findings emphasize the need to expand educational opportunities, modernize training programs, and ensure equitable access to airway devices.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' contributions

Jayme Marques dos Santos Neto: Refined the questionnaire; Secured ethical approval; coordinated survey distribution with the Brazilian Society of Anesthesiology, organized the dataset and wrote the manuscript.

Flávia Jatobá de Barros: Refined the questionnaire; secured ethical approval; coordinated survey distribution with the Brazilian Society of Anesthesiology, and organized the dataset.

Victor Macedo Lemos: Refined the questionnaire; secured ethical approval; coordinated survey distribution with the Brazilian Society of Anesthesiology, and organized the dataset.

Clístenes Cristian de Carvalho: Conceived the study idea; proposed the survey-based design; drafted the initial version of the questionnaire; conducted statistical analyses and wrote the manuscript.

All authors critically reviewed the final draft and approved the submitted version.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors gratefully acknowledge the Brazilian Society of Anesthesiology for its support in disseminating the survey to its full membership. We also thank the colleagues who contributed to the development and refinement of

the questionnaire through thoughtful feedback and pre-study review.

The authors used Gemini-based LLMs (Gemini 2.5) and Neural Machine Translation for language editing only; all scientific content and interpretation are the authors' own.

No funding organization has had any roles in the survey's design, implementation, and analysis.

Associate Editor

Mariana Fontes Lima Neville

References

1. Cook TM, Ogelsby F, Kane AD, Armstrong RA, Kursumovic E, Soar J. Airway and respiratory complications during anaesthesia and associated with peri-operative cardiac arrest as reported to the 7th National Audit Project of the Royal College of Anaesthetists. *Anaesthesia*. 2023;79(4).
2. Van Zundert AAJ, Gatt SP, Van Zundert TCRV, Hagberg CA, Pandit JJ. Supraglottic Airway Devices: Present State and Outlook for 2050. *Anesth Analg*. 2024;138:337–49.
3. Cook TM, Kelly FE. Time to abandon the “vintage” laryngeal mask airway and adopt second-generation supraglottic airway devices as first choice. *British Journal of Anaesthesia*. 2015;115:497–9.
4. de Carvalho CC, Kapsokalyvas I, El-Boghdady K. Second-Generation Supraglottic Airway Devices Versus Endotracheal Intubation in Adults Undergoing Abdominopelvic Surgery: A Systematic Review and Meta-Analysis. *Anesth Analg*. 2025;140:265–75.
5. Hummel R, Wollschläger D, Baldering Hans-Jürgen, Engelhard K, Wittenmeier E, Epp K, et al. Big data: Airway management at a university hospital over 16 years; a retrospective analysis. *PLoS One*. 2022;17:e0273549.
6. Werner J, Klementova O, Bruthans J, Macoun J, Gaszynski T, Henlin T, et al. Evaluation of the i-gel® Plus supraglottic airway device in elective surgery: a prospective international multi-centre study. *Anaesthesia*. 2024;79:1284–91.
7. Obeidat SS, Suleiman A, Ahrens E, Needham MJ, Stewart C, Khany M, et al. Real-World Evaluation of i-gel Introduction on Intraoperative Airway-Related Safety Events: A Retrospective Cohort Study from a New England Hospital Network. *Anesth Analg*. 2025;140:253–61.
8. Scheffer M. Demografia Médica no Brasil 2025. 1o ed Brasília, DF: Ministério da Saúde; 2025. [Internet][cited 2026 Apr 4]. Available from: http://bvsms.saude.gov.br/bvs/publicacoes/%20demografia_medica_brasil_2025.pdf.
9. Toy S, Daly Guris RJ. How to conduct survey-based research. *Anaesthesia*. 2023;78:902–5.
10. Sharma A, Minh Duc NT, Luu Lam Thang T, Nam NH, Ng SJ, Abbas KS, et al. A Consensus-Based Checklist for Reporting of Survey Studies (CROSS). *J Gen Intern Med*. 2021;36:3179–87.
11. Story D, Tait A. Readers' toolbox Understanding Research Methods. *Anesthesiology*. 2019;130:192–202.
12. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–81.
13. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Informatics*. 2019;95:103208.

14. Hugo A, Gutierrez A, Hugo M, Gutierrez A. Package ‘samplesize4surveys’. 2022. [Internet][cited 2026 Apr 6]. Available from: <https://cran.r-project.org/web/packages/samplesize4surveys/samplesize4surveys.pdf>.
15. Yang LQ, Zhu L, Shi X, Miao CH, Yuan HB, Liu ZQ, et al. Postoperative pulmonary complications in older patients undergoing elective surgery with a supraglottic airway device or tracheal intubation. *Anaesthesia*. 2023;78:953–62.
16. Forestell B, Ramsden S, Sharif S, Centofanti J, Al Lawati K, Fernando SM, et al. Supraglottic Airway Versus Tracheal Intubation for Airway Management in Out-of-Hospital Cardiac Arrest: A Systematic Review, Meta-Analysis, and Trial Sequential Analysis of Randomized Controlled Trials. *Crit Care Med*. 2024;52:E89–99.
17. Ullmann H, Renziehausen L, Geil D, Sponholz C, Thomas-Rüddel D, Völker MT, et al. The Influence of Positive End-Expiratory Pressure on Leakage and Oxygenation Using a Laryngeal Mask Airway: A Randomized Trial. *Anesth Analg*. 2022;135(4):769–76.
18. Hammer M, Santer P, Schaefer MS, Althoff FC, Wongtangman K, Frey UH, et al. Supraglottic airway device versus tracheal intubation and the risk of emergent postoperative intubation after general anaesthesia in adults: a retrospective cohort study. *Br J Anaesth*. 2021;126:738–45.
19. Yan C, Zhang Y, Chen Y, Qv Z, Zuo M. Comparison of SaCoVLM™ video laryngeal mask-guided intubation and i-gel combined with flexible bronchoscopy-guided intubation in airway management during general anesthesia: a non-inferiority study. *BMC Anesthesiol*. 2022;22:302.
20. Govardhane BT, Shinde AD, Gehdoo RP, Arora S. Current practice pattern among anaesthesiologists for difficult airway management: A nationwide cross-sectional survey. *Indian J Anaesth*. 2023;67:809–14.
21. Krug N, Thieme V, Voelker MT, Bercker S. Results of an international survey on the use and perceptions of the laryngeal mask airway (LMA). Luong CQ, editor. *PLOS One*. 2025;20:e0321421.
22. He Y, Zhang Z, Li R, et al. National survey on the current status of airway management in China. *Sci Rep*. 2024;14:15627.
23. Goodwin L, Samuel K, Schofield B, et al. Airway management during in-hospital cardiac arrest in adults: UK national survey and interview study with anaesthetic and intensive care trainees. *J Intensive Care Soc*. 2021;22:192–7.

ORIGINAL INVESTIGATION

Clinical impact of extubation in the operating room after cardiac surgery: a retrospective analysis of a prospective registry



Patricia Silva de Marco *, Pedro Henrique Romani Lauand , Luiz Carlos Volpi Júnior ,
Raphael Fernandes Cardoso Vianna Alves , Maria Paula Belini , Lilia Nigro Maia ,
Maurício Nassau Machado ¹, Marcelo Arruda Nakazone ¹

Faculdade de Medicina de São José do Rio Preto (FAMERP), Departamento de Cardiologia e Cirurgia Cardiovascular, São José do Rio Preto, SP, Brazil

Received 23 August 2025; accepted 9 April 2026
Available online 24 April 2026

KEYWORDS

Acute kidney injury;
Airway extubation;
Cardiac surgery
procedures;
Coronary artery
bypass;
Mortality;
Propensity score

Abstract

Background: The practice of immediate extubation in the Operating Room (OR) after cardiac surgery involving cardiopulmonary bypass remains debated. Concerns persist regarding its safety profile compared to conventional fast-track weaning in the Intensive Care Unit (ICU).

Methods: This retrospective analysis of a prospective single-center registry included 846 adult patients undergoing isolated Coronary Artery Bypass Grafting (CABG) or Heart Valve Surgery (HVS). To rigorously account for indication bias, we employed a doubly robust estimation approach, combining Stabilized Inverse Probability of Treatment Weighting (SIPTW) with multi-variable regression adjustment for residual confounding.

Results: Of the 846 patients, 115 (13.6%) were extubated in the OR and 731 (86.4%) in the ICU. The weighted analysis successfully balanced baseline characteristics. Extubation in the OR was associated with a significant reduction in 30-day mortality (adjusted Hazard Ratio [aHR] = 0.16; 95% CI 0.03–0.92; $p = 0.040$) and acute kidney injury (adjusted Odds Ratio [aOR] = 0.55; 95% CI 0.31–0.99; $p = 0.046$). Additionally, it reduced the likelihood of prolonged ICU stay (> 14 days) (aOR = 0.10; $p = 0.032$). Contrary to unadjusted analyses, there was no significant association with respiratory tract infection ($p = 0.163$), reintubation ($p = 0.103$), or other adverse events.

Conclusion: In this single-center, robustly adjusted analysis, extubation in the OR was associated with favorable outcomes, including lower 30-day mortality and reduced resource utilization in highly selected patients. Previous concerns regarding increased reintubation rates were not substantiated.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail: psilvademarco@gmail.com (P.S. de Marco).

¹ These authors contributed equally to this work and share Senior Authorship.

Introduction

Advances in anesthetic, surgical and Cardiopulmonary Bypass (CPB) techniques have significantly reduced duration of Mechanical Ventilation (MV) and clinical complications after cardiac surgeries.^{1,2}

Early extubation (typically defined as within 6 hours of Intensive Care Unit [ICU] arrival) is now a well-established quality metric that optimizes resource utilization and promotes faster recovery.^{3,4} However, the practice of immediate extubation in the Operating Room (OR) remains a subject of debate.

While OR extubation has been demonstrated to be feasible in selected low-risk populations, such as those undergoing Off-Pump Coronary Artery Bypass Grafting (OPCABG) or minimally invasive procedures,⁵⁻⁸ its safety and efficacy in a broader spectrum of cardiac surgery patients requiring CPB are less established. Proponents argue that avoiding ICU ventilation minimizes exposure to positive pressure ventilation and sedation, potentially reducing length of stay and costs.^{9,10} Conversely, critics caution that immediate extubation may mask early postoperative instability, potentially increasing the risk of respiratory failure, reintubation, and bleeding complications.^{11,12} Recent observational studies using propensity-score matching have yielded conflicting results, with some reporting safety and others suggesting increased hazards.¹²

A critical limitation of prior studies is the potential for indication bias, where healthier patients are preferentially selected for OR extubation, confounding the assessment of true clinical impact. To address this gap, we conducted a retrospective analysis of a prospectively collected registry using a doubly robust statistical approach.

The primary objective of this study was to evaluate the association between OR extubation and 30-day all-cause mortality in adult patients undergoing isolated Coronary Artery Bypass Grafting (CABG) or Heart Valve Surgery (HVS). Our prespecified primary hypothesis was that extubation in the OR is a feasible approach associated with satisfactory clinical results and reduced resource utilization when compared to conventional ICU extubation, after rigorous adjustment for baseline risk and operative characteristics.

Methods

Patients

This single-center study was designed as a retrospective analysis of data from a prospectively maintained registry. We evaluated all adult patients (≥ 18 years) who underwent isolated CABG or HVS in a tertiary hospital between September 2018 and August 2023 (60 months). Patients who underwent aortic surgeries, combined procedures, congenital heart disease correction, or heart transplantation were excluded. The study was conducted in accordance with the Declaration of Helsinki and approved by the local research ethics committee (CAAE: 81704424.0.0000.5415), which waived the need for individual informed consent due to the retrospective nature of the analysis.

Demographic and laboratory data were obtained from electronic medical records. The risk of postoperative death was assessed using the InsCor score, a locally validated risk model that includes 10 variables, each one with a different score: age > 70 years, female sex, CABG combined with HVS, myocardial infarction < 90 days, reoperation, aortic valve surgery, tricuspid valve surgery, baseline creatinine level $> 2 \text{ mg.dL}^{-1}$, left ventricular ejection fraction $< 30\%$ and critical preoperative state.¹³

The primary outcome was all-cause mortality within 30 days. Secondary outcomes included Acute Kidney Injury (AKI), defined according to the Kidney Disease: Improving Global Outcomes (KDIGO) criteria,¹⁴ respiratory tract infection, reintubation, ICU readmission, and prolonged ICU length of stay (> 14 days). The outcome “mechanical ventilation > 24 hours” was excluded from the comparative analysis to avoid structural bias relative to the intervention.

Postoperative management

Postoperative care followed a standardized institutional protocol applied uniformly to all patients, regardless of the extubation location. For patients admitted to the ICU requiring mechanical ventilation, sedation was typically maintained with propofol or dexmedetomidine and titrated to achieve a light sedation target (Richmond Agitation-Sedation Scale [RASS] -1 to 0) to facilitate neurological assessment and weaning. Patients extubated in the OR received supplemental oxygen and standard non-invasive monitoring. The decision of extubation in the OR was made by consensus of the attending anesthesiologist and surgeon based on real-time clinical assessment, reflecting routine practice rather than a predefined protocol. Analgesia for both groups was managed using a multimodal approach, including intravenous dipyrone and opioids as needed to maintain a Visual Analog Scale (VAS) score < 4 . A dedicated physiotherapy team initiated respiratory rehabilitation and early mobilization protocols on the first postoperative day, contingent on hemodynamic stability. This standardized pathway ensured that differences in outcomes were not attributable to variations in post-extubation nursing or physiotherapy care.

Missing data handling

The dataset was complete for all baseline covariates included in the propensity score model and outcome variables. Specifically, there were no missing values for age, sex, comorbidities, renal function, or operative times across the 846 patients included in the final analysis. Therefore, complete-case analysis was performed without the need for imputation methods.

Statistical analysis

Continuous variables were assessed for normality using the Shapiro-Wilk test and visual inspection of histograms. To ensure uniformity in data presentation, all continuous variables are reported as median (Interquartile Range, IQR) and were compared using the Mann-Whitney *U*-test. Categorical variables are presented as absolute numbers (*n*) and percentages (%) and compared using the Chi-Square test or Fisher's exact test. To address potential confounding by indication

and baseline differences between the groups (OR extubation vs. ICU extubation), we employed a doubly robust estimation approach utilizing Stabilized Inverse Probability of Treatment Weighting (SIPTW). A propensity score for the likelihood of being extubated in the OR was calculated using a multivariable logistic regression model including the following covariates: age (years), sex (reference: male), Chronic Obstructive Pulmonary Disease (COPD), hypertension, diabetes mellitus, baseline estimated Glomerular Filtration Rate (eGFR), calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) 2021 equation,¹⁵ left ventricular function (reference: Left Ventricular Ejection Fraction [LVEF] $\geq 50\%$), urgent/emergent surgery, type of surgery (reference: CABG), on-pump status, and duration of CPB (minutes). Stabilized weights were calculated to minimize the influence of extreme weights. Covariate balance after weighting was assessed using Standardized Mean Differences (SMD), with an SMD < 0.10 considered indicative of negligible imbalance (Supplementary Table S1).

Outcome analyses were performed on the weighted population using the Complex Samples module of IBM SPSS Statistics v.27 (IBM Corp, Armonk, NY). For the primary outcome (30-day mortality), a weighted Cox proportional hazards regression model was used. For secondary outcomes, weighted binary logistic regression models were applied. To ensure robustness against residual confounding, particularly for age, all baseline covariates used in the propensity score were also included as adjusters in the final outcome models (doubly robust adjustment). Results are presented as adjusted Hazard Ratios (aHR) or adjusted Odds Ratios (aOR) with 95% Confidence Intervals (95% CI). A two-tailed p-value < 0.05 was considered statistically significant.

Results

Study population and propensity score weighting

A total of 1,078 patients were assessed for eligibility. After excluding 232 patients (21.5%) primarily due to multi-procedure or aortic surgeries, the final study population comprised 846 patients (Fig. 1). Of these, 115 (13.6%) were

extubated in the OR and 731 (86.4%) in the ICU. In the original unadjusted cohort, patients extubated in the OR were significantly younger (median 60 vs 63 years, $p = 0.020$), had a lower prevalence of hypertension (66.1% vs 74.8%, $p = 0.048$), better baseline renal function (median eGFR 77 vs. 69 $\text{mL}\cdot\text{min}^{-1}/1.73\text{ m}^2$, $p = 0.001$), and shorter duration of CPB (median 80 vs. 89 min, $p < 0.001$) compared to the ICU group (Table 1).

To account for baseline imbalances and potential indication bias, SIPTW was applied. The weighting procedure successfully balanced the covariates between groups, reducing the SMD to < 0.10 for critical variables, including renal function and CPB duration (Supplementary Table S1). A slight residual imbalance remained for age, which was subsequently adjusted for using a doubly robust approach in the final outcome models. The weighted sample size ($n = 843$) represents the sum of stabilized weights and differs slightly from the original sample size ($n = 846$) due to statistical adjustments; no patients were excluded from the analysis. The baseline characteristics for both the unadjusted and weighted cohorts are presented in Table 1.

Primary outcome: 30-day mortality

The unadjusted intra-hospital mortality rate within 30 days was 1.7% (2/115) in the OR extubation group versus 6.6% (48/731) in the ICU extubation group ($p = 0.041$) (Table 2). In an exploratory unadjusted analysis, non-survivors were significantly older, had worse renal function, and longer CPB times compared to survivors (Supplementary Tables S2–S6). After applying SIPTW and adjusting for all baseline covariates (doubly robust Cox regression), extubation in the OR was associated with a significant reduction in the risk of 30-day all-cause mortality (aHR = 0.16; 95% CI 0.03 to 0.92; $p = 0.040$) (Table 3).

Subgroup analysis

In the subgroup of patients undergoing isolated CABG ($n = 446$), no deaths occurred in the OR extubation arm (0/58, 0.0%), preventing the calculation of a specific hazard ratio due to perfect separation of data. In the HVS subgroup ($n = 397$), mortality was 3.9% (2/51) in the OR group versus

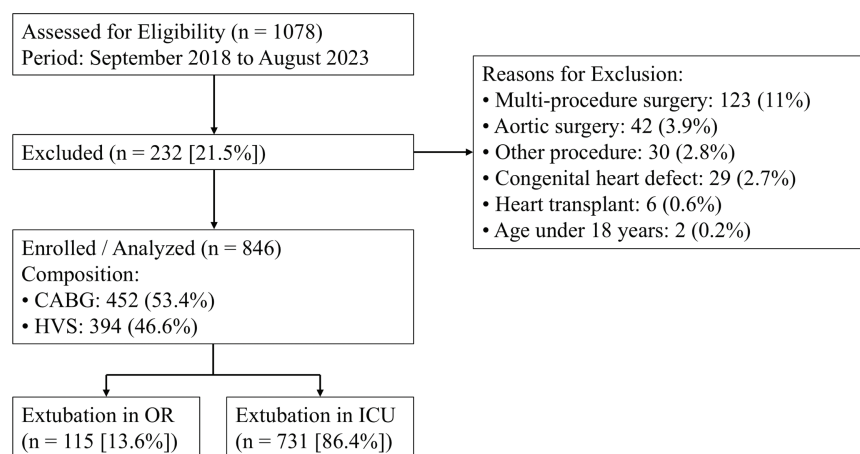


Figure 1 Flow diagram of the study population selection process. Note: CABG, Coronary Artery Bypass Grafting; HVS, Heart Valve Surgery; OR, Operating Room; ICU, Intensive Care Unit. Values are presented as number (n) and percentage (%).

Table 1 Baseline demographic and clinical characteristics of the study population before and after Stabilized Inverse Probability of Treatment Weighting (SIPTW).

	Overall n = 846	Unadjusted Cohort		Weighted Cohort (SIPTW)			
		OR extubation n = 115	ICU extubation n = 731	p-value	OR extubation n = 108	ICU extubation n = 735	p-value
Demographics & Risk							
Age (years)	62 (55–69)	60 (51–67)	63 (56–70)	0.02	60 (51–67)	62 (55–70)	0.024
Male sex	578 (68.3)	73 (63.5)	505 (69.1)	0.23	(66.0)	(67.6)	0.704
Body mass index (kg.m ⁻²)	27 (25–31)	27 (25–31)	27 (24–30)	0.434	28 (26–31)	27 (25–30)	0.136
InsCor	3 (2–5)	2 (2–5)	4 (2–7)	< 0.001	2 (2–4)	4 (2–6)	< 0.001
Comorbidities							
Chronic obstructive pulmonary disease	39 (4.6)	4 (3.5)	35 (4.8)	0.534	(2.7)	(4.3)	0.608
Hypertension	623 (73.6)	76 (66.1)	547 (74.8)	0.048	(74.7)	(73.7)	0.781
Diabetes mellitus	298 (35.2)	39 (33.9)	259 (35.4)	0.751	(36.0)	(34.9)	0.794
Cardiac & Renal Status							
Left ventricular ejection fraction							
> 50%	632 (74.7)	94 (81.7)	538 (73.6)	0.062	(73.2)	(75.3)	0.64
31% to 50%	164 (19.4)	18 (15.7)	146 (20.0)	0.276	(24.1)	(18.8)	0.196
≤ 30%	50 (5.9)	3 (2.6)	47 (6.4)	0.106	(2.7)	(5.9)	0.188
Baseline SCr (mg.dL ⁻¹)	1.10 (0.90–1.34)	1.10 (0.90–1.20)	1.10 (0.91–1.40)	0.003	1.10 (0.99–1.27)	1.10 (0.90–1.36)	0.984
Baseline e-GFR (CKD-EPI 2021; mL.min ⁻¹ /1.73 m ²)	70 (54–87)	77 (61–91)	69 (53–86)	0.001	71 (54–84)	70 (54–88)	0.726
Operative data							
Surgery incidence							
Previous cardiac surgery	78 (9.2)	7 (6.1)	71 (9.7)	0.212	(1.3)	(10.6)	0.001
Surgical Priority							
Urgent/Emergent surgery	350 (41.4)	32 (27.8)	318 (43.5)	0.002	(43.2)	(40.7)	0.576
Isolated CABG	452 (53.4)	43 (37.4)	409 (56.0)	< 0.001	(53.2)	(52.9)	0.956
On-pump surgery	826 (97.6)	99 (86.1)	727 (99.5)	< 0.001	(97.6)	(97.3)	1.000
CPB (min)	88 (75–102)	80 (68–95)	89 (76–104)	< 0.001	91 (74–101)	87 (74–102)	0.779

Footnotes: Values are presented as n (%) for categorical data or as median (IQR) for continuous variables. SIPTW, Stabilized Inverse Probability of Treatment Weighting; OR, Operating Room; ICU, Intensive Care Unit; SCr, Serum Creatinine; eGFR, Estimated Glomerular Filtration Rate (CKD-EPI); CABG, Coronary Artery Bypass Grafting; CPB, Cardiopulmonary Bypass.

*The p-values for the weighted cohort are derived from Complex Samples General Linear Models. Note: The weighted sample size (n = 843) represents the sum of stabilized weights and differs slightly from the original sample size (n = 846) due to statistical adjustments.

Table 2 Unadjusted clinical complications, ICU stay, and mortality of patients undergoing cardiac surgery according to extubation location.

	Overall n = 846	OR extubation n = 115	ICU extubation n = 731	p-value
Postoperative outcomes				
Acute kidney injury (KDIGO) ^a	446 (52.7)	46 (40.0)	400 (54.7)	0.003
KDIGO 1	373 (44.1)	43 (37.4)	330 (45.1)	0.120
KDIGO 2	36 (4.3)	3 (2.6)	33 (4.5)	0.460
KDIGO 3	37 (4.4)	0 (0.0)	37 (5.1)	0.014
RRT up to 7 days	14 (1.7)	0 (0.0)	14 (1.9)	0.237
Reoperation for bleeding/tamponade	24 (2.8)	1 (0.9)	23 (3.1)	0.234
Acute atrial fibrillation	137 (16.2)	17 (14.8)	120 (16.4)	0.659
Respiratory infection	188 (22.2)	14 (12.2)	174 (23.8)	0.005
Tracheal reintubation up to 7 days	32 (3.8)	2 (1.7)	30 (4.1)	0.296
Deep sternal wound infection	37 (4.4)	3 (2.6)	34 (4.7)	0.319
Type 1 neurological injury	36 (4.3)	4 (3.5)	32 (4.4)	0.807
Discharge and mortality				
CS-ICU readmission	48 (5.7)	4 (3.5)	44 (6.0)	0.274
CS-ICU LOS up to 30 days	4 (3 - 6)	3 (2 - 4)	4 (3 - 6)	< 0.001
Long LOS (> 14 days)	48 (5.7)	1 (0.9)	47 (6.4)	0.017
30-day mortality	50 (5.9)	2 (1.7)	48 (6.6)	0.041

Footnotes: Values are presented as n (%) for categorical variables and median (IQR) for continuous variables. The p-values calculated using Chi-Square test, Fisher's exact test, or Mann-Whitney U-test, as appropriate. These represent unadjusted comparisons.

^a Acute Kidney Injury defined according to KDIGO (Kidney Disease: Improving Global Outcomes definition and staging) criteria.

8.4% (29/346) in the ICU group. While a formal statistical test of interaction between extubation location and surgery type was limited by the absence of events in the CABG-OR arm, the direction of benefit appeared consistent across procedural types, with no mortality events observed in the revascularization group, although the small sample size precludes definitive safety conclusions (Table S7–S8).

Secondary outcomes

Results of the doubly robust logistic regression analyses for secondary outcomes are presented in Table 3. OR extubation

was associated with a significant reduction in the odds of developing AKI (KDIGO stage 1–3) (aOR = 0.55; 95% CI 0.31 to 0.99; p = 0.046) and a significant reduction in the likelihood of prolonged ICU stay (> 14 days) (aOR = 0.10; 95% CI 0.01 to 0.82; p = 0.032).

Regarding safety outcomes, there was no statistically significant association between OR extubation and the risk of reintubation within 7 days (aOR = 0.26; 95% CI 0.05 to 1.31; p = 0.103), postoperative atrial fibrillation (aOR = 1.11; 95% CI 0.58 to 2.14; p = 0.758), or ICU readmission (aOR = 0.37; 95% CI 0.11 to 1.31; p = 0.124). Unlike the unadjusted analysis, the robust adjusted model showed no statistically significant

Table 3 Independent predictors of 30-day mortality and postoperative complications: doubly robust analysis.

Outcome	Adjusted Effect Size ^a (95% CI)	p-value
Primary Outcome		
30-Day Mortality (aHR)	0.16 (0.03 – 0.92)	0.040
Secondary Outcomes (Renal & Utilization)		
Acute Kidney Injury (aOR)	0.55 (0.31 – 0.99)	0.046
Prolonged ICU Stay > 14 days (aOR)	0.10 (0.01 – 0.82)	0.032
ICU Readmission (aOR)	0.37 (0.11 – 1.31)	0.124
Secondary Outcomes (Safety)		
Respiratory Tract Infection (aOR)	0.59 (0.28 – 1.24)	0.163
Reintubation < 7 days (aOR) ^a	0.26 (0.05 – 1.31)	0.103
Atrial Fibrillation (aOR)	1.11 (0.58 – 2.14)	0.758

Footnotes: Values represent adjusted Hazard Ratio (aHR) for mortality (Cox Regression) and adjusted Odds Ratio (aOR) for secondary outcomes (Logistic Regression). All models were weighted by SIPTW (Stabilized Inverse Probability of Treatment Weighting) and further adjusted for the following covariates: Age, Sex, COPD, Hypertension, Diabetes, Baseline eGFR, Left Ventricular Function, Urgency of Surgery, Type of Surgery (CABG vs. Valve), On-pump status, and CPB duration. Doubly robust estimation combines SIPTW with regression adjustment for the same covariates, providing consistent estimates if either the propensity score or outcome model is correctly specified.

^a Due to the low number of reintubation events (n = 32), estimates should be interpreted with caution.
CI, Confidence Interval.

difference in the risk of respiratory tract infection (aOR = 0.59; 95% CI 0.28 to 1.24; $p = 0.163$), suggesting that previous unadjusted differences were likely driven by baseline patient characteristics (Table 3).

Discussion

This study evaluated the impact of OR extubation on clinical outcomes and 30-day mortality in 846 patients undergoing isolated CABG or HVS. Using a doubly robust analytical approach that combined SIPTW with covariate adjustment, our findings suggest that extubation in the OR may be associated with a significant reduction in 30-day all-cause mortality, AKI, and prolonged ICU length of stay. Conversely, after rigorous adjustment for baseline differences, we found no significant association between OR extubation and respiratory tract infection, reintubation rates, or other adverse events, suggesting that the apparent protective effect against infection observed in our unadjusted analysis – and potentially in prior observational research – was likely driven by baseline selection bias rather than the intervention itself.

Our primary finding – a significant association with reduced 30-day mortality (aHR = 0.16; 95% CI 0.03–0.92; $p = 0.040$) – is noteworthy. Although a doubly robust adjustment was employed to mitigate selection bias, the observed magnitude of effect may partly reflect that successful OR extubation acts as a surrogate marker for optimal intraoperative physiology and surgical performance. Nevertheless, the potential physiological benefits of avoiding prolonged positive pressure ventilation cannot be disregarded as a contributing factor to these favorable outcomes.¹⁶ Notably, in our subgroup analysis of isolated CABG patients, no deaths occurred in the OR extubation group (0/58). While this finding points toward the feasibility of the strategy in revascularization procedures, the small sample size and absence of events preclude definitive safety conclusions. Although formal statistical testing for interaction was limited by this “zero-event” rate in the intervention arm, the direction of benefit appeared consistent across procedure types.

A critical contribution of our study is the refutation of recent safety concerns. A large propensity-matched analysis by Hawkins et al.¹² suggested that OR extubation was associated with increased rates of reintubation and reoperation for bleeding. In contrast, our doubly robust analysis found no significant association with reintubation (aOR = 0.26; 95% CI 0.05–1.30; $p = 0.103$) (Table 3). Furthermore, we observed no signal of increased risk for adverse bleeding events requiring reoperation (0.9% vs. 3.1%, $p = 0.234$) (Table 2), differing from the findings of Hawkins et al.¹² Our findings align more closely with James et al.³ and e Silva et al.,⁹ who reported that immediate extubation does not compromise safety when performed in selected patients. The absence of increased bleeding or reintubation in our cohort likely reflects the careful intraoperative selection criteria (hemodynamic stability, normothermia, and hemostasis) used by our team, supporting the feasibility of this approach in tertiary centers.

A key finding of our doubly robust analysis is that the protective effect against respiratory infections observed in the unadjusted data (Table 2) was no longer statistically

significant after adjustment (aOR = 0.59; 95% CI 0.28–1.24; $p = 0.163$) (Table 3). In our initial univariable comparison, OR extubation appeared to significantly reduce the risk of infection. However, the robust SIPTW adjustment revealed that this association was likely influenced by the healthier baseline profile of patients selected for OR extubation (younger, better renal function). This result is methodologically reassuring and consistent with the randomized trial by Nicholson et al.,¹⁷ which found no difference in pulmonary function between early and delayed extubation. This underscores the importance of using advanced statistical methods to disentangle true clinical effects from baseline characteristics.¹⁸

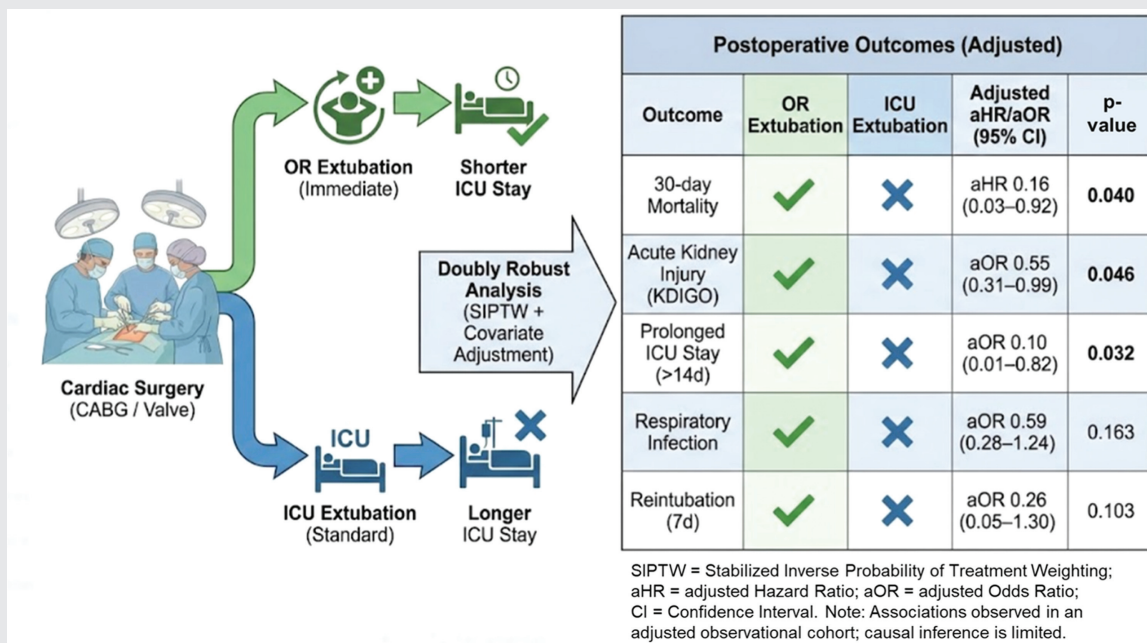
Consistent with the literature on fast-track protocols,^{4,10} OR extubation was strongly associated with reduced resource utilization, specifically a 90% reduction in the odds of prolonged ICU stay (> 14 days). Furthermore, we observed a reduction in AKI (aOR = 0.55; 95% CI 0.31–0.99; $p = 0.046$), which persisted even after adjusting for CPB duration and baseline function (Table 3). This suggests a potential hypothesis that the preservation of spontaneous ventilation and avoidance of positive pressure might contribute to improved renal perfusion, a hypothesis supported by prior physiological studies.⁶ While our findings are consistent with this physiological rationale, further prospective studies are required to confirm this mechanism.

Limitations

Our study has limitations inherent to its retrospective design. Despite robust statistical adjustment using SIPTW and doubly robust estimation, residual confounding by unmeasured variables (e.g., intraoperative ease of separation from bypass, specific surgeon preference) cannot be completely excluded. Second, the specific subgroup analysis for isolated CABG was limited by the absence of mortality events in the OR extubation arm, preventing the calculation of specific hazard ratios for that subset. Finally, the decision to extubate was based on clinical consensus rather than a rigid protocol, which introduces variability but also reflects real-world clinical judgement. We also acknowledge that urine output criteria for AKI were not available for all patients, limiting our definition to serum creatinine changes. Furthermore, although our weighting approach balanced patient acuity, the study period (2018–2023) encompassed the COVID-19 pandemic and potential variations in practice over time. Specifically, the COVID-19 pandemic period (2020–2021) may have affected patient selection, surgical volume, and postoperative management protocols, introducing a possibility of temporal confounding that could not be fully adjusted for. Finally, we acknowledge that the primary outcome of 30-day mortality was based on a low number of events in the OR extubation group. Consequently, the mortality estimates and its associated adjusted hazard ratio (95% CI 0.03–0.92) should be interpreted with caution due to the uncertainty inherent in small event counts and the resulting wide confidence interval.

Conclusion

This study indicates that, in a highly selected population, extubation in the OR is a safe strategy associated with



Central Illustration: Associations between operating room extubation on clinical outcomes.

reduced 30-day mortality, lower risk of AKI, and shorter ICU stay in selected patients undergoing cardiac surgery. Concerns regarding increased reintubation or bleeding were not substantiated in this analysis. These findings from a tertiary center support the feasibility of OR extubation for carefully selected candidates.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Gemini, a large language model from Google, in order to improve language and readability, assist in structuring the manuscript sections (Abstract, Introduction, Methods, Results, and Discussion), summarize statistical results into text, and refine the overall narrative clarity. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Authors' contributions

Patricia Silva de Marco: Substantial contributions to the conception of the work; drafting the work and critical

review of important intellectual content; acquisition and analysis of data for the study; consent to be responsible for all aspects of the work, ensuring that issues related to the accuracy or integrity of any part of the work are properly investigated and resolved.

Pedro Henrique Romani Lauand: Substantial contributions to the conception of the work; acquisition and analysis of data for the study.

Luiz Carlos Volpi Júnior: Substantial contributions to the conception of the work; acquisition and analysis of data for the study.

Raphael Fernandes Cardoso Vianna Alves: Substantial contributions to the conception of the work; acquisition and analysis of data for the study.

Maria Paula Belini: Substantial contributions to the conception of the work; acquisition and analysis of data for the study.

Lilia Nigro Maia: Substantial contributions to the conception of the work; critical review of important intellectual content.

Mauricio Nassau Machado: Substantial contributions to the conception of the work; drafting the work and critical review of important intellectual content; acquisition and analysis of data for the study; critical review of important intellectual content; final approval of the version to be published.

Marcelo Arruda Nakazone: Substantial contributions to the conception of the work; Drafting the work and critical review of important intellectual content; Acquisition and analysis of data for the study; Critical review of important intellectual content; Final approval of the version to be published.

Conflicts of interest

The authors declare no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjane.2026.844761.

Associate Editor

João Manoel Silva Jr.

References

- Wong W-T, VK Lai, Chee YE, Lee A. Fast-track cardiac care for adult cardiac surgical patients. *Cochrane Database Syst Rev*. 2016;9:CD003587.
- Chan JL, Miller JG, Murphy M, Greenberg A, Iraola M, Horvath KA. A Multidisciplinary Protocol-Driven Approach to Improve Extubation Times After Cardiac Surgery. *Ann Thorac Surg*. 2018;105:1684–90.
- James L, Smith DE, Galloway AC, et al. Routine Extubation in the Operating Room After Isolated Coronary Artery Bypass. *Ann Thorac Surg*. 2024;117:87–94.
- Grant MC, Crisafi C, Alvarez A, et al. Perioperative Care in Cardiac Surgery: A Joint Consensus Statement by the Enhanced Recovery After Surgery (ERAS) Cardiac Society, ERAS International Society, and The Society of Thoracic Surgeons (STS). *Ann Thorac Surg*. 2024;117:669–89.
- Horswell JL, Herbert MA, Prince SL, Mack MJ. Routine Immediate Extubation After Off-Pump Coronary Artery Bypass Surgery: 514 Consecutive Patients. *J Cardiothorac Vasc Anesth*. 2005;19:282–7.
- Dorsa AG, Rossi AI, Thierer J, et al. Immediate Extubation After Off-Pump Coronary Artery Bypass Graft Surgery in 1,196 Consecutive Patients: Feasibility, Safety and Predictors of When Not to Attempt It. *J Cardiothorac Vasc Anesth*. 2011;25:431–6.
- Nagre AS, Jambures NP. Comparison of Immediate Extubation Versus Ultrafast Tracking Strategy in the Management of Off-Pump Coronary Artery Bypass Surgery. *Ann Card Anaesth*. 2018;21:129–33.
- Kandasamy A, Ramalingam S, Simon H, Arumugham S, Reddy B, Krupananda H. Ultra fast-tracking versus a conventional strategy in valve replacement surgery. *Indian J Anaesth*. 2013;57:298.
- E Silva RAG, Borgomoni GB, Maia ADS, et al. Extubation in the Operating Room After Coronary Artery Bypass Graft Surgery Reduces Hospital Stay. *J Cardiothorac Vasc Anesth*. 2023;37:1938–45.
- Carnero-Alcázar M, Beltrao-Sial R, Montero-Cruces L, et al. Ultrafast track versus conventional fast track in patients undergoing cardiac surgery: a propensity score-matched analysis. *Interdiscip Cardiovasc Thorac Surg*. 2023;37:ivad143.
- Montes FR, Sanchez SI, Giraldo JC, et al. The Lack of Benefit of Tracheal Extubation in the Operating Room After Coronary Artery Bypass Surgery. *Anesth Analg*. 2000;91:776–80.
- Hawkins AD, Strobel RJ, Mehaffey JH, et al. Operating Room Versus Intensive Care Unit Extubation Within 6 Hours After On-Pump Cardiac Surgery: Early Results and Hospital Costs. *Semin Thorac Cardiovasc Surg*. 2024;36:195–208.
- Mejía OAV, Lisboa LAF, Puig LB, et al. InsCor: A Simple and Accurate Method for Risk Assessment in Heart Surgery. *Arq Bras Cardiol*. 2013;100:246–54.
- Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group (2012) KDIGO Clinical Practice Guideline for Acute Kidney Injury – Summary of Recommendation Statements. *Kidney Int Suppl*. 2012;2:8–12.
- Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med*. 2021;385:1737–49.
- Bouabdallaoui N, Stevens SR, Doenst T, et al. Impact of Intubation Time on Survival following Coronary Artery Bypass Grafting: Insights from the Surgical Treatment for Ischemic Heart Failure (STICH) Trial. *J Cardiothorac Vasc Anesth*. 2018;32:1256–63.
- Nicholson DJ, Kowalski SE, Hamilton GA, Meyers MP, Serrette C, Duke PC. Postoperative pulmonary function in coronary artery bypass graft surgery patients undergoing early tracheal extubation: A comparison between short-term mechanical ventilation and early extubation. *J Cardiothorac Vasc Anesth*. 2002;16:27–31.
- Shahsanaei F, Behrooz S, Petrudi NR, Khajehbahrami F. The Overall Prevalence and Main Determinants of Prolonged Mechanical Ventilation in Patients Undergoing Coronary Artery Bypass Grafting: A Systematic Review. *Heart Views*. 2023;24:188–93.

ORIGINAL INVESTIGATION

Effects of propofol versus sevoflurane induction on echocardiographic parameters in patients with mitral stenosis: a randomized clinical trial



Mohammed Jaffer Sherif ^a, Saravana Babu MS ^{b,*}, Prasanta Kumar Dash ^a,
Nagarjuna Panidappu ^c, Subin Sukesan ^a, Thomas Koshy ^a

^a Sree Chitra Tirunal Institute for Medical Sciences and Technology, Division of Cardiothoracic and Vascular Anesthesia, Trivandrum, India

^b All India Institute of Medical Sciences, Department of Anesthesiology, Madurai, India

^c Amrita Institute of Medical Sciences, Department of Cardiothoracic and Vascular Anesthesia, Kochi, India

Received 22 July 2025; accepted 15 May 2026

Available online 3 June 2026

KEYWORDS

Anesthetics;
Inhalation;
Anesthetics;
Intravenous;
Echocardiography;
Mitral valve stenosis;
Propofol;
Sevoflurane

Abstract

Background: The main aim of this study was to compare the effects of propofol and sevoflurane induction on the echocardiographic parameters of Mitral Stenosis (MS) patients.

Methods: Prospective, randomized, outcome-assessor-blinded clinical trial in 80 adults with MS undergoing mitral valve replacement. Patients were randomized to receive General Anesthesia (GA) induction with either propofol (Group P) or sevoflurane (Group S). The primary objective was to assess the changes in mean gradient after 3 minutes of GA induction. Secondary objectives were to assess the changes in echocardiographic Doppler parameters and hemodynamic variables.

Results: Compared to pre-induction values, there was significant reduction in the peak and mean gradients as well as peak velocity, in both the groups after GA induction. The Mitral Valve Area calculated by continuity equation (MVA-c) and by pressure half time (MVA-p) were increased in Group S after GA induction. On comparison between the two groups, Group S showed lower pressure half time and greater MVA-p ($p = 0.0099$) and MVA-c ($p = 0.0316$) than Group P after GA induction. Mean arterial pressure decreased at the 1st and 3rd minute following GA induction in both groups. The Heart Rate (HR) increased at the 1st and 3rd minute in Group P whereas it reduced at both the time points in group S.

Conclusion: Even though both anesthetic agents reduced transvalvular gradients in MS patients, sevoflurane was associated with a greater increase in estimated valve area and a reduction in heart rate compared with propofol. These findings suggest more favorable short-term hemodynamic and echocardiographic changes with sevoflurane at the studied time points.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail: saravanababu4u@gmail.com (S. Babu MS).

Mitral Stenosis (MS) is one of the most common valvular heart diseases and the major manifestation of rheumatic fever, the incidence of which is declining in developed countries.¹ Echocardiographic modalities such as M-mode, two-dimensional echo, three-dimensional echo, and Doppler evaluation are utilized for assessment of the severity of the lesion.² The echocardiographic parameters determining the severity of mitral valve stenosis are greatly influenced by the hemodynamic status of the patient.³ Patients with MS are subjected to General Anesthesia (GA) for various reasons like transesophageal echocardiographic evaluation, percutaneous balloon mitral valvotomy, and elective or emergency Mitral Valve Replacement (MVR).⁴ A clear understanding of cardiovascular dynamics in Mitral Stenosis (MS) is essential when selecting an anesthetic regimen for general anesthesia induction. Progressive narrowing of the mitral valve impairs left ventricular filling, elevating left atrial pressure and leading to atrial dilatation; although disease severity is defined by valve area, symptoms are primarily driven by left atrial pressure. Factors that increase transvalvular gradients, such as tachycardia or increased cardiac output, can precipitate pulmonary congestion and edema.^{5,6} Chronic left ventricular underfilling, compounded by interventricular septal shift from right ventricular hypertrophy, may result in systolic dysfunction.⁷ Because cardiac output in MS is fixed and preload dependent, induction of anesthesia is associated with significant hemodynamic instability. Therefore, anesthetic goals include avoiding tachycardia and preserving preload, afterload, and left ventricular contractility. Propofol and sevoflurane are commonly used agents for GA induction; however, their effects on the hemodynamic and echocardiographic profiles of patients with MS have not been extensively studied. Although sevoflurane is considered relatively more cardiostable than propofol, the extent of its influence on flow dynamics in MS and its clinical significance remain poorly understood.

We hypothesize that sevoflurane, compared with propofol, may be associated with more favorable hemodynamic parameters after GA induction in patients with MS, owing to its potential to reduce heart rate, which may in turn improve valvular flow dynamics. Therefore, we designed a prospective, randomized, comparative study to evaluate and compare the effects of propofol- and sevoflurane-based induction on echocardiographic parameters in patients with MS undergoing MVR. The primary objective of the study was to assess and compare the change in mean transmitral pressure gradient three minutes after induction of GA. Changes in other echocardiographic and Doppler parameters, as well as hemodynamic variables, were considered secondary objectives of the study.

Methods

This prospective, randomized, outcome-assessor-blinded single center clinical trial was conducted in accordance with the Declaration of Helsinki, with approval from the institutional ethics committee (SCT/IEC/1509/JAN-2023) and following prospective trial registration with the clinical trials registry of India (Registration Number: CTRI/2023/08/056224; Date: 07, August 2023, study period 10-Aug-2023 to 30-Dec-2023). Written informed consent was obtained from

all the study participants and the study adhered to the Consolidated Standards of Reporting Trials (CONSORT) guidelines for reporting parallel groups.⁸ Provisionally accepted patients were screened for eligibility to participate in the study. Patients aged more than 18 years of either sex, who were diagnosed with MS and scheduled to undergo elective MVR were enrolled sequentially into the study. Exclusion criteria were patients not willing to participate in the study, more than mild mitral regurgitation, associated aortic, pulmonary, and tricuspid valve disease, associated symptomatic coronary artery disease, on preoperative inotropic or mechanical circulatory supports, redo surgeries and emergency surgeries. We also planned to exclude patients with suboptimal pre-induction or post-induction Transthoracic Echocardiographic (TTE) images.

After screening and enrolling the patients for the study, all study participants were randomly allocated to one of the study groups based on a computer-generated random number table. The type of intervention the subjects were about to receive on the day of surgery was concealed in an opaque sealed envelope. Those envelopes were handed over to that particular anesthesiologist attending the case. The study participants received either propofol (Group P) or sevoflurane (Group S) for induction of GA based on randomization and group allocation. An independent echocardiographer performed the TTE assessment as per the study protocol. The hemodynamic and echocardiographic data were recorded and analyzed by an outcome assessor.

Patients included in the study were educated about the study during preoperative evaluation. In the operating room, after attaching the ASA standard monitors, invasive arterial cannulation was done under local anesthesia for all the patients. Additionally, Bi-Spectral Index (BIS; Medtronic, Minneapolis, MN, USA) sensors were placed on the forehead to monitor the depth of anesthesia. Induction of general anesthesia was performed using fentanyl ($5 \mu\text{g.kg}^{-1}$) and midazolam (0.05 mg.kg^{-1}) in both groups. In Group S, anesthesia was induced with sevoflurane, starting at 2% and gradually increasing to 6% in 100% oxygen, with patients breathing spontaneously until loss of the eyelash reflex. In Group P, induction was performed with propofol at 1 mg.kg^{-1} , followed by 10 mg boluses until loss of the eyelash reflex. In both groups, anesthetic depth was titrated to maintain a Bispectral Index (BIS) between 40 and 60. This was followed by administration of vecuronium 0.1 mg.kg^{-1} to facilitate tracheal intubation. Titrated boluses of $25 \mu\text{g}$ of injection phenylephrine was administered when the Mean Arterial Pressure (MAP) reduced more than 20% from the baseline. Anesthesia was maintained with sevoflurane, fentanyl infusion, and intermittent boluses of vecuronium.

Before GA induction (T-0) and after 3 minutes of GA induction, but before tracheal intubation (T-1), TTE examination of the heart was performed by an experienced perioperative echocardiographer using an ultrasound machine (S5-1; iE33; Philips, Bothell, USA) and the images were stored for analysis (Fig. 1). An independent outcome assessor blinded to the type of intervention analyzed the following parameters from the stored echocardiographic data: 1) Peak Velocity (PV), 2) Peak Pressure Gradient (PPG), 3) Mean Pressure Gradient (MPG), 4) Pressure Half Time (PHT), 5)

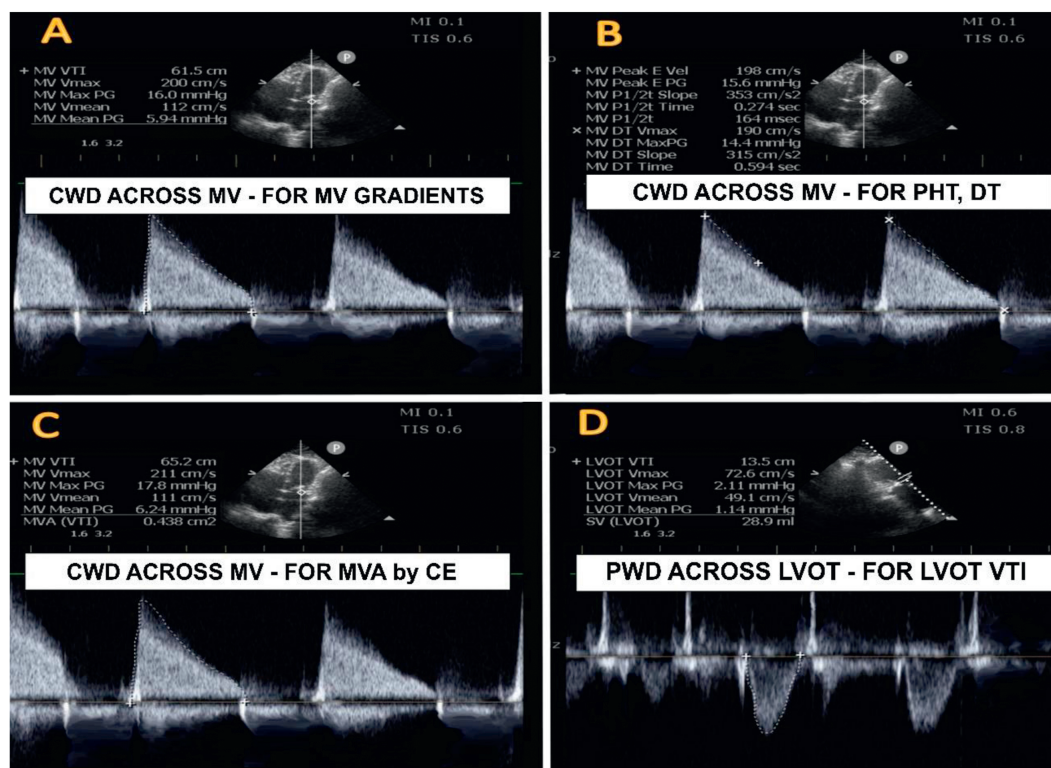


Figure 1 Images showing various echocardiographic parameters to determine the severity of MS. (A) CWD across mitral valve in transthoracic apical four-chamber view-spectral Doppler traced to determine the Vmax-maximum velocity, peak, and mean pressure gradients. (B) CWD across mitral valve in transthoracic apical four-chamber view-spectral Doppler traced along the slope starting from the peak to half the velocity for determination of PHT and till the baseline to derive DT. (C) CWD across mitral valve in transthoracic apical four-chamber view-spectral Doppler traced to obtain MV VTI for the derivation of MVA by continuity equation. (D) LVOT VTI was obtained by tracing the PWD spectral trace across LVOT in an apical five-chamber view. CE, Continuity Equation; CWD, Continuous Wave Doppler; DT, Deceleration Time; LVOT, Left Ventricular Outflow Tract; MV, Mitral Valve; MVA, Mitral Valve Area; PHT, Pressure Half Time; PLAX, Parasternal Long Axis view; PWD, Pulse Wave Doppler; VTI, Velocity Time Integral.

Deceleration Time (DT), 6) Mitral Valve Area by continuity equation (MVA-c) and 7) Mitral Valve Area by PHT (MVA-p). Parameters 1 to 5 were derived by tracing the transmitral continuous wave spectral Doppler trace in the apical 4 chamber view (Fig. 1A–B). MVA-c was determined from Velocity Time Integral (VTI) of the mitral valve (Fig. 1C), VTI of Left Ventricular Outflow Tract (LVOT) (Fig. 1D), and LVOT diameter [MVA-c = (LVOT area × LVOT VTI) / mitral valve VTI].^{9,10} The MVA-p is determined by the formula: $MVA-p = (220/PHT)$.⁶ All the echocardiographic assessments were performed according to the latest American Society of Echocardiography (ASE) guidelines by averaging 3 to 5 cardiac cycles in patients with sinus rhythm and 7 to 10 cardiac cycles in patients with atrial fibrillation.¹¹

The patient's hemodynamic parameters like MAP, Heart Rate (HR), and Cardiac Output (CO) were noted before induction (baseline), at 1-minute, and at 3 minutes after induction of GA. The total dose of phenylephrine required during induction to restore the MAP close to the baseline was recorded. To test the reproducibility of Doppler measurements 10 sets of data were collected randomly. The data were analyzed by a single assessor at two different time points for the intra-observer variability and by an additional assessor for interobserver variability.

Based on a pilot study involving 10 patients that demonstrated an effect size of 0.5 standard deviations (Cohen's $d = 0.5$), a sample size of 36 participants per group was calculated to be sufficient to detect a 25% change in MPG between pre- and post-induction values within each group, with 80% statistical power and a significance level (α) of 0.05. We included 40 patients in each group considering a 10% drop-out rate due to technical reasons like poor quality echocardiographic images. All statistical analyses were performed using the statistical software package SPSS version 23.0 (IBM Software Group, Chicago, IL 60606, USA). Categorical and continuous variables were expressed as number (proportions) and mean $\pm$ SD or median (Interquartile Range; IQR; 25th–75th percentiles) respectively. The Kolmogorov-Smirnov test was used to test the normality of the data. A paired t -test was used to compare quantitative parameters before and after intervention in each group. An independent t -test was used to compare quantitative parameters between the two groups. Pearson Chi-Square test was used for the analysis of categorical variables. The intra-class correlation coefficient was calculated for intra and inter-observer reliability assessment. An intra-class correlation coefficient of 0.7 to 1 is considered to be a good correlation. For all statistical interpretations, a p -value of < 0.05 was considered to be statistically significant.

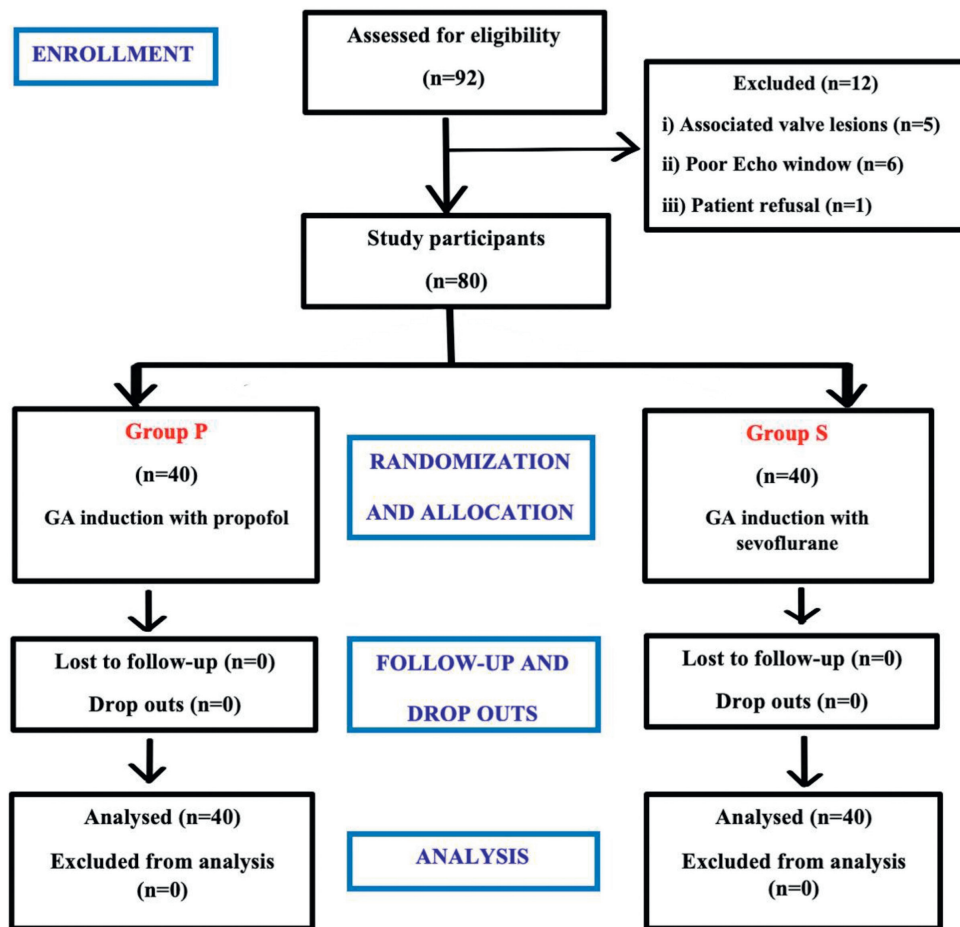


Figure 2 CONSORT diagram depicting various stages at which patients were screened and finally recruited into the study.

Results

A total of 92 patients diagnosed with MS and scheduled for elective MVR were assessed for eligibility. Of these, 12 patients were excluded based on predefined exclusion criteria. The remaining 80 patients were randomly allocated into two groups, with 40 patients in each group. No patients were excluded after randomization because of suboptimal pre- or post-induction TTE images and all randomized participants were included in the final analysis (Fig. 2). Both groups were comparable for demography, functional class, and preoperative echocardiographic data (Table 1). In Group P, after the induction of GA, there was a significant fall in MPG and PV when compared to the baseline pre-induction values (Table 2). In Group S, besides a significant decrease in post-induction MPG, PPG and PV, there was also a significant decrease in PHT and DT (Table 2). This significant decrease in PHT was reflected as an increase in post-induction MVA-p in comparison to pre-induction MVA-p in Group S (Table 2). The MVA-c also showed a significant increase in the post-induction period in comparison to pre-induction values in Group S (Table 2). The pre-induction and post-induction values of PHT, DT, MVA-c, and MVA-p did not show any significant change in Group P (Table 2).

On comparison of post-induction echocardiographic parameters between the two groups, the PHT was significantly lower in Group S than in Group P whereas the MVA-p and MVA-c were

significantly higher in Group S than in Group P (Table 3). The other post-induction echocardiographic parameters (PV, PPG, MPG, and DT) were comparable between the two groups. Baseline hemodynamic parameters (HR, MAP, and CO) were comparable between the two groups. The post-induction HR was significantly lower in Group S in the 1st minute and 3rd minute in comparison to Group P (Table 3). In both groups, the post-induction MAP and CO values decreased from the baseline and were comparable between the two groups (Table 3). The phenylephrine requirement to normalize the MAP was significantly higher in Group P (83%) in comparison to Group S 64% (50 ± 12.5 vs. 37.5 ± 12.5 μ g, median [IQR]- 50 [25] vs. 37.5 [25], $p < 0.0001$). Intra-class correlation coefficient analysis of data (PV, PPG, MPG, PHT, and MVA-p) measured by different observers (inter-observer variability) and by the same observer at two different time points (intra-observer variability) showed statistically significant good correlation (Table 4).

Discussion

In this study, both propofol and sevoflurane were associated with reductions in MPG and PPG following induction of GA. A greater increase in estimated valve area was observed in the sevoflurane group. Hemodynamic changes were more pronounced with propofol, whereas sevoflurane was associated with more favorable hemodynamic effects. Under normal

Table 1 Comparison of demographic data, functional class, and preoperative echocardiographic parameters.

Variables		Group P (n = 40)	Group S (n = 40)	p-value
Age (years)		50 ± 7.0	49.6 ± 7.2	0.8150 ^a
Weight (kg)		62.5 ± 6.8	66.2 ± 10	0.0540 ^a
Gender	Male	8 (20%)	16 (40.0%)	0.0510 ^b
	Female	32 (80%)	24 (60%)	
NYHA class	II	20 (50.0%)	22 (55.0%)	0.654 ^b
	III	20 (50.0%)	18 (45.0%)	
Heart rhythm	Sinus	24 (60.0%)	24 (60.0%)	1.0000 ^b
	Non- Sinus	16 (40.0%)	16 (40.0%)	
Preoperative TTE	MPG (mmHg)	12.09 ± 4.12	13.18 ± 5.10	0.2963 ^a
	MVA (cm ²)	0.93 ± 0.20	0.96 ± 0.24	0.5454 ^a
Pre-induction TTE	MPG (mmHg)	11.12 ± 3.02	11.30 ± 4.84	0.8424 ^a
	MVA (cm ²)	0.88 ± 0.21	0.91 ± 0.11	0.4259 ^a

MAP, Mean Arterial Pressure; MPG, Mean Pressure Gradient; MVA, Mitral Valve Area; NYHA, New York Heart Association; TTE, Transthoracic Echocardiography.

^a Using *t*-test, data expressed as mean and Standard Deviation (SD).

^b Using Chi-Square test, data expressed as frequency.

conditions, the flow velocity across the mitral valve is less than 1.3 m.s⁻¹.⁹ As mitral valve area decreases with disease progression, transvalvular flow velocity increases. Induction of general anesthesia with either propofol or sevoflurane reduces sympathetic and parasympathetic tone and increases venous capacitance, leading to a reduction in venous return.^{12,13} The resulting decrease in blood flow across the mitral valve accounts for the observed reductions in peak velocity, peak pressure gradient, and mean pressure gradient. However, the updated guidelines on valvular heart diseases did not consider either velocity or pressure gradients for the classification of disease severity.⁴ This is because the trans-mitral Doppler is heavily influenced by the flow across the valve, which in turn is determined by the diastolic filling time and compliance of the left atrium and left ventricle.^{14,15} Despite these limitations, the MPG serves as a very important and convenient parameter to decide upon the severity of stenosis and has been in use for a longer time.¹⁶ Transmitral MPG of less than 5 mmHg indicates mild or no stenosis whereas MPG of more than 10 mmHg indicates severe stenosis.^{17,18} When this criterion was considered, it was surprising to find that approximately 23% of the total study patients in our study, who were documented to have severe MS, had their MPG less than 5 mmHg after induction of GA (Group P: 12 [30%] and Group S: 6 [15%]). This finding was predominantly seen in patients induced with propofol. This is a clinically alarming phenomenon under anesthesia, where misclassification of MS severity could easily happen.

The PHT is the next important echocardiographic parameter that indicates the severity of MS. According to the American College of Cardiology (ACC)/American Heart Association (AHA) 2020 guidelines, a mitral valve PHT of ≥ 150 msec is considered severe MS.⁴ All patients enrolled in our study had a baseline PHT of more than 200 msec. After the induction of GA, the PHT dropped significantly in Group S compared to Group P. The slight reduction in PHT from baseline in Group P could be due to an increase in left atrial reservoir capacity caused by propofol.¹⁹ The limitation of the PHT measurement is that it is greatly influenced by the HR and net compliance of the cardiac chambers. Moreover, the

presence of more than mild mitral regurgitation could reduce the PHT and underestimate the true MS severity.^{14,15} Hence, we excluded the patients with associated moderate and severe mitral regurgitation.

Valve area is constant at varying degrees of hemodynamic alterations and hence it serves as a most reliable marker of MS severity. The estimated mitral valve area of less than 1.5 cm² is graded as severe MS.⁴ By echocardiography, the mitral valve area can be estimated by planimetry, continuity equation, PHT, and the Proximal Iso velocity Surface Area (PISA) method.²⁰ The valve area estimation by planimetry is considered gold standard but, both planimetry and PISA could not be performed uniformly in all patients due to poor alignment and sub-optimal image in supine lying patients which could have resulted in erroneous measurements. Moreover, the dense calcification of the mitral leaflets underestimates the mitral valve area measurement by the planimetry method.^{21,22} Hence, we estimated the mitral valve area by PHT and continuity equation for all the study participants in both groups. The significant drop in post-induction PHT in Group S increased the post-induction Mitral Valve Area estimated by PHT (MVA-p). Also, the post-induction MVA-c increased significantly in Group S. In comparison with propofol, the post-induction MVA-c and MVA-p were significantly higher after sevoflurane induction. These observations agree with the findings of the study by Kuperstein et al.²³ The exact reason for the increase in valve area under anesthesia is not completely understood. It could be the complex interplay between the HR, loading conditions of the heart, myocardial contractility, and interindividual responses to anesthetic agents.^{12,13} However, the observed increase in MVA-p and MVA-c after induction does not reflect a true anatomical change in valve area but rather the effect of altered loading conditions and HR.

As far as hemodynamic parameters are concerned, both sevoflurane and propofol are known to reduce mean arterial pressure.¹³ In the present study, MAP and cardiac output decreased after induction of GA in both groups. However, these variables were measured repeatedly over time and our analysis relied on separate pairwise *t* tests rather than a

Table 2 Comparison of pre-induction and post-induction echocardiographic and Doppler parameters.

Doppler parameters	Group P (n = 40)				Group S (n = 40)			
	Pre-induction	Post-induction	Δ (95% CI)	p-value	Pre-induction	Post-induction	Δ (95% CI)	p-value
PV (m.s^{-1})	2.2 $\pm$ 0.3	1.8 $\pm$ 0.4	0.4 (0.29 to 0.52)	<0.0001	2.3 $\pm$ 0.4	1.9 $\pm$ 0.3	0.4 (0.29 to 0.52)	<0.0001 ^a
PPG (mmHg)	20.4 $\pm$ 6.1	13.9 $\pm$ 5.9	6.5 (4.58 to 8.42)	<0.0001	20.4 $\pm$ 6.0	14.6 $\pm$ 4.0	5.8 (4.10 to 7.50)	<0.0001 ^a
MPG (mmHg)	11.1 $\pm$ 3.0	7.1 $\pm$ 2.8	4 (3.07 to 4.93)	<0.0001	11.3 $\pm$ 4.8	8.0 $\pm$ 3.1	3.3 (1.95 to 4.65)	<0.0001 ^a
PHT (msec)	261.9 $\pm$ 66.9	250.3 $\pm$ 55.8	11.6 (-8.2 to 31.4)	0.4023	240.6 $\pm$ 57.5	224.6 $\pm$ 58.5	16 (2.5 to 34.5)	<0.0001 ^a
DT (msec)	846.0 $\pm$ 192.8	831.4 $\pm$ 250.8	14.6 (-58.0 to 87.2)	0.7711	811.8 $\pm$ 201.7	767.6 $\pm$ 179.7	44.2 (17 to 105.4)	0.0030 ^a
MVA-c (cm^2)	0.90 $\pm$ 0.21	0.91 $\pm$ 0.26	0.01 (-0.066 to 0.086)	0.8504	0.99 $\pm$ 0.28	1.05 $\pm$ 0.31	0.06 (0.035 to 0.155)	0.0080 ^a
MVA-p (cm^2)	0.88 $\pm$ 0.21	0.92 $\pm$ 0.29	0.04 (-0.043 to 0.123)	0.4819	0.91 $\pm$ 0.11	1.08 $\pm$ 0.25	0.17 (0.101 to 0.239)	0.0002 ^a
Post induction MPG < 5 mmHg	12 (30%)				6 (15%)			0.1104 ^b

CI, Confidence Interval; DT, Deceleration Time; MPG, Mean Pressure Gradient; MVA-c, Mitral Valve Area by continuity equation; MVA-p, Mitral Valve Area by PHT; PHT, Pressure Half Time; PPG, Peak Pressure Gradient; PV, Peak Velocity. Δ = Pre-induction value-post induction value.

^a Using paired *t* test, Values are expressed as mean $\pm$ Standard Deviation.

^b Using Chi-Square test, data expressed as frequency.

Table 3 Comparison of post-induction echocardiographic, Doppler and hemodynamic parameters.

Parameters	Group P (n = 40)	Group S (n = 40)	Estimated mean difference (95% CI)	p-value
PV (m.s ⁻¹)	1.8 ± 0.4	1.9 ± 0.3	-0.1 (-0.257 to 0.0574)	0.2097 ^a
PPG (mmHg)	13.9 ± 5.9	14.6 ± 4.0	-0.7 (-2.944 to 1.544)	0.5364 ^a
MPG (mmHg)	7.1 ± 2.8	8.0 ± 3.1	-0.9 (-2.215 to 0.415)	0.1769 ^a
PHT (msec)	250.3 ± 55.8	224.6 ± 58.5	25.7 (0.252 to 51.148)	0.0478 ^a
DT (msec)	831.4 ± 250.8	767.6 ± 179.7	63.8 (-33.32 to 160.92)	0.1948 ^a
MVA-c (cm ²)	0.91 ± 0.26	1.05 ± 0.31	-0.14 (-0.267 to 0.0126)	0.0316 ^a
MVA-p (cm ²)	0.92 ± 0.29	1.08 ± 0.25	-0.16 (-0.281 to -0.0395)	0.0099 ^a
HR (Beats.min ⁻¹) Baseline	83.8 ± 15.1	82.4 ± 19.5	1.4 (-6.363 to 9.163)	0.7210 ^{a,b}
HR (Beats.min ⁻¹) 1 st minute	86.2 ± 14.8 (Δ = 2.4, 95% CI -2.37 to 7.17)	78.1 ± 18.1 (Δ = 4.3, 95% CI -1.72 to 10.32)	8.1 (0.74 to 15.46)	0.0310 ^{a,b}
HR (Beats.min ⁻¹) 3 rd minute	89.1 ± 16.1 (Δ = 5.3, 95% CI 0.31 to 10.29)	76.7 ± 16.9 (Δ = 5.7, 95% CI -0.18 to 11.58)	12.4 (5.053 to 19.747)	0.0010 ^{a,b}
MAP (mmHg) Baseline	105.0 ± 7.3	105.7 ± 10.5	-0.7 (-4.726 to 3.326)	0.7131 ^{a,b}
MAP (mmHg) 1 st minute	91.6 ± 6.5 (Δ = 13.4, 95% CI 11.18 to 15.62)	89.1 ± 8.1 (Δ = 16.6, 95% CI 13.55 to 19.65)	2.5 (-0.769 to 5.769)	0.1263 ^{a,b}
MAP (mmHg) 3 rd minute	85.8 ± 6.4 (Δ = 19.2, 95% CI 17.0 to 21.4)	85.4 ± 8.8 (Δ = 20.3, 95% CI 17.17 to 23.43)	0.4 (-3.025 to 3.825)	0.8173 ^{a,b}
CO (L.min ⁻¹) Baseline	4.1 ± 0.8	4.0 ± 0.6	0.1 (-0.215 to 0.415)	0.5289 ^{a,b}
CO (L.min ⁻¹) post-induction	3.1 ± 0.8 (Δ = 1, 95% CI 0.74 to 1.26)	2.9 ± 0.5 (Δ = 1.1, 95% CI 0.92 to 1.280)	0.2 (-0.097 to 0.497)	0.1839 ^{a,b}

CI, Confidence Interval; CO, Cardiac Output; DT, Deceleration Time; HR, Heart Rate; MAP, Mean Arterial Pressure; MPG, Mean Pressure Gradient; MVA-c, Mitral Valve Area by continuity equation; MVA-p, Mitral Valve Area by PHT; PHT, Pressure Half Time; PPG, Peak Pressure Gradient; PV, Peak Velocity.

Estimated mean difference = mean ± standard deviation (Group P – Group S).

Δ = Baseline value – value at that time point (within the same group).

^a Using *t*-test, data expressed as mean and Standard Deviation (SD).

^b p-values (unadjusted) were obtained from separate pairwise comparisons and do not account for within-subject correlation (no repeated-measures model) and were not adjusted for multiple comparisons; therefore, results – especially marginal p-values – should be interpreted cautiously.

repeated-measures approach like repeated measures ANOVA or a mixed effects model that would better account for within subject correlation and formally assess time-group interaction. Hence, p-values, especially those near the significance threshold, should be interpreted cautiously, as multiple comparisons across time points and outcomes increase the risk of Type I error. Therefore, findings are best considered time-point specific. The underlying mechanism

for the hemodynamic change differs slightly between the two anesthetic agents. Reduced venous return and fall in systemic vascular resistance reduce the MAP with sevoflurane induction, whereas the added myocardial depression reduces the MAP with propofol induction.¹³ Theoretically, reduced MAP should augment the CO. Since MS is a fixed output condition, the cardiac output may fail to increase with the fall in MAP. Sevoflurane, a regularly used induction

Table 4 Intra-class correlation coefficient analysis for inter and intra-observer reliability of data.

Intraclass correlation coefficient analysis					
	Inter-observer coefficient	95% CI		p-value	Intra-observer coefficient
		Lower Bound	Upper Bound		
PV	0.994	0.976	0.999	< 0.0001	0.995
PPG	0.992	0.968	0.998	< 0.0001	0.990
MPG	0.991	0.965	0.998	< 0.0001	0.997
PHT	0.992	0.970	0.998	< 0.0001	0.998
MVA-p	0.986	0.942	0.996	< 0.0001	0.984

CI, Confidence Interval; MPG, Mean Pressure Gradient; MVA-c, Mitral Valve Area by continuity equation; MVA-p, Mitral Valve Area by PHT; PHT, Pressure Half Time; PPG, Peak Pressure Gradient; PV, Peak Velocity. Intra-class correlation coefficient of 0.7 to 1 is a good correlation.

anesthetic agent, is not well studied for its cardio-dynamic effects. However, the available literature supports the reduced response of sevoflurane to vasopressors during induction.²⁴ In our study, sevoflurane induction reduced the HR significantly in comparison to propofol. Both sevoflurane and propofol depress the baroreceptor reflex and the net change in HR depends on the balance between the sympathetic and para-sympathetic tone.^{25,26} As a reflex phenomenon, heart rate should increase with a fall in the MAP. Sevoflurane blunts the baroreceptors more often than propofol to produce a significant reduction in HR.^{23,27} Also, previous studies have shown that sevoflurane anesthesia resulted in a more stable fall in MAP with a low normal HR.²⁸ In MS patients, lower HR is more favorable during the induction of GA.

Heart rhythm and respirophasic changes are the other important influencing factors in assessing the Doppler parameters.^{29,30} The echocardiographic Doppler parameters vary between the patients in sinus rhythm and those patients with atrial fibrillation due to the abolition of the 'atrial kick' required for adequate left ventricular filling.³¹ Additionally, as part of the induction procedure, the patients were ventilated with positive pressure after administration of the neuromuscular blocking drug, which again could have potentially altered the Doppler parameters. We overcame these two confounding factors by averaging the Doppler measurement over several cardiac cycles as per the ASE guideline recommendations.¹¹ It could have nullified the changes in heart rhythm as well as the respiratory cycle which could have otherwise occurred.

We acknowledge that our study has a few limitations. On the outset, we restricted our study to a single ethnic group with two commonly used anesthetic induction agents. Probably, a multinational trial involving a diverse ethnic population group with additional anesthetic agents will be superior in generalizing study results. Next, ACC/AHA recommends the estimation of valve area by planimetry and Pulmonary Artery Systolic Pressure (PASP) for the classification of MS severity.⁴ We could not obtain satisfactory images uniformly to perform the planimetry mitral valve area measurement and not all patients had tricuspid regurgitation for measuring PASP. Also, we did not study the effect of these two agents on postoperative outcomes like duration of mechanical ventilation, length of intensive care unit stay, hospital stay, and mortality. The possible influence of concomitant anesthetic medications and fluid/vasopressors is undeniable. Finally, we could have performed subgroup analysis in those patients with atrial fibrillation. Despite these limitations, our study highlights the short-term hemodynamic effects of sevoflurane as an induction agent in patients with mitral stenosis, suggesting potential areas for further clinical studies.

Conclusion

In conclusion, both inhalational and intravenous anesthetic induction agents were associated with reductions in trans-mitral pressure gradients following induction of general anesthesia in patients with mitral stenosis. Sevoflurane was associated with more favorable short-term hemodynamic and echocardiographic changes, including an increase in

estimated valve area and a reduction in heart rate at the studied time points. As these findings are based on surrogate outcomes from a single center study, further multicenter studies incorporating clinical outcomes are required to establish their clinical significance.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

AI disclosure

The authors did not use generative artificial intelligence in the preparation of this manuscript.

Institutional ethics committee approval

(SCT/IEC/1509/JAN-2023)

Clinical trial registration number

Clinical trials registry of India (Registration Number: CTRI/2023/08/056224; Date: 07, August 2023).

Funding

None.

Conflicts of interest

The authors declare no conflicts of interest.

Associate Editor

Luiz Marcelo Malbouisson

References

1. Luo T, Han J, Meng X. Features of rheumatic mitral valves and a grading system to identify suitable repair cases in China. *J Thorac Dis.* 2017;9:3138–47.
2. Chandrashekar Y, Westaby S, Narula J. Mitral stenosis. *Lancet.* 2009;374:1271–83.
3. Baumgartner H, Falk V, Bax JJ, et al. 2017 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur Heart J.* 2017;38:2739–91.
4. Otto CM, Nishimura RA, Bonow RO, et al. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: executive summary. *Circulation.* 2021;143:e35–71.
5. Maeder MT, Weber L, Buser M, et al. Pulmonary hypertension in aortic and mitral valve disease. *Front Cardiovasc Med.* 2018;5:40.
6. Imran TF, Awtry EH. Severe mitral stenosis. *N Engl J Med.* 2018;379:e6.
7. Klein AJ, Carroll JD. Left ventricular dysfunction and mitral stenosis. *Heart Fail Clin.* 2006;2:443–52.
8. Schulz KF, Altman DG, Moher D, Group CONSORT. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332.
9. Baumgartner H, Hung J, Bermejo J, et al. Echocardiographic assessment of valve stenosis: EAE/ASE recommendations for clinical practice. *J Am Soc Echocardiogr.* 2009;22:1–23.
10. Pandian NG, Kim JK, Arias-Godinez JA, et al. Recommendations for the Use of Echocardiography in the Evaluation of Rheumatic Heart Disease: A Report from the American Society of Echocardiography. *J Am Soc Echocardiogr.* 2023;36:3–28.
11. Mitchell C, Rahko PS, Blauwet LA, et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults. *J Am Soc Echocardiogr.* 2019;32:1–64.
12. Kanaya N, Hirata N, Kurosawa S, Nakayama M, Namiki A. Differential effects of propofol and sevoflurane on heart rate variability. *Anesthesiology.* 2003;98:34–40.
13. de Wit F, van Vliet AL, de Wilde RB, et al. The effect of propofol on haemodynamics. *Br J Anaesth.* 2016;116:784–9.
14. Diwan A, McCulloch M, Lawrie GM, Reardon MJ, Nagueh SF. Doppler estimation of left ventricular filling pressures in patients with mitral valve disease. *Circulation.* 2005;111:3281–9.
15. Oh JK, Park S, Pislaru SV, Nkomo VT. Native valvular heart disease. In: Oh JK, Kane GC, eds. *The Echo Manual*, 4th ed., Philadelphia: Wolters Kluwer; 2018:103–32.
16. El Sabbagh A, Reddy YNV, Barros-Gomes S, et al. Low-gradient severe mitral stenosis. *J Am Heart Assoc.* 2019;8:e010736.
17. Harb SC, Griffin BP. Mitral valve disease: a comprehensive review. *Curr Cardiol Rep.* 2017;19:73.
18. Essandoh M. Intraoperative echocardiographic assessment of mitral valve area after degenerative mitral valve repair. *J Cardiothorac Vasc Anesth.* 2016;30:1364–8.
19. Kehl F, Kress TT, Mraovic B, et al. Propofol alters left atrial function evaluated with pressure-volume relations in vivo. *Anesth Analg.* 2002;94:1421–6.
20. Kostibas MP, Brady MB, Koch CG. Mitral valve stenosis. In: Perino CA Jr, Reeves ST, eds. *A Practical Approach to Transesophageal Echocardiography*, 4th ed., Philadelphia: Wolters Kluwer; 2020:220–37.
21. Omran AS, Arifi AA, Mohamed AA. Echocardiography of the mitral valve. *J Saudi Heart Assoc.* 2010;22:165–70.
22. Cherry AD, Maxwell CD, Nicoara A. Intraoperative evaluation of mitral stenosis by transesophageal echocardiography. *Anesth Analg.* 2016;123:14–20.
23. Kuperstein R, Raibman-Spector S, Canetti M, et al. Influence of anesthesia on hemodynamic assessment of mitral stenosis severity. *Cardiol J.* 2022;29:245–51.
24. Kanaya N, Satoh H, Seki S, Nakayama M, Namiki A. Propofol anesthesia enhances the pressor response to intravenous ephedrine. *Anesth Analg.* 2002;94:1207–11.
25. Nagasaki G, Tanaka M, Nishikawa T. Recovery profile of baroreflex control of heart rate after isoflurane or sevoflurane anesthesia in humans. *Anesth Analg.* 2001;93:1127–31.
26. Husedzinović I, Tonković D, Barisin S, Bradić N, Gasparović S. Hemodynamic differences in sevoflurane versus propofol anesthesia. *Coll Antropol.* 2003;27:205–12.
27. Dhande K, Kshirsagar J, Dhande A, Patil N, Jr VP. Hemodynamic stability, patient acceptance and cost of propofol and sevoflurane for induction of anesthesia. *Cureus.* 2020;12:e7687.

28. Fredman B, Nathanson MH, Smith I, Wang J, Klein K, White PF. Sevoflurane for outpatient anesthesia. *Anesth Analg.* 1995; 81:823–8.
29. Soni N, Williams P. Positive pressure ventilation: what is the real cost? *Br J Anaesth.* 2008;101:446–57.
30. Michard F. Changes in arterial pressure during mechanical ventilation. *Anesthesiology.* 2005;103:419–28.
31. Arı H, Arı S, Karakuş A, et al. Impact of cardiac rhythm on mitral valve area and gradient in mitral stenosis. *Anatol J Cardiol.* 2017;18:90–8.

ORIGINAL INVESTIGATION

Effects of shoulder rotation and table tilt maneuvers on the lumbar and thoracic paramedian acoustic window in older adults: a quasi-experimental study



Marcos V. Nunes de Souza ^{a,b,*}, Ana M.M. Caetano ^a, Nádia M.C. Duarte ^a,
Camila T. Nunes de Souza ^a, Artur J.L. Bezerra ^a, Emmanuelle T.A.G.B.B. Silva ^a

^a Universidade Federal de Pernambuco, Medical Science Center, Recife, PE, Brazil

^b Hospital da Restauração, Department of Anesthesiology, Recife, PE, Brazil

Received 29 July 2025; accepted 20 May 2026

Available online 1 June 2026

KEYWORDS

Aged;
Epidural anesthesia;
Posterior longitudinal
ligament;
Spinal anesthesia;
Ultrasonography

Abstract

Background: Shoulder rotation and table tilt have been reported to increase the acoustic window for spinal ultrasonography. However, evidence in older adults remains scarce. This study evaluated their anatomical effects on lumbar (L3–L4) and thoracic (T10–T11) paramedian acoustic windows, quantified by posterior longitudinal ligament (PLL) length, in older adults (≥ 65 years).

Methods: This quasi-experimental study enrolled 45 older adults scheduled for elective surgery at two tertiary hospitals in Brazil. The primary outcome was PLL ultrasonographic length (paramedian sagittal oblique plane) assessed at the L3–L4 and T10–T11 interspaces in three positions: F (standard spinal flexion), S (F plus 45° leftward shoulder rotation), and T (F plus 10° dorsal table tilt). Data were analyzed using a Linear Mixed-Effects Model with Bonferroni-adjusted post-hoc contrasts.

Results: The model revealed significant effects of position ($p = 0.003$) and spinal level ($p < 0.001$) on PLL length, with no Level $\times$ Position interaction ($p = 0.693$). PLL length was significantly greater at L3–L4 than at T10–T11. Table tilt produced a significant increase relative to the flexion position (difference = 2.2 mm; 95% CI 0.9–3.5; $p = 0.002$), whereas shoulder rotation did not (difference = 1.2 mm; 95% CI 0.0–2.5; $p = 0.158$), with homogeneous effects across both levels. Fixed effects explained 27% of total variance ($R^2_m = 0.27$; $R^2_c = 0.55$).

Conclusion: Table tilt produced a significant anatomical enlargement of the neuraxial acoustic window at both spinal levels studied, whereas shoulder rotation did not reach statistical significance. These findings establish an anatomical foundation for future trials evaluating their impact on neuraxial procedural outcomes in older adults.

* Corresponding author.

E-mail: marcos.vnsouza94@gmail.com (M.V. Nunes de Souza).

Trial registration: <https://ensaiosclinicos.gov.br/rg/RBR-3p6t9pm>.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Neuraxial anesthesia offers several key advantages, including maintenance of consciousness, spontaneous ventilation, and airway patency, while providing effective sympathetic and motor blockade, benefits that are particularly relevant for older adults.^{1,2} However, comorbidities and age-related physiological changes, such as ligament calcification, osteophyte formation, and facet joint hypertrophy, narrow the spinal puncture window, increasing the technical difficulty of neuraxial anesthesia.^{3,4} Consequently, advanced age directly correlates with a greater number of puncture attempts,⁵ which not only cause patient discomfort but also increase the risk of severe complications, including post-dural puncture headache, neuraxial hematoma, and neurological injury.⁶ In this scenario, Ultrasound (US) has become a safe and widely available tool that can help predict and manage difficulties that may arise during neuraxial anesthesia.^{7–10} Weed et al. identified a correlation between the ease of spinal anesthesia and the size and quality of the Posterior Longitudinal Ligament (PLL), as visualized by US.¹¹

Dorsal table tilt and shoulder rotation maneuvers have been reported to increase thoracic PLL length in adults, thereby expanding the spinal acoustic window.^{12–14} However, these maneuvers have shown inconsistent results in certain populations, such as pregnant women,¹⁵ and remain understudied in the elderly.

Additionally, conventional lumbar flexion provides limited expansion of the acoustic window in older patients, especially at L3–L4, where it may not differ from the neutral position.¹⁶ Given these challenges, maneuvers designed to optimize the acoustic window and facilitate neuraxial blockade in older adults are of utmost clinical relevance. However, before their impact on clinical outcomes can be assessed, their anatomical and biomechanical effects must first be characterized.

The aim of the present study was to evaluate the influence of shoulder rotation and dorsal table tilt on the lumbar (L3–L4) and thoracic (T10–T11) acoustic windows in older adults, by measuring PLL length using the Paramedian Sagittal Oblique (PSO) ultrasound plane.

Methods

Study design and participants

The study was approved by the Institutional Review Board (IRB) (CAAE: 66608522.1.0000.8807; 66608522.1.3001.5198) and was registered at the Brazilian Registry of Clinical Trials – ReBEC platform (<https://ensaiosclinicos.gov.br/rg/RBR-3p6t9pm>). This quasi-experimental study was conducted at two tertiary hospitals in Recife, Brazil, and evaluated 50 volunteers (first enrollment: June 19, 2024; last enrollment: December 20, 2024) scheduled for elective surgery.

Inclusion criteria

The inclusion criteria were ≥ 65 years of age and willingness to provide written informed consent.

Exclusion criteria

The exclusion criteria were: inability to maintain the designed positions; allergy to ultrasound gel; history of shoulder or spine surgery, trauma, or anomaly (congenital or acquired); and poor ultrasound image quality, defined as a Weed score ≤ 8 .¹¹ For calculation of the Weed score, each participant was evaluated at four lower lumbar interspaces (L5–S1, L4–L5, L3–L4, L2–L3), bilaterally. Each site was scored 0, 1, or 2 (0 = PLL not visible; 1 = Hazy; 2 = Clear visualization of PLL and its limits), resulting in a total score ranging from 0 to 16 (higher scores, in particular scores above 8, indicate better PLL visualization).¹¹

Data collection and sequence positioning

For each study participant, age, sex, weight, height, Body Mass Index (BMI), American Society of Anesthesiologists (ASA) physical status classification, and comorbidities were recorded. Before surgery, each volunteer was seated on the operating table with hips and knees flexed at 90°. Prior to image acquisition, the sonographic quality of the four lower lumbar interspaces was evaluated using the Weed score. This assessment served strictly as an objective exclusion criterion; participants presenting with low-quality image (Weed score ≤ 8) were excluded to ensure the reliability of all subsequent measurements. Spinal ultrasonography in the Paramedian Sagittal Oblique (PSO) plane was then performed at the L3–L4 and T10–T11 levels in three consecutive positions (F→S→T), yielding six images per participant.¹⁷ Each data collection session lasted approximately 15 minutes per participant. This fixed sequence was adopted to ensure accurate assignment of each image to the corresponding position, spinal level, and participant, given that the stored images contained no identifying information. An assistant instructed and assisted patients in maintaining each position, as described below (Fig. 1).

F (Flexion): The shoulders are held in a ‘slouched’ or flexed posture, with cervical and lumbar flexion and exaggeration of the thoracic kyphosis, while the arms rest on the legs.

S (Shoulder rotation): The same as in position F, but with 45° leftward shoulder rotation along the axis of the vertebral column, which was confirmed by a Protractor (Keuwlsoft, London, UK).

T (Table tilt): the patient is placed in position F, but with a 10° dorsal table tilt, confirmed by a Protractor (Keuwlsoft, London, UK).

Shoulder rotation and dorsal table tilt angles were chosen based on previous studies.^{12–15} We chose the L3–L4 and T10–



Figure 1 Sitting positions and maneuvers. (a) Normal sitting position. (b) Position F – standard spinal flexed position. Individuals were instructed to relax their shoulders, flex their cervical and lumbar spines, and exaggerate thoracic kyphosis, while resting their arms on their legs. (c) Position S - shoulder rotation maneuver. Same as in position F, and 45° leftward shoulder rotation on the same axis of the vertebral column. (d) Position T – table tilt maneuver. Same as in position F, and 10° dorsal table tilt.

T11 levels because they are commonly used at our institution for administration of anesthesia or analgesia during lower-limb, abdominal, and thoracic surgeries.

Image acquisition

Starting from the sacral plateau, a scan of the paramedian sagittal oblique plane was performed using a 2–5 MHz curvilinear transducer (GE HealthCare, Chicago, Illinois, USA). Scanning was continued cranially, counting each lamina until the L3–L4 and T10–T11 levels were reached. The position was confirmed by manual palpation of spinal processes, beginning with the sacral plateau, and by ultrasonographic identification of the T12 vertebra, which is the first vertebra to articulate with the rib when scanning from bottom to top. The ‘sawtooth’ pattern was used to identify the laminae, while the anterior complex, consisting of the PLL and vertebral body, was visualized as a deeper, brighter structure (Fig. 2).¹⁷ The best PLL image, optimized by careful probe tilting, was then saved to a USB drive.

Bias attenuation and image masking

An experienced anesthesiologist (M.S.) with over 100 neuraxial ultrasound scans and three years of practical experience performed all US scans and measurements for the present study. Although the anesthesiologist could not be blinded to the patient’s position, the measurement process was conducted under blinded conditions. An assistant who did not participate in recording the US scans anonymized the stored images (USB drive) by assigning them a random number from 1 to 270 (using a simple random draw) and sent them to the

investigator (M.S.) in this new random sequence. The investigator then used these anonymized US images to determine PLL length along the largest axis (Fig. 2). The procedure was executed using the Radiant DICOM Viewer application (Medixant, Poznan, Wielkopolska, Poland). No patient identification, positioning, or level was present in the file name or image (Fig. 2). The examiner compiled all PLL measurements into a spreadsheet, and after completion, the assistant assigned each PLL value to the corresponding participant, spinal level, and position.

Sample size and statistical analysis

Previous studies enrolled 30–34 subjects, considering a 1 mm increase in PLL as anatomically meaningful.^{12–15} Based on data derived from these studies, a statistician calculated that a minimum sample size of 38 (effect size = 0.7, α = 0.05, power = 90%, and 10% dropout) would be ideal for the present study. A target of 50 subjects was therefore established.

A Linear Mixed-Effects Model (LMM) was fitted to account for the hierarchical structure of repeated measurements nested within subjects across positions and spinal levels. Position (F, S, and T; reference: F), spinal level (L3–L4 and T10–T11; reference: L3–L4), and their two-way interaction (Level $\times$ Position) were entered as categorical fixed effects, with a random intercept specified for each patient to account for between-subject variability in baseline PLL length. Parameters were estimated using Restricted Maximum Likelihood (REML) with Satterthwaite degrees of freedom approximation. The overall significance of fixed effects was first assessed using Type III F-tests. As the Level $\times$ Position interaction was not statistically significant, pairwise position

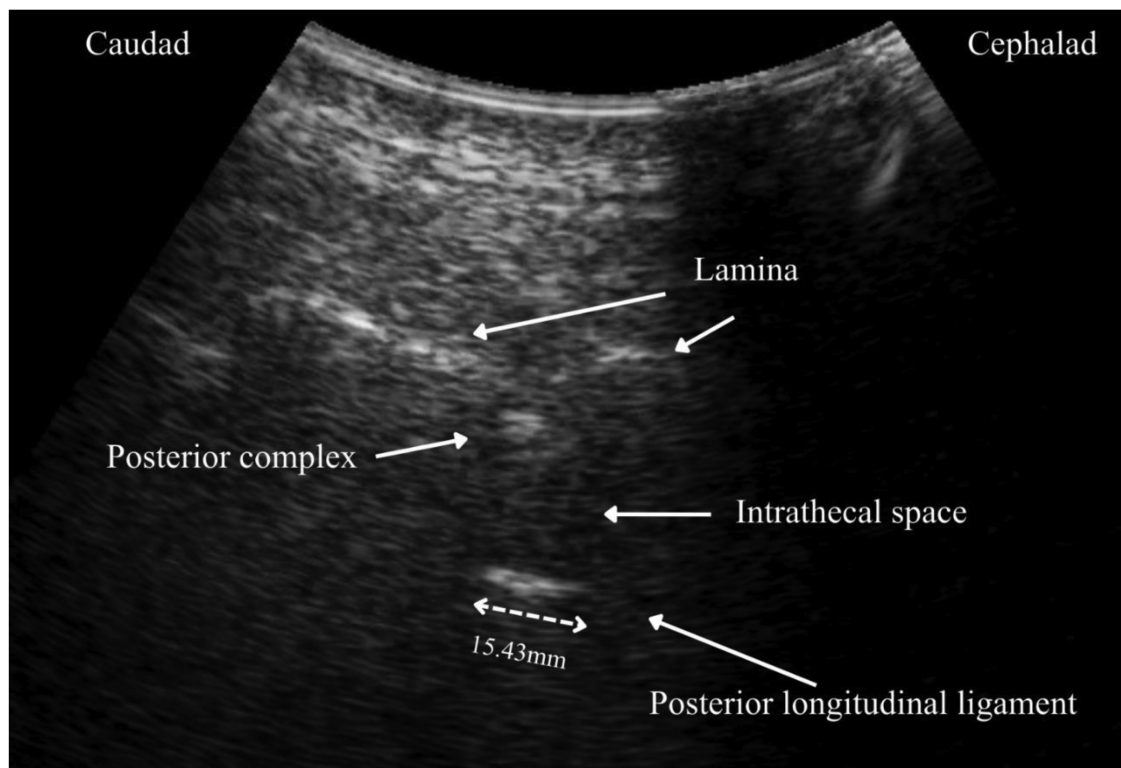


Figure 2 Structures of neuraxial ultrasound on paramedian sagittal oblique plane on L3–L4 intervertebral space and PLL length along the largest axis. No patient identification, positioning, or level was present in the file name or image.

contrasts (S vs. F, T vs. F, and T vs. S) were subsequently computed as marginal comparisons collapsed across spinal levels and adjusted by the Bonferroni method. Effect sizes were reported as marginal R^2 (R^2_m) and conditional R^2 (R^2_c), reflecting variance explained by fixed effects alone and by the full model, respectively.

Model assumptions were satisfied. Residual normality was confirmed by the Shapiro-Wilk test ($p = 0.403$) and quantile-quantile plots, with no evidence of heteroscedasticity on inspection of residuals versus fitted values. The random effects followed an approximately normal distribution.

The primary outcome was PLL length at the L3–L4 and T10–T11 interspaces across positions F, S, and T. No secondary outcomes were pre-specified. All analyses were performed in R software (version 4.4.2, R Core Team, Vienna, Austria).

Results

Figure 3 summarizes the enrollment process. Initially, 50 volunteers met the study eligibility criteria. However, four had to be excluded from the present study because they could not maintain the required positions: of these, three reported pain related to a recent unrelated surgery, and one required supplemental oxygen and experienced shortness of breath. In addition, imaging data for one participant were lost due to a power outage. Baseline characteristics of the five excluded participants were not formally compared with those of the included sample owing to the small number of exclusions; however, no systematic differences in age, sex, or ASA classification were apparent from available records. Thus, a total of 45 participants were included in the final analysis, and their demographic characteristics are presented in Table 1.

Table 1 Demographic characteristics of participants.

Characteristic	Data (n = 45)
Sex, n (%)	
Male	22 (49)
Female	23 (51)
Age, years (mean $\pm$ SD)	72.0 $\pm$ 6.0
Height, m (mean $\pm$ SD)	1.63 $\pm$ 0.09
Body mass index, kg.m ⁻² (mean $\pm$ SD)	24.7 $\pm$ 5.4
Weight, kg (mean $\pm$ SD)	65.4 $\pm$ 13.3
ASA classification ^a , n (%)	
I	2 (4)
II	14 (31)
III	28 (62)
IV	1 (2)
Comorbidities ^b , n (%)	
Diabetes mellitus	32 (71)
Systemic arterial hypertension	32 (71)
Peripheral arterial disease	19 (42)
Smoking	19 (42)
Overweight	14 (31)
Alcoholism	9 (20)
Obesity	6 (13)
Coronary artery disease	3 (7)
Cancer	2 (4)
Others	9 (20)

ASA, American Society of Anesthesiologists. Data are expressed as mean $\pm$ SD, and number (%).

^a Sum of percentages was 99% due to rounding.

^b Sum of percentages exceeded 100% because it is a prevalence measure, and each participant might have more than one comorbidity.

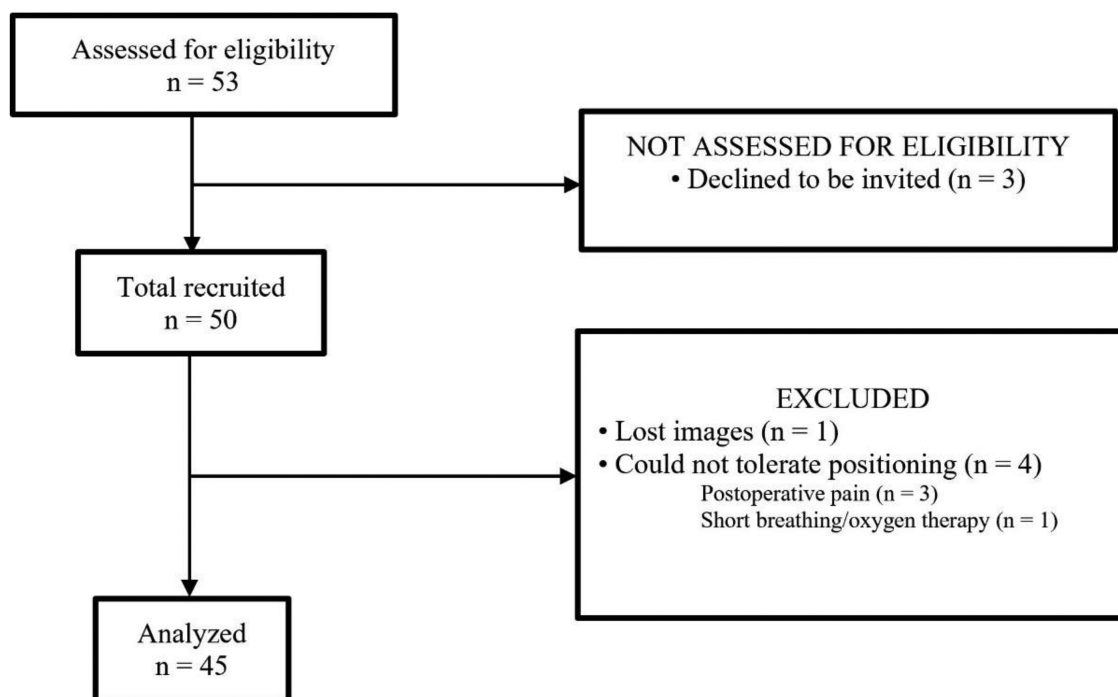


Figure 3 Enrollment of patients in the study.

Table 2 Linear mixed-effects model results: estimated marginal means of PLL length, omnibus F-tests, and post-hoc pairwise comparisons (n = 45).

Estimated Marginal Means			
Level	Position		
	F [95% CI]	S [95% CI]	T [95% CI]
L3-L4 (mm)	15.2 ± 0.8 [13.6, 16.9]	16.4 ± 0.8 [14.8, 18.0]	17.9 ± 0.8 [16.2, 19.5]
T10-T11 (mm)	13.0 ± 0.8 [11.3, 14.6]	14.3 ± 0.8 [12.7, 15.9]	14.8 ± 0.8 [13.1, 16.4]

Fixed Effects Omnibus Tests				
	F-statistic	df	df (res)	p-value
Position	5.976	2	220	0.003
Level	22.536	1	220	< 0.001
Position × Level	0.367	2	220	0.693

Post-hoc pairwise comparison ^a		
Level	Difference [95% CI]	p-value Bonferroni
L3-L4 – T10-T11 (mm)	2.5 ± 0.5 [1.45, 3.51]	< 0.001
Position		
S – F (mm)	1.2 ± 0.6 [0.0, 2.5]	0.158
T – F (mm)	2.2 ± 0.6 [0.9, 3.5]	0.002
T – S (mm)	1.0 ± 0.6 [-0.3, 2.2]	0.405

PLL, Posterior Longitudinal Ligament; 95% CI, 95% Confidence Interval; df, Degrees of freedom; df(res), Residual degrees of freedom; F, Standard spinal flexion; S, 45° leftward shoulder rotation; T, 10° dorsal table tilt.

Coefficients of determination $R^2_m = 0.27$; $R^2_c = 0.55$.

Data are expressed as mean ± SE; 95% Confidence Intervals are model-derived from the Linear Mixed-Effects Model.

^a Post-hoc comparisons for Position are reported as marginal means collapsed across spinal levels, as the Level × Position interaction was not statistically significant (p = 0.693); p-values adjusted by Bonferroni correction.

Four measurements (two at L3–L4 and two at T10–T11) yielded a PLL length of 0 mm in position F, as the PLL could not be visualized despite repeated scanning; notably, in these same participants, the PLL became visible after application of the maneuvers.

Table 2 and Figure 4 summarize the key outcomes. The LMM revealed significant overall effects of position and spinal level on PLL length, with no significant Level × Position interaction (p = 0.693), indicating that the effects of the maneuvers were statistically homogeneous across spinal levels. In the post-hoc analysis, PLL length was significantly greater at L3–L4 than at T10–T11. Table tilt was the only maneuver to produce a significant increase in PLL length relative to the flexion position, an effect considered applicable to both spinal levels; shoulder rotation did not reach statistical significance. The difference between table tilt and shoulder rotation was also not statistically significant.

Fixed effects explained 27% of the total variance in PLL length ($R^2_m = 0.27$), rising to 55% when between-subject variability was accounted for ($R^2_c = 0.55$). The random intercept variance was 11.2 mm² (SD = 3.34 mm), reflecting substantial individual variability in baseline PLL length across patients.

Discussion

In the present study, we hypothesized that shoulder rotation and table tilt maneuvers would enlarge the acoustic window at the lumbar (L3–L4) and thoracic (T10–T11) levels in older adults. The LMM revealed a significant overall effect of position on PLL length, driven by the table tilt maneuver, which reached statistical significance in the post-hoc analysis. The Level × Position interaction was not significant, indicating that the anatomical response to the maneuvers was statistically homogeneous across spinal levels; the effect of table tilt was therefore interpreted as applicable to both L3–L4 and T10–T11. Shoulder rotation did not reach statistical significance after Bonferroni correction. These findings establish an anatomical foundation for future clinical trials evaluating the impact of these maneuvers on clinical outcomes in older adults.

Sebbag et al. evaluated shoulder rotation and table tilt in full-term pregnant women at the L3–L4 interspace, but found no significant increase in PLL dimensions, likely owing to the limited lumbar flexion imposed by the gravid uterus.¹⁵ In contrast, the present study demonstrated a statistically significant anatomical effect of table tilt across both spinal

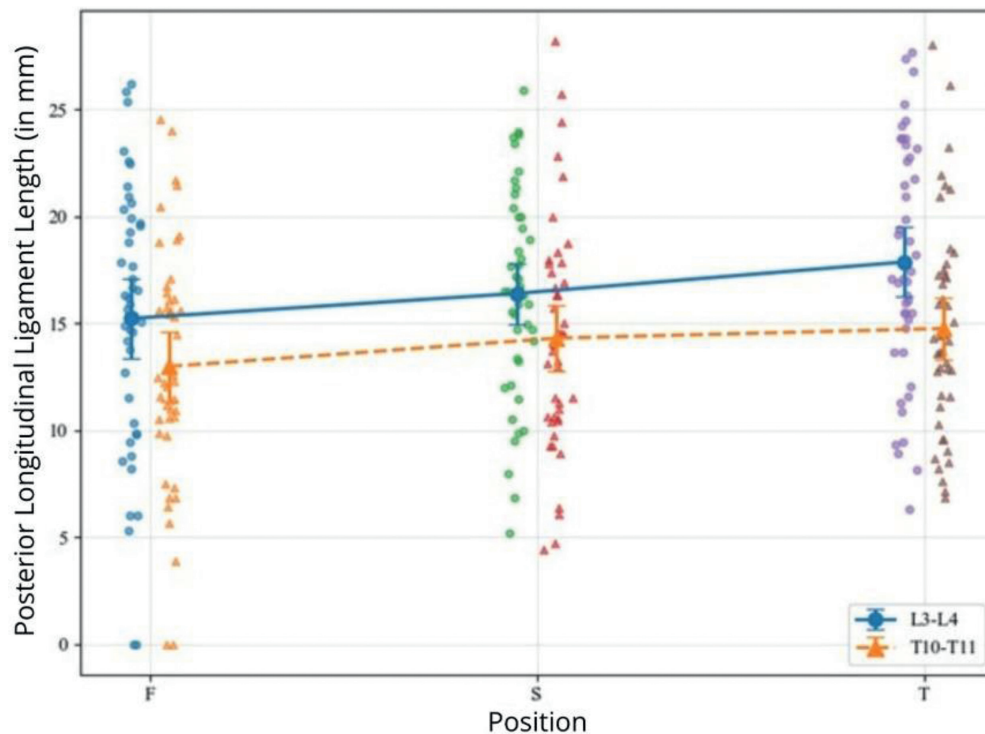


Figure 4 Estimated Marginal Means and 95% confidence interval of PLL length over position F (standard spinal Flexed position), S (Shoulder rotation maneuver), and T (Table tilt maneuver) for each level.

levels, highlighting the need for further research in challenging patient populations.

Unlike previous investigations, which reported significant increases in PLL length following shoulder rotation at thoracic levels,¹²⁻¹⁴ the present study found no statistically significant effect of this maneuver at either spinal level. In contrast, table tilt resulted in an increase in PLL length consistent with or exceeding values reported in prior investigations,¹²⁻¹⁴ suggesting that this maneuver may be particularly appropriate for elderly patients.

The significantly greater PLL length at L3–L4 compared with T10–T11 ($p < 0.001$) is consistent with the known anatomical differences between lumbar and thoracic vertebral morphology, with wider intervertebral spaces at lower spinal levels.

Biomechanically, shoulder rotation induces vertebral torsion by rotating the upper trunk with respect to the hips, which are held fixed, resulting in a more pronounced enlargement of the interlaminar spaces at higher thoracic levels.^{18,19} Therefore, the diminished and non-significant response at both levels may also be attributed to reduced mobility, dehydration, and age-related narrowing of the intervertebral spaces.^{3,4} Table tilt, by contrast, enhances both lumbar and thoracic flexion simultaneously, promoting a more uniform expansion of the intervertebral spaces across levels.^{18,19}

A difference of at least 1 mm in PLL length has been proposed as an anatomically meaningful threshold, approximating the outer diameter of a standard epidural needle.¹²⁻¹⁴ In the present study, all observed mean differences between positions exceeded this threshold at both spinal levels;

however, only table tilt reached statistical significance. Future studies with larger sample sizes may help clarify whether shoulder rotation produces a consistently detectable anatomical effect in this population.

The effects of the maneuvers observed in the present study can be further characterized by the effect size metrics derived from the LMM. Fixed effects (spinal level and position) explained 27% of the total variance in PLL length ($R^2_m = 0.27$), a moderate proportion indicating that position and vertebral level are meaningful determinants of PLL length. When individual between-subject variability was incorporated through the random intercept, the model explained 55% of total variance ($R^2_c = 0.55$). The gap between R^2_m and R^2_c reflects the substantial contribution of individual anatomical differences, quantified as a random intercept variance of 11.2 mm² (SD = 3.34 mm), underscoring that baseline PLL dimensions vary considerably across older adults, independently of position or spinal level.

Interestingly, while the PLL could not be identified in four initial images obtained in position F, it became visible after applying the proposed maneuvers. This finding highlights the potential of these techniques; however, whether this translates into any procedural advantage remains to be established in future clinical studies.

The present study has some limitations. All ultrasound examinations were performed by a single experienced investigator (over 100 US scans), precluding interobserver reliability analysis. This limitation reduces the external validity of the present findings, given the operator-dependent nature of ultrasound imaging. The same investigator who performed the ultrasound scans subsequently measured the PLL lengths; as

blinding of the examiner was not feasible, potential measurement bias was minimized by anonymizing the images and presenting them in a random sequence, ensuring the examiner was unaware of patient identity, maneuver, or spinal level. The fixed sequential order of the positions (F→S→T) without randomization may have introduced an order bias in both directions: while progressive muscle relaxation over time could have favored a greater expansion of the acoustic window during the final maneuvers, cumulative fatigue arising from sustained non-physiological positions could have conversely hindered the volunteers' ability to maintain the optimal posture.

Although fatigue or discomfort could interfere with maintaining the required positions, 90% of the study participants tolerated the maneuvers well. Among the 50 volunteers who met the inclusion criteria, five were excluded: four due to pain or discomfort that prevented proper positioning, and one because of image loss during a power outage. Based on the present study, the maneuvers appear to be well-tolerated by older adults, although specific clinical conditions may limit their general applicability. Nevertheless, the final sample size still exceeded the minimum required for statistical validity.

Finally, and most importantly, the present study assessed PLL length as a surrogate anatomical marker and was not designed to evaluate clinical outcomes. Whether enlargement of the acoustic window translates into improved procedural outcomes requires prospective clinical investigation.

The present study is notable for evaluating an elderly patient population, which has often been excluded from previous investigations.¹²⁻¹⁵ Older individuals have biomechanical constraints that may limit expansion of the intervertebral spaces and reduce the effectiveness of the maneuvers. Nevertheless, table tilt produced a statistically significant and anatomically meaningful increase in PLL length applicable to both spinal levels studied.

The study was conducted at two tertiary public hospitals within a developing country, reflecting a population with limited access to health-promoting resources and treatments. This is supported by the high prevalence of patients classified as ASA III-IV, indicating significant comorbidity burden, which may have adversely affected spinal mobility and attenuated the observed effects. Accordingly, the present findings may represent a conservative estimate of the anatomical response to table tilt, and this response may be more pronounced in healthier elderly individuals; however, whether comorbidity burden directly affects PLL dimensions remains to be investigated.

Conclusion

Table tilt enlarged the acoustic window, as reflected by PLL enlargement, at both the lumbar (L3–L4) and thoracic (T10–T11) levels in older adults. Shoulder rotation did not significantly increase PLL length at either level. As the present study assessed PLL length as a surrogate anatomical marker, these findings provide anatomical evidence that table tilt may enlarge the neuraxial acoustic window in older adults. Future clinical trials should assess the ability of these maneuvers to reduce the technical challenges associated with the administration of neuraxial anesthesia and related complication rates.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Chat GPT and Claude in order to improve fluency, language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Funding

This research did not receive any specific funding from funding agencies in the public, private, or non-profit sectors.

Acknowledgements

We are thankful to Kamila Cunha for providing assistance during data collection.

Conflicts of interest

The authors declare no conflicts of interest.

Authors' contributions

Marcos V. Nunes de Souza: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ana M.M. Caetano:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. **Nádia M.C. Duarte:** Conceptualization, Methodology, Writing – review & editing. **Camila T. Nunes de Souza:** Data curation, Formal analysis, Investigation, Resources, Visualization, Writing – original draft, Writing – review & editing. **Artur J.L. Bezerra:** Data curation, Investigation, Writing – review & editing. **Emmanuelle T.A.G.B. Silva:** Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Associate Editor

Eric Benedet Lineburger

References

1. Barbosa FT, Jucá MJ, Castro AA. Neuroaxis Block Compared to General Anesthesia for Revascularization of the Lower Limbs in the Elderly. A Systematic Review with Metanalysis of Randomized Clinical Studies. *Braz J Anesthesiol.* 2009;59:234–43.
2. Nordquist D, Halaszynski TM. Perioperative multimodal anesthesia using regional techniques in the aging surgical patient. *Pain Res Treat.* 2014;2014:1–10.
3. Fecine BRA. O processo de envelhecimento: as principais alterações que acontecem com o idoso com o passar dos anos. *InterSciencePlace.* 2012;1:106–32.
4. Albuquerque M, Mattos S, Nunes R. Educação Continuada em Anestesiologia. 1st ed. Rio de Janeiro: Sociedade Brasileira de Anestesiologia; 2018.
5. Atallah MM, Demian AD, Shorrab AA, et al. Development of a difficulty score for spinal anaesthesia. *Br J Anaesth.* 2004;92:354–60.
6. Duits FH, Martinez-Lage P, Paquet C, et al. Performance and complications of lumbar puncture in memory clinics: Results of the multicenter lumbar puncture feasibility study. *Alzheimers Dement.* 2016;12:154–63.
7. Poots C, Chin KJ. Strategies for successful lumbar neuraxial anesthesia and analgesia in patients with challenging anatomy. *BJA Educ.* 2024;24:46–56.
8. Perlas A, Chaparro LE, Chin KJ. Lumbar neuraxial ultrasound for spinal and epidural anesthesia: A systematic review and meta-analysis. *Reg Anesth Pain Med.* 2016;41:251–60.
9. Chin KJ, Ramlogan R, Arzola C, et al. The Utility of Ultrasound Imaging in Predicting Ease of Performance of Spinal Anesthesia in an Orthopedic Patient Population. *Reg Anesth Pain Med.* 2013;38:34–8.
10. Chin KJ, Perlas A, Chan V, et al. Ultrasound Imaging Facilitates Spinal Anesthesia in Adults with Difficult Surface Anatomic Landmarks. *Anesthesiology.* 2011;115:94–101.
11. Weed JT, Taenzer AH, Finkel KJ, et al. Evaluation of pre-procedure ultrasound examination as a screening tool for difficult spinal anesthesia. *Anaesthesia.* 2011;66:925–30.
12. Ramsay N, Walker J, Tang R, et al. Flexion-rotation manoeuvre increases dimension of the acoustic target window for paramedian thoracic epidural access. *Br J Anaesth.* 2014;112:556–62.
13. Byon HJ, Song JH, Song KC, et al. The effects of shoulder rotation on the acoustic window for thoracic paramedian epidural approach in the lateral decubitus position. *Reg Anesth Pain Med.* 2016;41:572–5.
14. Byon HJ, Hong SJ, Kim DY, et al. Comparison of the acoustic windows for the thoracic paramedian epidural approach after shoulder rotation: The lateral decubitus versus the sitting position. *Anaesth Crit Care Pain Med.* 2019;38:53–6.
15. Sebbag I, Tang R, Gunka V, et al. Effect of table tilt and spine flexion–rotation on the acoustic window of the lumbar spine in pregnant women. *Braz J Anesthesiol.* 2018;68:280–4.
16. Byon HJ, Han JU, Jung JK, et al. Comparison of the Effects of Lumbar Spine Flexion on the Acoustic Windows Between Young and Elderly Patients. *Pain Physician.* 2022;25:E759–65.
17. Chin KJ, Karmakar MK, Peng P, et al. Ultrasonography of the Adult Thoracic and Lumbar Spine for Central Neuraxial Blockade. *Anesthesiology.* 2011;114:1459–85.
18. Gregersen GG, Lucas DB. An in vivo study of the axial rotation of the human thoracolumbar spine. *J Bone Joint Surg Am.* 1967;49:247–62.
19. White AA, Panjabi MM. *Clinical Biomechanics of the Spine.* 2nd ed. Boston: Lippincott Williams & Wilkins; 1990. p. 102–23.

ORIGINAL INVESTIGATION

Continuous erector spinae plane block for robotic-assisted thoracic surgery in lung cancer: a dual-center retrospective propensity-score weighted cohort study



Matheus Arraes ^{a,*}, Sara Amaral ^b, Saullo Queiroz Silveira ^c, Leopoldo Muniz da Silva ^{a,d}, Rafael Sousa Fava Nersessian ^a, Francisco Jose Lucena Bezerra ^a, Manoel de Souza Neto^a, Caroline Machado Nunes ^a, Helidea de Oliveira Lima ^d, Fernando Nardy Bellicieri ^c, Glenio B. Mizubuti ^{e,1}, Rafael Lombardi ^{f,1}

^a São Luiz Hospital – ITAIM/Rede, Department of Anesthesiology, Brazil

^b Duke University Medical Center, Department of Anesthesiology, Durham, NC, USA

^c Vila Nova Star Hospital/Rede D'Or, Department of Anesthesiology, São Paulo, SP, Brazil

^d D'Or Institute for Research and Education (IDOR), São Paulo, SP, Brazil

^e Queen's University, Department of Anesthesiology and Perioperative Medicine, Kingston, ON, Canada

^f University of Nebraska Medical Center, Department of Anesthesiology, Omaha, NE, USA

Received 7 October 2025; accepted 20 May 2026

Available online 6 June 2026

KEYWORDS

Analgesia;
Erector spinae plane
block;
Lung neoplasms;
Propensity score;
Robotic surgical
procedures;
Thoracic surgical
procedures

Abstract

Background: Postoperative analgesia remains challenging in thoracic surgery, particularly with the rise of Robotic-Assisted Thoracoscopic Surgery (RATS). The Erector Spinae Plane Block (ESPB) is a promising opioid-sparing technique, but its role in RATS is unclear.

Methods: In this retrospective dual-center cohort study, we included 78 patients with lung cancer undergoing RATS, who received general anesthesia with or without ESPB. ESPB involved a single shot of 30 mL of 0.5% ropivacaine injection at T7, followed by a 24-hour infusion of 0.2% ropivacaine (7 mL.h⁻¹). The primary outcome was opioid rescue within 36 hours. Secondary outcomes included hospital length of stay and chest tube duration. Confounders were balanced with Inverse Probability Treatment Weighting (IPTW).

Results: In unadjusted analyses, there was no significant difference in rescue opioid consumption between groups (ESPB 37.8% vs. GA 27.3%; $p = 0.46$). After propensity score weighting to improve covariate balance, ESPB use was associated with lower postoperative opioid requirements (OR = 0.57; 95% CI 0.32–0.79; $p = 0.01$). This finding may reflect improved adjustment for measured confounders, although residual confounding cannot be excluded. In contrast, chest

* Corresponding author.

E-mail: matheusaarraes@gmail.com (M. Arraes).

¹ These authors have contributed equally to this work and share senior authorship.

tube presence within 24 hours was associated with higher opioid requirements (OR = 2.32; 95% CI 1.22–4.20; $p = 0.04$). ESPB patients also required fewer scheduled opioids during the first 24–36h postoperatively ($p = 0.03$).

Conclusion: In this observational cohort, ESPB use was associated with lower postoperative opioid requirements in RATS. These findings suggest a potential role for ESPB within ERAS pathways, although randomized studies are needed to confirm efficacy and define its role in minimally invasive thoracic surgery.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Postoperative pain management remains a critical challenge in thoracic surgery, in which inadequate analgesia can lead to severe complications, including atelectasis, pneumonia, respiratory failure, and chronic pain.¹ Over the past decade, minimally invasive techniques such as Video-Assisted Thoracic Surgery (VATS) and Robotic-Assisted Thoracoscopic Surgery (RATS) have largely replaced traditional thoracotomy, reducing surgical trauma and enhanced recovery.^{2–5} However, optimal pain control strategies must evolve alongside these advancements.

While thoracic epidural analgesia has long been considered the gold standard for thoracic procedures, its use has gradually declined due to well-recognized adverse effects, including hypotension and urinary retention, particularly when continuous local anesthetic-opioid combinations are employed.⁶ In this context, opioid-sparing multimodal analgesia strategies and ultrasound-guided fascial plane blocks have gained increasing attention. Among these, the Erector Spinae Plane Block (ESPB) has emerged as a promising alternative.^{4–7} First described in 2016,⁴ ESPB is characterized by technical simplicity, a favorable safety profile, and a lower risk of neuraxial-related complications compared with traditional epidural techniques.⁶

Despite its growing adoption, key uncertainties remain regarding ESPB's mechanism of action. Studies report inconsistent spread to the paravertebral, epidural, and intercostal spaces, with significant variability in block distribution among patients.^{7–14} This raises questions about its reliability and optimal application, particularly in RATS, where evidence remains scarce.

Evidence specifically addressing the use of ESPB in RATS for lung cancer remains very limited, particularly regarding continuous ESPB techniques using catheter-based local anesthetic infusion. Available data are largely restricted to small case series or extrapolated from non-robotic or single-injection studies, leaving an important gap in evidence for this increasingly common surgical setting.^{15–17}

In this context, the primary aim of this study was to evaluate the effectiveness of continuous ESPB as part of a multimodal, opioid-sparing analgesic strategy in patients undergoing RATS lung resection using a dual-center retrospective cohort design with propensity score weighting. We hypothesized that adding continuous ESPB to general anesthesia would reduce the proportion of patients requiring postoperative rescue opioids within 36 hours, compared with general anesthesia alone.

Methods

This investigation was approved by the Institutional Research Ethics Board (IREB) (protocol 4.518.838, CAAE 42052821.2.0000.0087), which waived informed consent. STROBE guidelines were followed, and a checklist is attached in the supplementary material. This study complied with Resolution 466/2012 of the Brazilian National Health Council. The authors received no financial support for the research, authorship, and/or publication of this article. No funding bodies had any role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

In this retrospective cohort, we reviewed electronic medical records at two tertiary hospitals between March 2019 and March 2022. All consecutive patients who underwent RATS for lung cancer during the study period were screened for eligibility. A total of 86 patients were initially identified. Six patients were excluded due to incomplete medical records, primarily related to missing postoperative pain assessments and/or incomplete documentation of rescue opioid administration, which precluded reliable classification of the primary outcome. One patient who received thoracic epidural analgesia and one patient with unintended ESPB catheter dislodgement were also excluded. No intraoperative conversions to open thoracotomy occurred during the study period. Patients with incomplete medical records were excluded prior to analysis, ensuring complete data availability for all covariates and outcome variables. No partial missingness remained, and therefore complete case analysis was performed without imputation.

During the study period, robotic thoracic surgeries were managed by a dedicated anesthesia team that routinely covered all eligible cases. The assignment of anesthesiologists to operating rooms was determined by standard departmental scheduling, independent of patient characteristics and without prior knowledge of which specific surgical cases would be performed. As a result, ESPB was not selectively offered based on individual patient factors or perceived pain risk.

For patients included in this study, thoracic procedures were consistently performed by the same group of anesthesiologists, who were instructed to follow an institutional protocol consisting of either general anesthesia alone or general anesthesia combined with ESPB. Although the study design was neither blinded nor systematically randomized, anesthesiologists became aware of individual case details only after assignment. This approach aimed to reduce selective application of ESPB, although residual confounding by indication cannot be entirely excluded.

We included patients aged ≥ 18 years, undergoing RATS with GA, with and without ESPB. We excluded patients with chronic opioid consumption, chronic pain, substance abuse, emergency surgeries, palliative care, Patient-Controlled Analgesia (PCA) with the installation of an epidural pump, single-shot epidural anesthesia, catheter dislodgement in patients undergoing ESPB, patients who were not extubated within 12 hours after surgery, combined surgical procedures, reoperation, comatose patients, those with impaired ability to assess pain or request analgesia, and incomplete medical records.

The primary outcome was the incidence of postoperative opioid rescue within 36 hours after awakening from the GA. Rescue opioids were administered only if patients experienced moderate to severe pain, either at rest or with movement. Pain was assessed using the Numerical Rating Scale (NRS), where 0 indicates no pain and 10 indicates extreme pain. Scores ≥ 4 were considered moderate or severe pain. The secondary outcomes included hospital length of stay and chest tube duration (in days). Although the NRS was used clinically to guide rescue analgesia according to a standardized institutional protocol, pain scores were not consistently documented in a structured manner in some of the medical records, and in some cases only analgesic administration was recorded. Therefore, longitudinal analysis of NRS values was not feasible, and the occurrence of rescue opioid administration was used as a surrogate marker of clinically significant pain.

The studied variables included age, gender, Body Mass Index (BMI), comorbidities, anesthetic technique, type of surgery, duration of surgery, postoperative use of opioids and pregabalin, chest tube duration, and Intensive Care Unit (ICU) length of stay.

The general anesthetic management and block technique were based on our hospital's standard of care. Anesthetic management followed institutional standards and was individualized when clinically indicated, according to the judgment of the attending anesthesiologist.

Anesthesia was induced with propofol 2 mg.kg^{-1} , and selective tracheal intubation was facilitated with rocuronium $0.6\text{--}1.2 \text{ mg.kg}^{-1}$. Fentanyl or methadone was used as the opioid for induction, with dexmedetomidine and magnesium sulfate administered as analgesic adjuncts. Anesthesia was maintained with continuous infusions of propofol and remifentanyl; propofol was titrated to maintain Bispectral Index (BIS) values between 40 and 60, while remifentanyl was administered to provide intraoperative analgesia according to clinical judgment. As part of a standardized multimodal analgesia protocol, parecoxib 40 mg , pantoprazole 40 mg , and dexamethasone 0.1 mg.kg^{-1} (maximum 10 mg) were administered before surgical incision. Twenty minutes before surgical completion, intravenous metemazole 2000 mg was given.

Intraoperative monitoring consisted of Non-Invasive Blood Pressure (NIBP), ECG, Pulse Oximetry (SpO_2), capnography, temperature, and BIS. Controlled ventilation was performed in a closed circuit with a flow up to 2 L.min^{-1} , tidal volume up to 7 mL.kg^{-1} , with a Positive end Expiratory Pressure (PEEP) of $6\text{--}8 \text{ cm H}_2\text{O}$, and respiratory rate to maintain an end-tidal CO_2 between $35\text{--}40 \text{ mmHg}$. Neuromuscular blockade was maintained at a post-tetanic count < 5 (i.e., deep block). Reversal of neuromuscular blockade with sugammadex was

performed in all patients. A convective air circulation heating system maintained normothermia. The same surgical team conducted all the included surgeries.

The ESPB and catheter placement were performed at the end of surgery, before emergence from general anesthesia and prior to tracheal extubation, under sterile conditions and ultrasound guidance. The block was intentionally performed at this time point to avoid interference with intraoperative management and patient positioning, as preoperative placement of a continuous ESPB catheter could interfere with the surgical field during robotic thoracic surgery and increase the risk of unintended catheter displacement. Given the use of a continuous catheter, the primary aim of the block was postoperative analgesia rather than preemptive intraoperative analgesia.

The ESPB was performed in a standardized manner using a high-frequency linear ultrasound transducer. A single-shot unilateral ESPB was administered by injecting 30 mL of 0.5% ropivacaine into the fascial plane between the deep surface of the erector spinae muscle and the transverse process of the seventh thoracic vertebra. A 20-gauge epidural catheter was then inserted through the Tuohy needle and advanced 5 cm into the fascial plane. Continuous infusion of ropivacaine 0.2% was administered via the ESPB catheter at a basal rate of 7 mL.h^{-1} , with the option of up to two additional bolus doses of 5 mL.h^{-1} , for the first 24 hours. After this period, the catheter was removed, and analgesia continued with standard or rescue analgesics at the discretion of the surgical team.

The surgical team determined postoperative analgesic management. The regimen included oral oxycodone 10 mg every 12 hours, IV tenoxicam 40 mg every 24 hours, and IV metamizole 50 mg.kg^{-1} every 6 hours.

Pain was assessed at 15-minute intervals during the first postoperative hour and then every 6 hours after discharge from the Post-Anesthesia Care Unit (PACU). Patients were instructed to promptly inform the nursing staff if they experienced moderate to severe pain. Intravenous morphine was administered as rescue analgesia for both groups when the pain score was ≥ 4 , with dosing set at 1 mg every 10 minutes for moderate pain (scores $4\text{--}7$) and 2 mg for severe pain (> 7), aiming to maintain a score below 4. Rescue opioid administration followed a standardized institutional protocol applied uniformly in both the Post-Anesthesia Care Unit (PACU) and the surgical ward. Nursing staff administered intravenous morphine based exclusively on patient-reported pain intensity using the NRS. Due to the presence of an indwelling ESPB catheter, blinding of nursing staff to group allocation was not feasible. This lack of blinding may have introduced performance or detection bias; however, the use of objective pain thresholds ($\text{NRS} \geq 4$) and a standardized rescue opioid protocol aimed to minimize subjective influence on analgesic administration.

Statistical analysis

The sample size was based on the available data, i.e., all patients who underwent RATS between March 2019 and March 2022. No statistical power calculation was performed before the study.

Data distribution normality was evaluated through a normal Quantile–Quantile (Q–Q) plot. For categorical variables,

Table 1 Summary of propensity weighted analysis.

	Propensity weighted analysis		
	Estimate (OR)	95% CI	p-value
Groups			0.01
GA	1.00		
ESPB	0.57	0.32 – 0.79	
Dose of Fentanyl (mcg)	0.96	0.91 – 1.08	0.06
Dose of Methadone (mg)	1.12	0.87 – 1.15	0.09
Chest tube in the First 24h	2.32	1.22 – 4.2	0.03

Estimates correspond to Odds Ratio (OR); CI, Confidence Interval; Weighted analysis evaluating the association between ESPB use and postoperative rescue opioid requirements.

† Postoperative rescue opioid requirement was defined as the occurrence of moderate to severe pain with the need for opioid analgesics (morphine) in the first 36 hours. General Anesthesia without Erector Spinae Plane Block (GA); General Anesthesia with Erector Spinae Plane Block (ESPB). The propensity score was estimated using a logistic regression model based on pre-treatment variables only (demographics, comorbidities, surgical characteristics, and intraoperative analgesic exposure). The weighted outcome model additionally included selected postoperative variables (e.g., chest tube presence within the first 24h) to account for clinically relevant factors associated with postoperative pain.

the Pearson Chi-Squared test was applied, with partitioned Chi-square analyses performed when $p < 0.05$. Comparisons between two independent groups were conducted using the Mann-Whitney *U*-test when the dependent variable was ordinal or continuous but did not meet normality assumptions. Propensity Scores (PS) were used to calculate weights: $1/PS$ for patients in the treatment group (ESPB) and $1/(1-PS)$ for those in the control group (GA), following the method of Inverse Probability of Treatment Weighting (IPTW).

To mitigate confounding inherent to the retrospective design, a propensity score-weighted analysis using Inverse Probability of Treatment Weighting (IPTW) was performed to estimate Odds Ratios (OR) with corresponding 95% Confidence Intervals (95% CI). The propensity score was derived from covariates selected *a priori* based on biological plausibility and prior evidence of association with postoperative pain, opioid consumption, and treatment allocation. These included demographic variables (age, sex, and body mass index), baseline clinical status (ASA physical status, diabetes mellitus, arterial hypertension, and chronic obstructive pulmonary disease), surgical characteristics (type and duration of surgery), and perioperative analgesic exposure (intraoperative doses of fentanyl, methadone, dexmedetomidine, and magnesium sulfate). Variables with a direct impact on postoperative pain and analgesic demand, including chest tube presence/duration, length of ICU stay, and scheduled opioid prescriptions during the first 36 postoperative hours, were not incorporated. Covariate balance after weighting was assessed using absolute standardized differences, with values < 0.1 indicating adequate balance.

The propensity score model included only pre-treatment variables selected *a priori* based on clinical relevance and prior literature, namely demographic characteristics, comorbidities, surgical factors, and intraoperative analgesic exposure. Variables that occurred after treatment allocation or were potentially influenced by the intervention (e.g., presence of a chest tube within the first 24 hours, ICU stay, and postoperative analgesic prescriptions) were intentionally excluded from the propensity score model to avoid adjusting for mediators.

The final weighted outcome model (Table 1) was constructed to estimate the adjusted association between ESPB use and postoperative opioid rescue, and additionally included clinically relevant postoperative variables (e.g., chest tube presence within the first 24 hours) to account for factors directly associated with postoperative pain and opioid requirements. In particular, the presence of chest tube within the first 24 hours was included as it is a well-recognized contributor to postoperative pain and opioid consumption, although it was not included in the propensity score model as it may lie on the causal pathway between treatment and outcome.

Absolute Standardized Differences (SD) were computed to evaluate residual imbalances in measured covariates after cohort weighting. A threshold of $SD < 0.1$ was interpreted as adequate covariate balance between groups. All *p*-values were two-sided, and $p < 0.05$ was considered statistically significant. Overlap assumptions were evaluated by visually inspecting the distribution of estimated propensity scores across treatment groups, confirming the presence of a common region of support. The distribution of inverse probability treatment weights was examined to identify potential instability or excessively large weights. Because no extreme weights were observed, weight truncation was not required. All statistical analyses were performed using R Statistical Software (v4.0.1, The R Foundation, Austria).

Results

A total of 86 patients underwent RATS lung resection during the study period. After excluding 8 cases (6 for incomplete records, 1 epidural anesthesia case, and 1 ESPB catheter dislodgement), 78 patients were analyzed (45 in the ESPB group and 33 in the GA alone group). The cohort had a mean age of 62 ± 12 years and was 51.3% female. Demographic and clinical characteristics are detailed in Table 2. Surgical procedures included lobectomies (57.7%) and segmentectomies (43.3%) (Fig. 1).

Table 2 Patient characteristics, surgical procedures and outcomes.

Age (years)	62.62 ± 12.69
BMI (kg.m ⁻²)	28.01 ± 5.14
ASA-PS Classification % (n)	
I	3.8 (3)
II	78.2 (61)
III	17.9 (14)
Groups % (n)	
General Anesthesia without Erector Spinae Plane Block (GA)	42.3 (33)
General Anesthesia with Erector Spinae Plane Block (ESPB)	57.7 (45)
Diabetes % (n)	
No	79.5 (62)
Yes	20.5 (16)
Chronic Obstructive Pulmonary Disease (COPD) % (n)	
No	85.9 (67)
Yes	14.1 (11)
Arterial Hypertension % (n)	
No	44.9 (35)
Yes	55.1 (43)
Type of surgery % (n)	
Thoracic Segmentectomy	42.3 (33)
Thoracic Lobectomy	57.7 (45)
Surgical duration (hours)	2.71 ± 1.02
Intensive Care Unit (ICU) in the Post-operative Period % (n)	
No	83.1 (64)
Yes	16.9 (13)
Duration of chest tube (Days)	1.88 ± 1.72
Chest tube in the first 24h % (n)	
No	44.9 (35)
Yes	55.1 (43)
Chest tube in the first 48h % (n)	
No	70.5 (55)
Yes	29.5 (23)
Rescue Opioids in the first 36h % (n)	
No	66.7 (52)
Yes	33.3 (26)
Length of Hospital Stay (Days)	3.39 ± 2.69

The primary outcome occurred in 33.3% of patients (26/78) requiring rescue opioids within 36 hours. Unadjusted analysis showed no significant difference between groups (ESPB 37.8% vs. GA 27.3%, $p = 0.46$) (Table 3). Univariate regression analysis showed no statistically significant associations between postoperative rescue opioid requirements and the evaluated clinical variables; chest tube presence within the first 24 hours did not reach statistical significance ($p = 0.08$) (Table 4). Propensity-weighted analysis identified significant associations: ESPB use was associated with a lower likelihood of postoperative opioid rescue (OR = 0.57; 95% CI 0.32–0.79, $p = 0.01$), and chest tube presence within the first 24 hours was associated with increased odds of postoperative opioid rescue (OR = 2.32; 95% CI 1.22–4.20, $p = 0.03$) (Table 1). Based on marginal predicted probabilities derived from the IPTW-weighted model, ESPB was associated with a 9.7% absolute reduction (95% CI 3.0

to 16.0) in the risk of postoperative rescue opioid consumption within 36 hours. Covariate balancing after propensity score weighting yielded standardized differences that remained within accepted thresholds (Fig. 2).

Notably, the ESPB group received significantly fewer scheduled opioids in the first 24 hours (33% vs. 72%, $p < 0.05$) and 36 hours (40% vs. 75%, $p < 0.05$) (Table 3). Moreover, 42.2% of ESPB patients required no opioids on postoperative day one versus 9% in the GA group ($p = 0.003$). Early moderate-severe pain (first hour) occurred in 11.9% of cases, with no group difference (GA 14.8% vs. ESPB 10%, $p = 0.83$).

No significant intergroup differences were observed in intraoperative medications (fentanyl, methadone, dexmedetomidine, magnesium sulfate; all $p > 0.20$), hospital stay duration (mean 3.5 ± 2.4 GA vs. 3.3 ± 2.9 ESPB, $p = 0.70$), and chest tube days (mean 1.98 ± 1.82 GA vs. 1.80 ± 1.67 ESPB, $p = 0.65$) (Table 3).

Discussion

The present dual-center retrospective study evaluated the effectiveness of the ESPB as part of an opioid-sparing strategy in patients undergoing RATS. Our findings suggest that ESPB use was associated with lower postoperative opioid requirements in the first 24 and 36 hours.

Epidural analgesia, traditionally considered the gold standard for thoracic procedures, presents well-documented complications, including hypotension, urinary retention, and technical difficulties, particularly in minimally invasive surgeries where the risk-benefit balance has shifted.⁶ Paravertebral blocks, although rare, also carry risks of pleural puncture and pneumothorax, which are less acceptable in the current surgical context.⁶ In this scenario, ESPB emerges as a promising alternative, with a favorable safety profile, ease of execution, and minimal risk of serious complications, as evidenced in large case series and systematic reviews.^{5,14,18}

However, despite its increasing popularity, ESPB's exact mechanism of action remains a subject of debate. Several cadaver and Magnetic Resonance Imaging (MRI) studies have demonstrated inconsistent spread patterns of local anesthetics, with variable diffusion into the paravertebral, epidural, and intercostal spaces.^{7,12} The lack of reliable paravertebral or epidural spread raises questions about the primary mechanisms through which ESPB exerts its analgesic effects. While MRI studies (e.g., Schwartzmann et al.,⁸ Sørenstua et al.)¹² have confirmed some degree of paravertebral dissemination, clinical sensory blockade is often unpredictable.¹³ This suggests that mechanisms such as dampening of dorsal rami nociceptive input and systemic absorption of local anesthetics may also play significant roles in ESPB's analgesic effect.

Our study contributes to this discussion by demonstrating that even in RATS, where surgical incisions often extend beyond the expected dermatomal coverage of ESPB, the block was associated with opioid-sparing outcomes. These findings are consistent with a prior case series by Cavaleri et al.¹⁵ and with the randomized clinical trial by Finnerty et al.,¹⁶ both of which reported improved postoperative analgesia with ESPB in minimally invasive thoracic procedures. Notably, Finnerty et al.'s randomized comparison between ESPB and Serratus Anterior Plane (SAP) block demonstrated

STROBE Flow Diagram

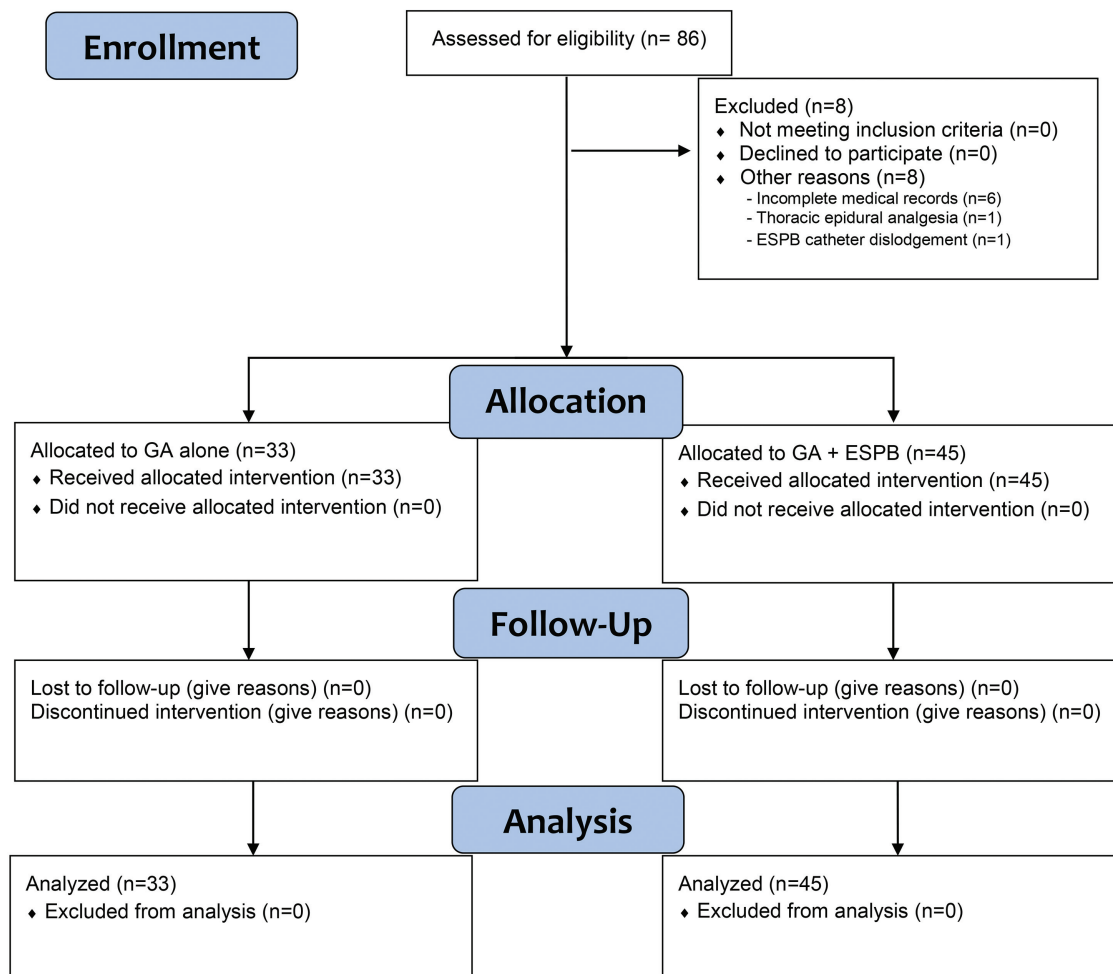


Figure 1 STROBE flow diagram illustrating patient screening, exclusions, and group allocation in this retrospective dual-center cohort study.

superior opioid-sparing effects with ESPB, supporting its preferential use in thoracic surgery settings where deeper analgesia is desired.

Propensity score weighting was employed to balance confounders (e.g., age, comorbidities, intraoperative medications), isolating the effect of ESPB on rescue opioid requirements. The adjusted analysis showed that ESPB use was associated with 43% lower odds of requiring rescue opioids (OR = 0.57; 95% CI 0.32–0.79, $p = 0.01$) compared to General Anesthesia (GA) alone (reference group: OR = 1.00). Notably, this opioid-sparing effect persisted despite a comprehensive multimodal regimen that included systemic agents like dexmedetomidine, magnesium sulfate, and methadone, suggesting that ESPB may provide additive benefits beyond systemic pharmacologic interventions.

The discrepancy between the unadjusted and IPTW adjusted analyses was expected and likely reflected confounding inherent to the retrospective design and the non-randomized allocation of ESPB. Patients were not randomized to receive ESPB, and although they were not selected according to baseline clinical characteristics, the choice of

multimodal analgesic strategy, including the decision to perform ESPB, was made at the discretion of the attending anesthesiologist. This clinical decision-making process, influenced by anticipated surgical pain and intraoperative considerations, may have contributed to differences in perioperative management that obscured the true association in the crude analysis. Variability in intraoperative opioid titration, use of adjunct analgesics, and postoperative prescribing practices may therefore have biased the unadjusted estimates toward the null. IPTW successfully improved covariate balance across measured pre-treatment factors, enabling a more accurate estimation of the independent association between ESPB and postoperative opioid requirements. Although the unadjusted analysis did not demonstrate statistical significance, the adjusted results were consistent with clinical plausibility and aligned with previously published evidence.^{15-17,19}

Methadone should be specifically addressed in the Discussion due to its pharmacological properties, particularly its long elimination half-life and potential prolonged analgesic effects. In this study, methadone was administered as a

Table 3 Patient characteristics, surgical procedures and outcomes of 78 patients undergoing robotic thoracic surgery according to whether they had received Erector Spinae Plane Block with general anesthesia (ESPB) or not (GA).

Variables	Groups		p-value
	GA (33)	ESPB (45)	
Age (years) ^a	65.0 ± 10.8	60.8 ± 13.7	0.15
Gender %(n)			0.88
Female	51.5 (17)	51.1 (23)	
Male	48.5 (16)	48.9 (22)	
BMI (kg.m ⁻²) ^a	27.1 ± 5.3	28.6 ± 4.9	0.22
ASA-PS classification ^b			
I	0.0 (0)	6.7 (3)	0.31
II	81.8 (27)	75.6 (34)	
III	18.2 (6)	17.8 (8)	
Hypertension ^b	57.6 (19)	53.3 (24)	0.32
Diabetes ^b	27.3 (9)	15.6 (7)	0.88
COPD ^b	15.2 (5)	13.3 (6)	0.99
Postoperative prescription of pregabalin ^b	87.9 (29)	80.0 (36)	0.53
Type of surgery ^b			0.98
Thoracic Segmentectomy	42.4 (14)	42.2 (19)	
Thoracic Lobectomy	57.6 (19)	57.8 (26)	
Surgical duration (hours) ^a	2.60 ± 1.3	2.8 ± 0.78	0.38
ICU in the postoperative period ^b	27.3 (9)	9.1 (4)	0.07
Duration of chest tube (days) ^a	1.9 ± 1.8	1.8 ± 1.6	0.65
Chest tube in the first 24 hours ^b	51.5 (17)	57.8 (26)	0.75
Chest tube in the first 48 hours ^b	36.4 (12)	24.4 (11)	0.37
Scheduled opioid prescription (first 24h) ^{#,b}	72.7 (24)	33.3 (15)	0.001
Scheduled opioid prescription (first 36h) ^{#,b}	75.8 (25)	40.0 (18)	0.004
Postoperative opioids rescue ^b			0.46
No	72.7 (24)	62.2 (28)	
Yes	27.3 (9)	37.8 (17)	
Dose of fentanyl ^c (mcg) ^a	258.3 ± 97	235 ± 41.1	0.50
Dose of methadone ^c (mg) ^a	12.4 ± 4.8	13.8 ± 4.1	0.26
Dose of dexmedetomidine ^c (mcg) ^a	95.6 ± 69.5	105.5 ± 56.8	0.53
Dose of magnesium sulfate ^c (mg) ^a	4000 ± 1137.5	3555.6 ± 1318.8	0.22
Hospital length of stay (days) ^a	3.5 ± 2.3	3.2 ± 2.9	0.70

The American Society of Anesthesiologists Physical Status (ASA-PS classification); Confidence Interval (CI); General Anesthesia without Erector Spinae Plane Block (GA); General Anesthesia with Erector Spinae Plane Block (ESPB).

† Postoperative rescue opioid requirement was defined as the occurrence of moderate to severe pain with the need for opioid analgesics (morphine).

^a Values expressed as mean ± standard deviation; #2 Standardized prescription with pre-defined scheduled dosing of oxycodone.

^b Values expressed as % (n); p-values are based on regression analysis.

^c Total cumulative dose of anesthetics administered during the intraoperative period.

single intraoperative dose and was similarly distributed between groups, providing a comparable baseline analgesic effect rather than differential exposure. Moreover, methadone dose was not independently associated with postoperative rescue opioid consumption and was adjusted for in the propensity score model. Notably, its concomitant administration in both groups may have provided sustained background analgesia over the early postoperative period, potentially attenuating differences in pain intensity between groups and, consequently, reducing the observed need for rescue opioids.

After adjustment, ESPB use was independently associated with a significantly lower likelihood of postoperative rescue opioid requirements, consistent with a clinically opioid-sparing association. These findings are in line with Enhanced

Recovery After Surgery (ERAS) protocols for thoracic surgery, which emphasize opioid-sparing strategies aimed at enhancing recovery and minimizing opioid-related adverse events, such as respiratory depression, nausea, and ileus.¹⁹

Notably, scheduled opioid prescriptions in the first 36 hours differed significantly between groups ($p = 0.004$), reinforcing the importance of standardized postoperative protocols. The difference in scheduled opioid prescriptions likely reflects provider-driven analgesic strategies determined after treatment allocation, with ESPB use intentionally associated with reduced fixed opioid dosing. Although patients managed without ESPB and without scheduled opioids could theoretically require more rescue analgesia due to less robust baseline pain control, scheduled opioid prescription represents a post-treatment variable and is better interpreted as a mediator

Table 4 Patient characteristics, surgical procedures and outcomes of 78 patients undergoing curative robotic thoracic surgery according to the postoperative rescue opioids requirements (univariate regression analysis).

Variables	Postoperative Rescue Opioids requirements ^a				p-value
	Total (n = 78)	No (n = 52)	Yes (n = 26)	OR (95% CI)	
Age (years)	62.6 ± 12.6	63.2 ± 12	61.3 ± 12.5	0.99 (0.95 – 1.3)	0.54
Gender					
Female				1.00	
Male	48.7 (38)	53.8 (28)	38.5 (10)	0.54 (0.2 – 1.38)	0.22
BMI (kg.m ⁻²)	28 ± 5.1	28 ± 4.5	27.8 ± 6.2	0.99 (0.9 – 1.09)	0.86
ASA-PS Classification					
I	3.85 (3)	3.8 (2)	3.8 (1)	1.00	
II	78.2 (61)	76.9 (40)	80.8 (21)	1.05 (0.1 – 23.37)	0.97
III	17.9 (14)	19.2 (10)	15.4 (4)	0.8 (0.06 – 20.11)	0.87
Hypertension	55.1 (43)	55.8 (29)	53.8 (14)	0.93 (0.36 – 2.4)	0.87
Diabetes	20.5 (16)	23.1 (12)	15.4 (4)	0.61 (0.15 – 1.98)	0.43
COPD	14.1 (11)	13.5 (7)	15.4 (4)	1.17 (0.28 – 4.31)	0.82
Postoperative pregabalin prescription	16.7 (13)	17.3 (9)	15.4 (4)	0.87 (0.22 – 3)	0.83
Groups					0.13
GA	42.3 (33)	46.2 (24)	34.6 (9)	1.00	
ESPB	57.7 (45)	53.8 (28)	65.4 (17)	1.62 (0.62 – 4.42)	
Type of surgery					0.15
Thoracic Segmentectomy	42.3 (33)	48.1 (25)	30.8 (8)	1.00	
Thoracic Lobectomy	57.7 (45)	51.9 (27)	69.2 (18)	2.08 (0.79 – 5.87)	
Surgical duration (hours)	2.58 (2.19–3.17)	2.58 (1.9–3.19)	2.71 (2.42–3.15)	1.17 (0.73 – 1.9)	0.50
ICU in the postoperative period	16.8 (13)	19.2 (10)	11.5 (3)	0.53 (0.11 – 1.96)	0.38
Duration of Chest Tube (days)	1.8 ± 1.7	1.9 ± 1.8	1.8 ± 1.6	0.99 (0.74 – 1.3)	0.96
Chest tube in the first 24h	55.1 (43)	48.1 (25)	69.2 (18)	2.43 (0.92 – 6.85)	0.08
Chest tube in the first 48h	29.4 (23)	38.4 (10)	25 (13)	1.87 (0.68 – 5.17)	0.22
Schedule opioid prescription (first 24h) ^b	50 (39)	50 (26)	50 (13)	1 (0.39 – 2.58)	1.00
Schedule opioid prescription (first 36h) ^b	55.1 (43)	53.8 (28)	57.7 (15)	1.17 (0.45 – 3.07)	0.75
Dose of fentanyl (mcg)	328.7 ± 229.8	346 ± 260.5	275 ± 65.4	0.48 (0.17 – 1.27)	0.15
Dose of methadone ^c (mg)	13.4 ± 4.3	12.9 ± 4.6	14.6 ± 3.6	1.1 (0.97 – 1.26)	0.16
Dose of dexmedetomidine ^c (mcg)	101.5 ± 61.8	100.5 ± 58.7	103.8 ± 69.4	1 (0.99 – 1.01)	0.84
Dose of Magnesium Sulfate ^c (mg)	3703.7 ± 1268.2	3562.5 ± 1389.6	3909 ± 1064.9	1 (0.99 – 1.11)	0.32

Body Mass Index (BMI) (kg.m⁻²); The American Society of Anesthesiologists Physical Status (ASA-PS classification); Confidence Interval (CI); General Anesthesia without Erector Spinae Plane Block (GA); General Anesthesia with Erector Spinae Plane Block (ESPB).

^a Postoperative rescue opioid requirement was defined as the occurrence of moderate to severe pain with the need for opioid analgesics (morphine) in the first 36h.

^b Standardized prescription with pre-defined scheduled dosing of oxycodone.

^c Total cumulative dose of anesthetics administered during the intraoperative period.

¹Values expressed as mean ± standard deviation.

²Values expressed as %(n); p-values are based on regression analysis.

rather than a traditional confounder. Importantly, scheduled opioid consumption was not independently associated with rescue opioid requirements in univariate analysis (Table 4), suggesting that postoperative prescribing practices alone are unlikely to explain the adjusted opioid-sparing effect of ESPB, while a residual influence cannot be entirely excluded.

The lack of early (first hour) differences in moderate-to-severe pain incidence may be attributed to the multimodal analgesia administered intraoperatively, consistent with previous findings.¹⁷ Conversely, the presence of chest tubes in the first 24 hours significantly increased the odds of requiring opioid rescue, underscoring their well-known

contribution to postoperative discomfort.²⁰ While ESPB provided a protective effect, its ability to fully counteract chest tube-induced pain may be limited, reinforcing the need for comprehensive multimodal approaches.

Limitations

This study has several limitations. Its retrospective design inherently introduced the risk of selection bias, information bias, and heterogeneity in perioperative management. Although propensity score weighting was used to mitigate measured confounding, residual confounding from unmeasured

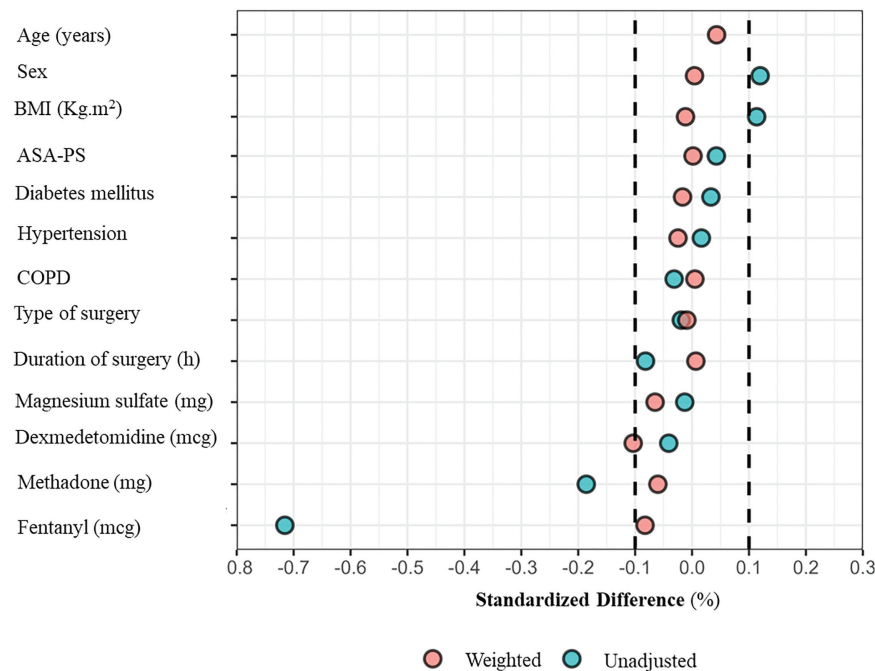


Figure 2 Standardized differences, prior to and following inverse probability-weighting. Abbreviations: Body Mass Index (BMI) (kg.m⁻²); Chronic Obstructive Pulmonary Disease (COPD); The American Society of Anesthesiologists Physical Status (ASA-PS classification).

variables could not be fully excluded. Patients were not randomized to receive ESPB; instead, anesthesiology coverage was determined by standard departmental scheduling, which may have introduced differences in perioperative management practices across providers. Furthermore, the relatively small sample size may have increased the risk of unstable propensity score weights and model sensitivity, potentially influencing effect estimates. Despite efforts to minimize selective application of ESPB, confounding by indication related to the choice of ESPB versus general anesthesia alone remained a potential limitation inherent to this observational study.

Sensory block mapping was not performed, precluding direct correlation between dermatomal coverage and analgesic efficacy. Intraoperative opioid selection was not fully standardized; although methadone was similarly distributed between groups and adjusted for in the propensity score model, its prolonged and variable pharmacokinetics may have contributed to background analgesia in both groups. Scheduled postoperative opioid prescriptions represented post treatment variables and may have acted as mediators rather than traditional confounders, although they were not independently associated with rescue opioid requirements in univariate analysis. In addition, pain scores were not consistently documented in the medical records, and in some cases only analgesic administration was available, precluding detailed analysis of pain trajectories over time.

All analyses were conducted using complete case data because patients with incomplete medical records were excluded prior to analysis; although the number of exclusions was small, this approach may still have introduced selection bias and should be acknowledged as a methodological limitation. Finally, although no formal sensitivity analyses were performed due to sample size constraints,

inspection of the propensity score distributions and inverse probability treatment weights did not suggest instability or undue influence of extreme weights. Nevertheless, the absence of additional formal robustness assessments should be considered when interpreting the results.

Generalizability

The generalizability of these findings should be interpreted with caution. This retrospective analysis included a limited sample of patients undergoing elective RATS for lung cancer at two high-volume tertiary centers, all managed by experienced surgical and anesthesiology teams following ERAS-aligned protocols. The ESPB technique was standardized in terms of puncture level, dosing, and continuous infusion, which strengthens internal validity but may limit applicability to lower-volume institutions or settings with less expertise in ultrasound-guided regional anesthesia. Furthermore, the exclusion of emergency cases and patients with chronic pain or prior opioid consumption restricts extrapolation to more complex populations or larger thoracic procedures. Prospective multicenter studies are warranted to validate these findings across diverse practice environments.

Conclusion

In this retrospective cohort, ESPB was associated with reduced postoperative opioid requirements in patients undergoing robotic thoracic surgery. Although these findings support ESPB as a potentially valuable component of multimodal, opioid-sparing analgesic strategies aligned with ERAS principles, causality cannot be established due to the observational

design. Residual confounding and variability in perioperative analgesic management may have influenced the observed associations. Therefore, the role of ESPB in robotic thoracic surgery should be further evaluated in prospective randomized trials with standardized analgesic protocols, detailed sensory mapping, and long-term follow-up to confirm efficacy and clarify underlying mechanisms.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies

The authors declare that no generative artificial intelligence or AI-assisted technologies were used in the preparation of this manuscript.

Disclosures

All authors (1) report no relationships that could be construed as a conflict of interest, (2) take responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation, and (3) have approved the final article.

Authors' contributions

All authors have made substantial contributions to (1) The conception and design of the work, (2) The acquisition, analysis, and interpretation of data, and (3) The manuscript draft and review of important intellectual content. Matheus Arraes, Sara Amaral, Saullo Queiroz Silveira, Leopoldo Muniz da Silva and Rafael Sousa Fava Nersessian were responsible for study conception, design, data screening/acquisition, analysis, interpretation, manuscript draft, and figures/tables creation. Caroline Machado Nunes contributed to study with data collection. Francisco Jose Lucena Bezerra and Manoel de Souza Neto contributed with manuscript/figures/tables review and manuscript draft. Glenio B. Mizubuti and Rafael Lombardi contributed with data interpretation/review, and manuscript draft.

Conflicts of interest

The authors declare no conflicts of interest.

Associate Editor

Eric Benedet Lineburger

References

- Cheng D, Downey RJ, Kernstine K, et al. Video-Assisted Thoracic Surgery in Lung Cancer Resection: A Meta-Analysis and Systematic Review of Controlled Trials. *Innov Technol Tech Cardiothorac Vasc Surg.* 2007;2:261–92.
- Geraci TC, Sasankan P, Luria B, Cerfolio RJ. Intraoperative Anesthetic and Surgical Concerns for Robotic Thoracic Surgery. *Thorac Surg Clin.* 2020;30:293–304.
- Zhang LY, Zhang YH, Shen J, Luo Y. Effects of dexmedetomidine on post-operative recovery and mental status in patients receiving robotic-assisted thoracic surgery. *Ann Palliat Med.* 2019;8:469–75.
- Forero M, Adhikary SD, Lopez H, Tsui C, Chin KJ. The Erector Spinae Plane Block: A Novel Analgesic Technique in Thoracic Neuropathic Pain. *Reg Anesth Pain Med.* 2016;41:621–7.
- Tsui BCH, Fonseca A, Munshey F, McFadyen G, Caruso TJ. The erector spinae plane (ESP) block: A pooled review of 242 cases. *J Clin Anesth.* 2019;53:29–34.
- Feray S, Lubach J, Joshi GP, Bonnet F, Van De Velde M. the PROSPECT Working Group *of the European Society of Regional Anaesthesia and Pain Therapy. PROSPECT guidelines for video-assisted thoracoscopic surgery: a systematic review and procedure-specific postoperative pain management recommendations. *Anaesthesia.* 2022;77:311–25.
- Ivanusic J, Konishi Y, Barrington MJ. A Cadaveric Study Investigating the Mechanism of Action of Erector Spinae Blockade. *Reg Anesth Pain Med.* 2018;43:567–71.
- Schwartzmann A, Peng P, Maciel MA, Alcarraz P, Gonzalez X, Forero M. A magnetic resonance imaging study of local anesthetic spread in patients receiving an erector spinae plane block. *Can J Anaesth J Can Anesth.* 2020;67:942–8.
- Ahiskalioglu A, Tulgar S, Celik M, Ozer Z, Alici HA, Aydin ME. Lumbar Erector Spinae Plane Block as a Main Anesthetic Method for Hip Surgery in High Risk Elderly Patients: Initial Experience with a Magnetic Resonance Imaging. *Eurasian J Med.* 2020;52:16–20.
- Schwartzmann A, Peng P, Maciel MA, Forero M. Mechanism of the erector spinae plane block: insights from a magnetic resonance imaging study. *Can J Anesth Can Anesth.* 2018;65:1165–6.
- Celik M, Tulgar S, Ahiskalioglu A, Alper F. Is high volume lumbar erector spinae plane block an alternative to transforaminal epidural injection? Evaluation with MRI. *Reg Anesth Pain Med.* 2019;44:906–7.
- Sørenstua M, Zantalis N, Raeder J, Vamnes JS, Leonardsen ACL. Spread of local anesthetics after erector spinae plane block: an MRI study in healthy volunteers. *Reg Anesth Pain Med.* 2023;48:74–9.
- Byrne K, Smith C. Human volunteer study examining the sensory changes of the thorax after an erector spinae plane block. *Reg Anesth Pain Med.* 2020;45:761–2.
- Kot P, Rodriguez P, Granell M, et al. The erector spinae plane block: a narrative review. *Korean J Anesthesiol.* 2019;72:209–20.
- Cavaleri M, Tigano S, Nicoletti R, et al. Continuous Erector Spinae Plane Block as Postoperative Analgesic Technique for Robotic-Assisted Thoracic Surgery: A Case Series. *J Pain Res.* 2021;14:3067–72.
- Finnerty DT, McMahon A, McNamara JR, Hartigan SD, Griffin M, Buggy DJ. Comparing erector spinae plane block with serratus anterior plane block for minimally invasive thoracic surgery: a randomized clinical trial. *Br J Anaesth.* 2020;125:802–10.
- Zhang J, Liu TX, Wang WX, Zhou SZ, Ran X, He P. Effects of ultrasound-guided erector spinae plane block on postoperative acute pain and chronic post-surgical pain in patients underwent video-assisted thoracoscopic lobectomy: a prospective randomized, controlled trial. *BMC Anesthesiol.* 2023;23:161.

18. De Cassai A, Geraldini F, Carere A, Sergi M, Munari M. Complications Rate Estimation After Thoracic Erector Spinae Plane Block. *J Cardiothorac Vasc Anesth*. 2021;35:3142–3.
19. Batchelor TJP, Rasburn NJ, Abdelnour-Berchtold E, et al. Guidelines for enhanced recovery after lung surgery: recommendations of the Enhanced Recovery After Surgery (ERAS[®]) Society and the European Society of Thoracic Surgeons (ESTS). *Eur J Cardiothorac Surg*. 2019;55:91–115.
20. Zengin M, Ulger G, Baldemir R, Sazak H, Alagoz A. Is there a relationship between body mass index and postoperative pain scores in thoracotomy patients with thoracic epidural analgesia? *Medicine (Baltimore)*. 2021;100:e28010.

ORIGINAL INVESTIGATION

Coronary artery bypass grafting in Jehovah's Witness patients: a retrospective propensity matched cohort study



Henrique Ryu Yamanaka Nakano ^{a,*}, Guilherme Mota Filho ^a,
Bruno Camargo Calabria Araújo ^a, Heloísa Zogheib ^a,
Enzo Camargo Calabria Araújo ^a, Leonardo Cardoso Saraiva ^a, Jessica Silva Nicolau ^b,
Suely Pereira Zeferino ^b, José Otávio Costa Auler Junior ^b

^a Hospital de Clínicas da Faculdade de Medicina da Universidade de São Paulo, Anesthesiology Institute of Surgery Department, São Paulo, SP, Brazil

^b Hospital de Clínicas da Faculdade de Medicina da Universidade de São Paulo, Heart Institute, São Paulo, SP, Brazil

Received 28 September 2025; accepted 3 May 2026

Available online 13 May 2026

KEYWORDS

Blood component
transfusion;
Coronary artery
bypass;
Jehovah's Witnesses;
Perioperative care;
Transfusion-free
surgery

Abstract

Background: Coronary Artery Bypass Grafting (CABG) in Jehovah's Witness (JW) patients raises concerns regarding perioperative safety because of refusal of allogeneic blood transfusions. This study evaluated the feasibility and short-term outcomes of CABG in JW patients and described perioperative blood management practices.

Methods: This retrospective observational study included 26 JW patients and 78 matched non-Jehovah's Witness (non-JW) patients who underwent CABG at InCor HC-FMUSP between 2015 and 2023. Propensity score matching at a 3:1 ratio based on age, sex, and preoperative hemoglobin was used to improve comparability. The primary outcome was 30-day all-cause mortality. Secondary outcomes included 30-day stroke, myocardial infarction, and acute kidney injury; ICU length of stay; perioperative red blood cell transfusion rates; postoperative hemoglobin and hematocrit; intraoperative blood loss; and cardiopulmonary bypass time.

Results: No 30-day deaths occurred among JW patients, whereas three deaths occurred in the non-JW group (OR = 0.45; 95% CI 0.003–5.09; p = 0.575). No JW patient experienced stroke, and one myocardial infarction occurred. JW patients had higher postoperative hemoglobin (OR = 1.50; 95% CI 1.03–2.28; p = 0.032) and lower intraoperative blood loss (OR = 0.55; 95% CI 0.27–0.98; p = 0.046). No JW patient received transfusion, whereas 17.9% and 12.8% of non-JW patients required intraoperative and postoperative transfusions, respectively. ICU length of stay was similar between groups.

* Corresponding author.

E-mail: nakanohenrique@gmail.com (H.R. Nakano).

Conclusion: In this cohort, CABG in JW patients was feasible with acceptable short-term outcomes. Outcomes were generally comparable between groups, acknowledging imprecision related to the small JW sample.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Advances in cardiac surgery have substantially improved survival and quality of life among patients undergoing coronary artery bypass grafting (CABG).¹ Despite these advances, clinical management may be influenced by patient-specific factors, including religious beliefs that restrict certain therapeutic options. Jehovah's Witnesses (JW), for example, refuse allogeneic blood transfusions for religious reasons, a decision that can pose complex clinical and ethical challenges in procedures associated with significant bleeding risk.^{2,3} In the context of CABG, this refusal has traditionally raised concerns regarding perioperative safety and, in some settings, has led to delays or even denial of surgical treatment.⁴⁻⁶

Although prior studies have reported acceptable CABG outcomes in JW patients, most originate from selected populations or high-resource centers. Data from public health-care systems remain limited, particularly in contexts characterized by prolonged waiting times and variable access to blood conservation technologies. Our findings contribute to evidence from such a setting.

Blood loss during CABG is a major determinant of perioperative morbidity and mortality.⁷ While transfusion practices were historically liberal, contemporary evidence supports restrictive strategies in cardiac surgery, emphasizing meticulous hemostasis, judicious blood product use, and structured perioperative care.^{8,9} For patients who refuse transfusion, these principles are essential and require coordinated blood conservation approaches, including preoperative anemia management, intraoperative cell salvage, and pharmacologic optimization.

This study evaluated perioperative management, implementation of Patient Blood Management (PBM) strategies, and short-term outcomes of JW patients undergoing CABG in a contemporary public hospital using a retrospective, propensity-matched observational design, with focus on 30-day morbidity and mortality.

Methods

Study design

This was a single-center, retrospective observational cohort study comparing JW and non-JW patients undergoing isolated CABG. This manuscript was prepared in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline. A completed STROBE checklist is provided as supplementary material.

Ethics approval and consent

All JW patients underwent individualized preoperative informed consent, during which acceptable and non-acceptable blood products and hemostatic agents were discussed

and documented in the medical record, in accordance with patient autonomy and institutional ethical standards. Owing to the retrospective design, no standardized registry allowed systematic determination of how many patients explicitly consented to transfusion in life-threatening situations, although such discussions were part of routine care.

This retrospective observational study used anonymized medical record data without direct patient contact. The requirement for informed consent was waived by the Ethics Committee (Approval n° 7.441.177; CAAE 85677324.8.0000.0068; March 14, 2025).

Participants

We included 26 JW and 78 matched non-JW patients who underwent isolated CABG at the Heart Institute, University of São Paulo, between January 2015 and December 2023.

Inclusion and exclusion criteria

Adults (> 18 years) undergoing first-time isolated CABG were included. Patients undergoing reoperation or combined procedures (e.g., valve or aortic surgery) were excluded.

Patient selection

The JW cohort (n = 26) comprised all consecutive JW patients undergoing isolated CABG during the study period. No a priori sample size calculation was performed, as this represented the total eligible JW population. Controls were selected in a 1:3 ratio from 566 non-JW patients who underwent isolated CABG by the same surgeon during the same period. As this surgeon is the only operator performing cardiac surgery in JW patients at our institution, restricting controls to his case list was intended to reduce surgeon-related variability and improve comparability before propensity score matching. The screening and matching process is summarized in the Figure 1 below.

Statistical analysis

To address the sample size disparity between JW and non-JW patients, Propensity Score Matching (PSM) was performed using age, sex, and preoperative hemoglobin.^{10,11} We used nearest-neighbor matching, without replacement and without a caliper restriction. The case-to-control ratio was selected via a simulation-based power analysis evaluating ratios from 1:1 to 5:1.¹² In each simulation, a continuous predictor and a binary outcome with fixed log-odds were generated using 26 cases and varying control ratios. Firth-penalized logistic regression was fit in 500 simulations per scenario, with power defined as the proportion of p-values < 0.05.¹³ As shown in Figure 2, power plateaued at a 3:1 ratio, which was adopted.

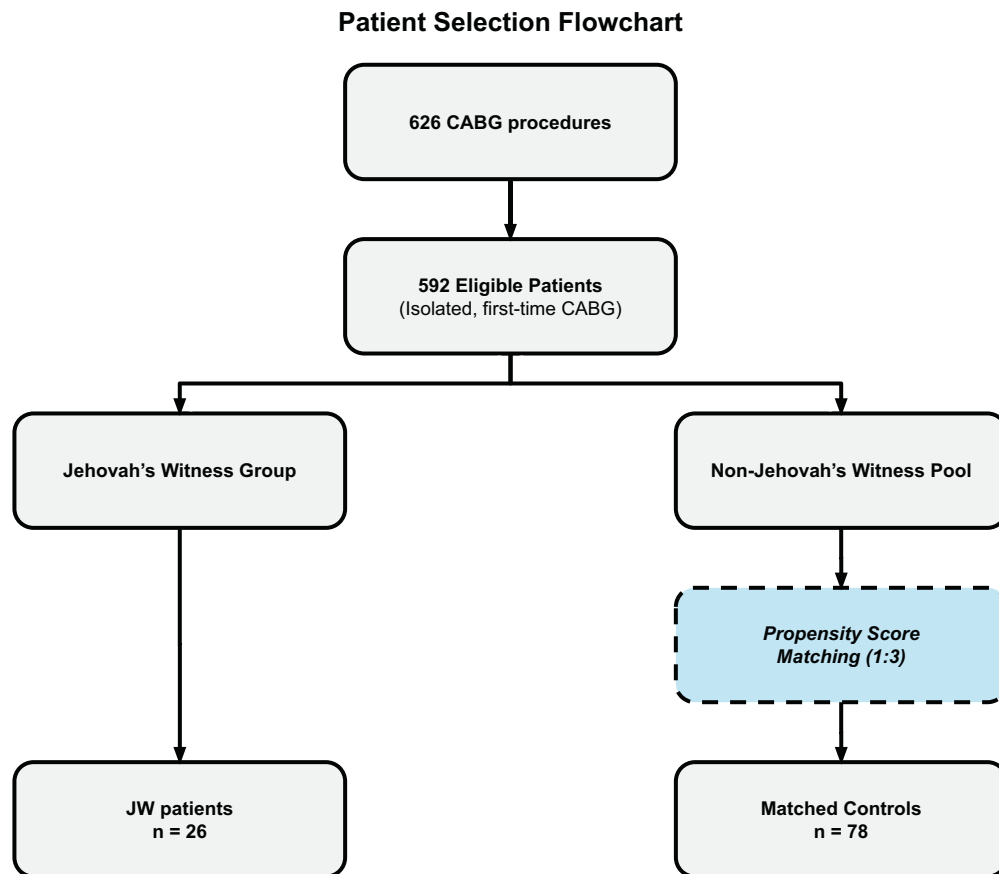


Figure 1 Flow diagram of patient selection and matching CABG, Coronary artery bypass grafting. Patients excluded from the initial screening ($n = 34$) consisted of reoperations or combined procedures. Propensity score matching was performed using a 1:3 ratio to ensure balanced baseline characteristics.

Baseline demographic and clinical characteristics were compared between JW and non-JW patients using Chi-square tests for categorical variables and t -tests or Mann-Whitney U -tests for normal and non-normal continuous variables, respectively. Logistic regression compared groups across pre-, intra-, and postoperative outcomes. More specifically, in each model, group was modeled as the dependent variable and the respective clinical outcome as the independent variable. Models were additionally adjusted for age, sex, and preoperative hemoglobin to account for residual imbalance after matching.¹⁴ Firth's correction was used due to binary outcomes with rare events (including zero counts) like transfusion and death.¹³ The Shapiro-Wilk test assessed normality of continuous variables; non-normal variables were log-transformed. Odds Ratios (ORs) were calculated by exponentiating model coefficients, with 95% Confidence Intervals estimated via profile likelihood; p -values were not adjusted for multiple comparisons to preserve statistical power.

We performed a sensitivity analysis to assess the potential impact of clinical characteristics on the assessed outcomes. This was done by including all of the available clinical characteristics (EuroSCORE II, ASA status, LVEF < 50%, diabetes, and hypertension) in the logistic regression models and re-running them. Given its role as a validated composite predictor of surgical mortality and morbidity in CABG, adjustment for EuroSCORE II was considered the main

analytic framework for outcome estimation. This approach was selected to provide clinically meaningful risk adjustment beyond the matching variables while preserving model stability in a small sample.

Because of the small number of JW patients and the risk of model overparameterization, the propensity score was restricted to key baseline variables directly associated with transfusion exposure and perioperative risk. EuroSCORE II, as a composite risk index, was included in regression adjustment models rather than in the matching algorithm to maintain balance and model stability.

There were no missing data for variables included in the analysis. All analyses were conducted using R (R Core Team, 2025) using the MatchIt¹⁵ and Logistf packages.¹⁶

Predictors

Preoperative predictors

Included age, sex, American Society of Anesthesiologists (ASA) status (1–5), EuroSCORE II, preoperative Hemoglobin (Hb) and Hematocrit (Ht) measured the day before surgery, and comorbidities (diabetes mellitus, hypertension, dyslipidemia, Coronary Artery Disease [CAD], and heart failure with Left Ventricular Ejection Fraction [LVEF] < 50%). Reported Hb values reflect measurements obtained immediately before surgery and therefore incorporate any

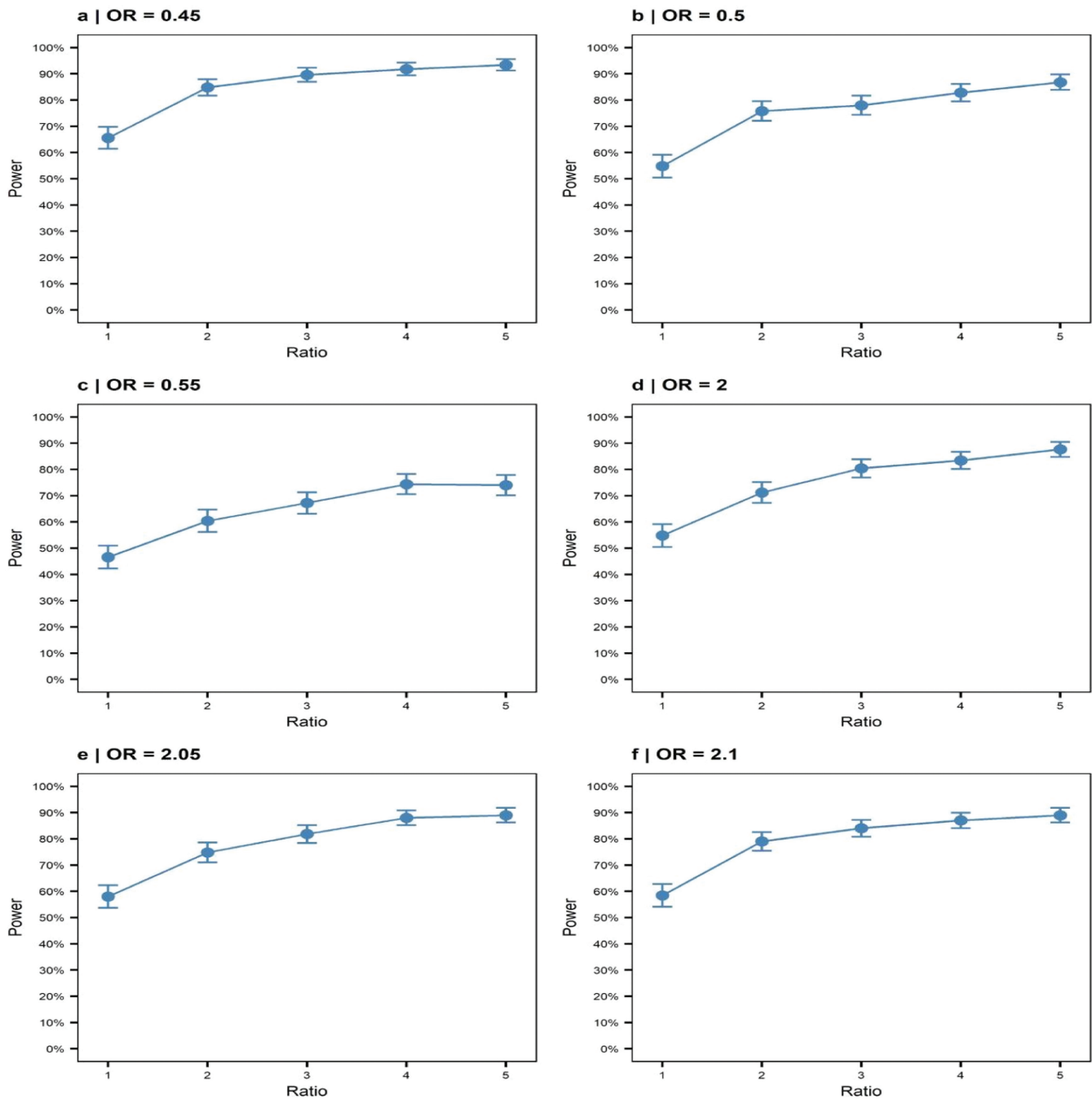


Figure 2 Simulation-based power analysis for defining the best control-to-case ratio to be used in propensity score matching. Each panel shows power estimates ($\pm 95\%$ CI) as a function of the control-to-case ratio (x-axis), assuming a fixed number of cases ($n=26$) and a corresponding fixed number of controls ($n = \text{ratio} \times 26$). Panels (a–c) show scenarios with $OR < 1$ (protective effects), and panels (d–f) show $OR > 1$ (risk effects). Each curve is based on 500 simulations. Power increases with larger control-to-case ratios, with diminishing gains beyond a 1:3 ratio.

preoperative optimization, including iron or erythropoietin when used. Diabetes, hypertension, and dyslipidemia were defined by prior diagnosis or ongoing pharmacologic treatment. CAD was defined as prior myocardial infarction, coronary revascularization, or angiographically confirmed disease. Left ventricular dysfunction was defined as $LVEF < 50\%$.

Perioperative blood management

Preoperative anemia optimization differed between groups. In JW patients, iron and erythropoietin were used selectively based on Hb levels, surgical urgency, and patient consent. When erythropoietin was administered, surgery was performed after achieving clinically acceptable Hb levels (typically $\geq 12.0 \text{ g.dL}^{-1}$), balancing optimization with urgency. In

non-JW patients, anemia management followed routine institutional practice without a standardized protocol. Tranexamic acid was routinely administered intraoperatively according to institutional CABG protocols. No additional hemostatic agents were systematically used preoperatively in either group.

Transfusion policy

In the non-JW group, transfusion decisions were based on multifactorial clinical assessment rather than a fixed Hb threshold, incorporating active bleeding, hemodynamic instability, tissue perfusion, and overall clinical context. In this study, “transfusion” refers exclusively to allogeneic Red Blood Cell (RBC) transfusion; use of plasma, platelets, or other hemostatic agents (e.g., fibrinogen or prothrombin complex concentrate) was not analyzed, representing a limitation of the study. No JW patient required transfusion. Per institutional policy, management prioritized blood conservation and avoidance of transfusion in accordance with documented patient preferences. A contingency plan allowed RBC transfusion in life-threatening situations with imminent risk of death after multidisciplinary discussion and ethical review; no such cases occurred.

Intraoperative predictors

Included Cardiopulmonary Bypass (CPB) duration (< 60, 60–120, > 120 minutes), estimated blood loss, and intraoperative transfusion. All JW patients had access to intraoperative cell salvage. Among non-JW controls, cell-saver use was available in 43 of 78 cases, reflecting institutional resource availability rather than clinical selection. Perioperative anticoagulant and antiplatelet management followed institutional CABG protocols, including preoperative discontinuation and postoperative resumption according to bleeding and thrombotic risk. Blood loss was estimated by sponge weight and suction volume, a commonly used method with recognized limitations, including estimation error, occult loss, and dilution by irrigation fluids.

Outcomes

The primary outcome was 30-day all-cause mortality. Secondary outcomes included 30-day stroke, Myocardial Infarction (MI), and Acute Kidney Injury (AKI); duration of mechanical ventilation; Intensive Care Unit (ICU) length of stay; perioperative Red Blood Cell (RBC) transfusion; use of vasoactive drugs; postoperative hemoglobin and hematocrit; intraoperative blood loss; and Cardiopulmonary Bypass (CPB) time.

The first hemoglobin and hematocrit measurement after ICU admission was used as the postoperative reference to reduce confounding from repeated sampling or non-surgical blood loss, acknowledging potential volumetric shifts. Death was defined as intraoperative or all-cause mortality within 30 days. Vasopressor use included norepinephrine, epinephrine, or vasopressin; inotropes included dobutamine, epinephrine, or milrinone. Postoperative vasodilators (e.g., nitroprusside, nitroglycerin, hydralazine) were used for blood pressure control.

Peri-procedural MI was defined according to contemporary CABG criteria based on biomarker elevation with electrocardiography or imaging evidence. Stroke was defined as a new neurological deficit lasting > 24 hours, confirmed by

neuroimaging after clinical suspicion. AKI was defined per Kidney Disease: Improving Global Outcomes (KDIGO) criteria, using baseline creatinine measured within 48 hours preoperatively and assessed during the first 7 postoperative days or until discharge.

Results

Demographic characteristics (Table 1)

Baseline demographic and clinical characteristics were comparable between groups. Mean age was similar in JW and non-JW patients (63.3 vs. 63.1 years), and women represented 30.8% and 32.1% of each group, respectively. EuroSCORE II values were also comparable (1.35 vs. 1.34), indicating similar low-risk surgical profiles.

Ten JW patients received preoperative iron supplementation or erythropoietin, whereas none of the non-JW patients underwent anemia optimization. Preoperative anemia, defined according to WHO criteria (hemoglobin < 13.0 g.dL⁻¹ in males and < 12.0 g.dL⁻¹ in females), was present in four JW patients (all male) and in 22 non-JW patients (15 males and 7 females). In some cases, particularly among non-JW patients, hemoglobin thresholds were not fully achieved; however, values were close to reference ranges, and surgical fitness was confirmed after multidisciplinary assessment considering clinical stability, urgency, and public healthcare constraints.

Group matching

PSM matched all 26 Jehovah's Witness patients to 78 controls. The median propensity score difference between each matched case and their respective controls was < 0.001 (maximum: < 0.001), indicating near-exact matching. Additionally, balance diagnostics revealed substantially improved standardized mean differences between groups after matching (distance: -0.002; age: 0.020; male sex: 0.013; preoperative hemoglobin: 0.025), whereas the corresponding pre-matching differences were larger (up to 0.092). These results are shown in Figure 3.

Primary outcome: 30-day all-cause mortality

Regarding the primary outcome, no deaths occurred among JW patients within 30 days, whereas three deaths were observed in the non-JW group (Table 2). Estimates were imprecise due to low event counts.

Secondary outcomes: major clinical events

Within 30 days, none of the JW patients experienced stroke, and one MI (3.8%) was observed. AKI occurred in seven JW patients (26.9%). In the non-JW group, one MI (1.3%), two strokes (2.6%), and 29 cases of AKI (37.2%) were recorded (Table 2 and Fig. 4). Differences between groups for these outcomes were not statistically significant.

Table 1 Baseline characteristics and preoperative data of Jehovah's Witnesses and matched Control patients.

Characteristics	Jehovah's Witnesses (n = 26)	Controls (n = 78)	p-value
Demographics			
Age (yr)	63.4 (10.35)	63.2 (9.11)	0.647
Male sex, n (%)	18 (69.2)	53 (67.9)	0.903
Clinical Status			
EuroSCORE II	1.34 (0.87)	1.35 (1.40)	0.925
Preoperative hemoglobin (g.dL ⁻¹)	13.7 (1.01)	13.3 (1.58)	134
LVEF < 50%, n (%)	5 (19.2)	15 (19.2)	1
ASA Physical Status, n (%)			
II	0 (0.0)	11 (14.5)	0.387
III	19 (73.1)	44 (57.9)	
IV	7 (26.9)	21 (27.6)	
Comorbidities, n (%)			
Hypertension	22 (84.6)	70 (89.7)	0.478
Diabetes mellitus	14 (53.8)	41 (52.6)	0.910

Values are mean (SD) or number (proportion). p-values denote differences between groups using Student's *t*-test or Mann-Whitney *U*-test for continuous variables and Chi-Square or Fisher's exact test for categorical variables. ASA, American Society of Anesthesiologists; LVEF, Left Ventricular Ejection Fraction.

Secondary outcomes: transfusion and perioperative blood management (Fig. 4)

None of the 26 JW patients required intraoperative or postoperative RBC transfusion. In contrast, 14 non-JW patients (17.9%) received intraoperative RBC transfusions (OR = 0.056;

95% CI 0–0.707; *p* = 0.02) and 10 non-JW (12.8%) received postoperative RBC transfusions (OR = 0.101; 95% CI 0.0001–0.884; *p* = 0.035), respectively; Table 2 and Fig. 2). Cell-saver use was universal among JW patients (26/26, 100%) and was available for 43 of 78 non-JW patients (55.1%).

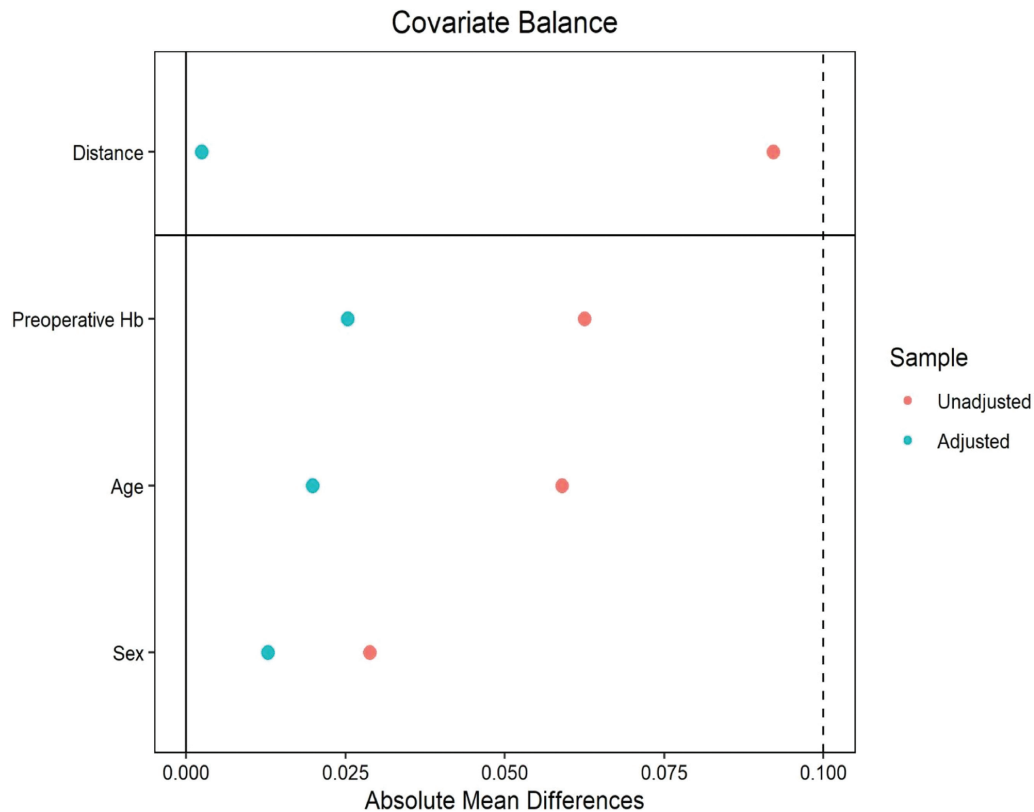


Figure 3 Balance between covariates before and after propensity score matching. Absolute Standardized Mean Differences (SMDs) are shown for each covariate comparing Jehovah's Witnesses vs controls before matching (Unadjusted, red) and after matching (Adjusted, blue; 1:3 ratio). The vertical dashed line denotes the conventional balance threshold ($|SMD| = 0.10$), with values closer to 0 indicating better covariate balance. "Distance" denotes the propensity score (matching distance).

Table 2 Logistic regression results adjusted for matching variables (age, sex, preoperative hemoglobin) + EuroSCORE II.

Variable	Study Group: Jehovah's Witnesses (n = 26)	Study Group: Controls (n = 78)	OR (95% CI)	p-value
CPB duration (hours) ^a	2.04 (0.61)	2.11 (0.54)	0.610 (0.202–1.601)	0.287
Intraoperative bleeding (mL) ^a	241.76 (105.00)	383.39 (230.51)	0.557 (0.272–0.996)	0.048 ^b
Intraoperative transfusion	0 (0.0%)	14 (17.9%)	0.081 (0.001–0.804)	0.028 ^b
Postoperative death	0 (0.0%)	3 (3.8%)	0.427 (0.003–4.741)	0.543
Postoperative hb (g.dL ⁻¹)	11.37 (1.26)	10.49 (1.70)	1.491 (1.028–2.257)	0.035 ^b
Postoperative hct (%)	34.77 (3.84)	31.65 (4.87)	1.189 (1.048–1.374)	0.006 ^c
Postoperative ICU stay (days) ^a	4.50 (3.68)	4.59 (4.58)	1.160 (0.565–2.335)	0.679
Postoperative myocardial infarction	1 (3.8%)	1 (1.3%)	2.806 (0.208–37.974)	0.403
Postoperative renal dysfunction	7 (26.9%)	29 (37.2%)	0.625 (0.217–1.673)	0.355
Postoperative stroke	0 (0.0%)	2 (2.6%)	0.636 (0.004–9.738)	0.773
Postoperative transfusion	0 (0.0%)	10 (12.8%)	0.097 (0.001–0.850)	0.032 ^b

Data: mean [SD] for continuous variables, n (%) for categorical. OR, Odds Ratio; CI, Confidence Interval. All models adjusted for matching variables (age, sex, preoperative hemoglobin) + EuroSCORE II.

^a Log transformation applied when either group showed non-normality (Shapiro-Wilk $p < 0.05$). p-values are uncorrected.

^b $p < 0.05$.

^c $p < 0.01$.

Secondary outcomes: hematologic and intraoperative variables (Fig. 4)

Preoperative hemoglobin levels were similar between groups (13.70 vs. 13.20 g.dL⁻¹; $p = 0.134$). Postoperatively, JW patients had higher hemoglobin (11.37 vs. 10.49 g.dL⁻¹; OR = 1.49; 95% CI 1.02–2.25; $p = 0.035$) and hematocrit levels (34.77% vs. 31.65%; OR = 1.18; 95% CI 1.04–1.37; $p = 0.006$). Cardiopulmonary bypass time was comparable between groups. Intraoperative blood loss was lower in JW patients (241 mL vs. 383 mL; OR = 0.55; 95% CI 0.27–0.99; $p = 0.048$; Table 2 and Fig. 2).

Secondary outcomes: postoperative course (Fig. 4)

Duration of mechanical ventilation and intensive care unit length of stay were similar between groups (Table 2).

Sensitivity analysis

Sensitivity analyses adjusting for additional clinical covariates (EuroSCORE II, ASA status, LVEF < 50%, diabetes, and hypertension; Table 1) yielded results consistent with the primary models, without meaningful changes in effect estimates (Supplementary Tables 1–5).

Discussion

This study describes perioperative management and short-term outcomes of JW patients undergoing CABG within a public tertiary healthcare system. Rather than providing definitive estimates of safety for rare events, our findings primarily offer an insight into the feasibility and implementation of structured PBM strategies in a transfusion-restricted population under routine public-hospital conditions.

The primary finding of this study concerns 30-day all-cause mortality, with no deaths observed among JW patients and no

clear difference in estimated risk compared with non-JW patients, acknowledging the limited power to detect rare events. Beyond mortality, JW patients had lower exposure to RBC transfusion, a finding consistent with prior observational data and randomized evidence supporting restrictive transfusion strategies in cardiac surgery. Studies such as those by Chambault et al.¹⁷ and the TRICS III trial⁹ have demonstrated that restrictive transfusion approaches are not associated with worse clinical outcomes and may reduce transfusion-related risks. Within this framework, our results support the feasibility of CABG in transfusion-restricted patients while reinforcing the relevance of PBM principles, noting that these secondary findings remain exploratory. Furthermore, the well-established risks associated with allogeneic transfusion, including infection, immune modulation, and increased morbidity and mortality, support the rationale for restrictive transfusion strategies, as demonstrated in landmark trials by Carson et al.⁸ and Hébert et al.¹⁸

In addition, JW patients presented with higher postoperative hemoglobin levels and lower intraoperative blood loss compared with matched controls. These findings are consistent with structured blood conservation efforts and enhanced perioperative vigilance. As highlighted by Shander et al.,¹⁹ the JW population played a key role in the development of PBM principles, with implications extending beyond religious contexts.²⁰ For patients who categorically refuse transfusion, strategies such as meticulous hemostasis, avoidance of unnecessary hemodilution, intraoperative cell salvage, and pharmacologic optimization are essential. When applied systematically, these principles may also benefit the broader CABG population, although their direct impact on hard clinical outcomes remains uncertain.

At our institution, surgical eligibility for JW patients did not rely on a fixed hemoglobin cutoff and decisions were individualized. We agree that higher preoperative hemoglobin levels are desirable and may offer greater physiological reserve in transfusion-restricted patients. However, optimal Hb targets

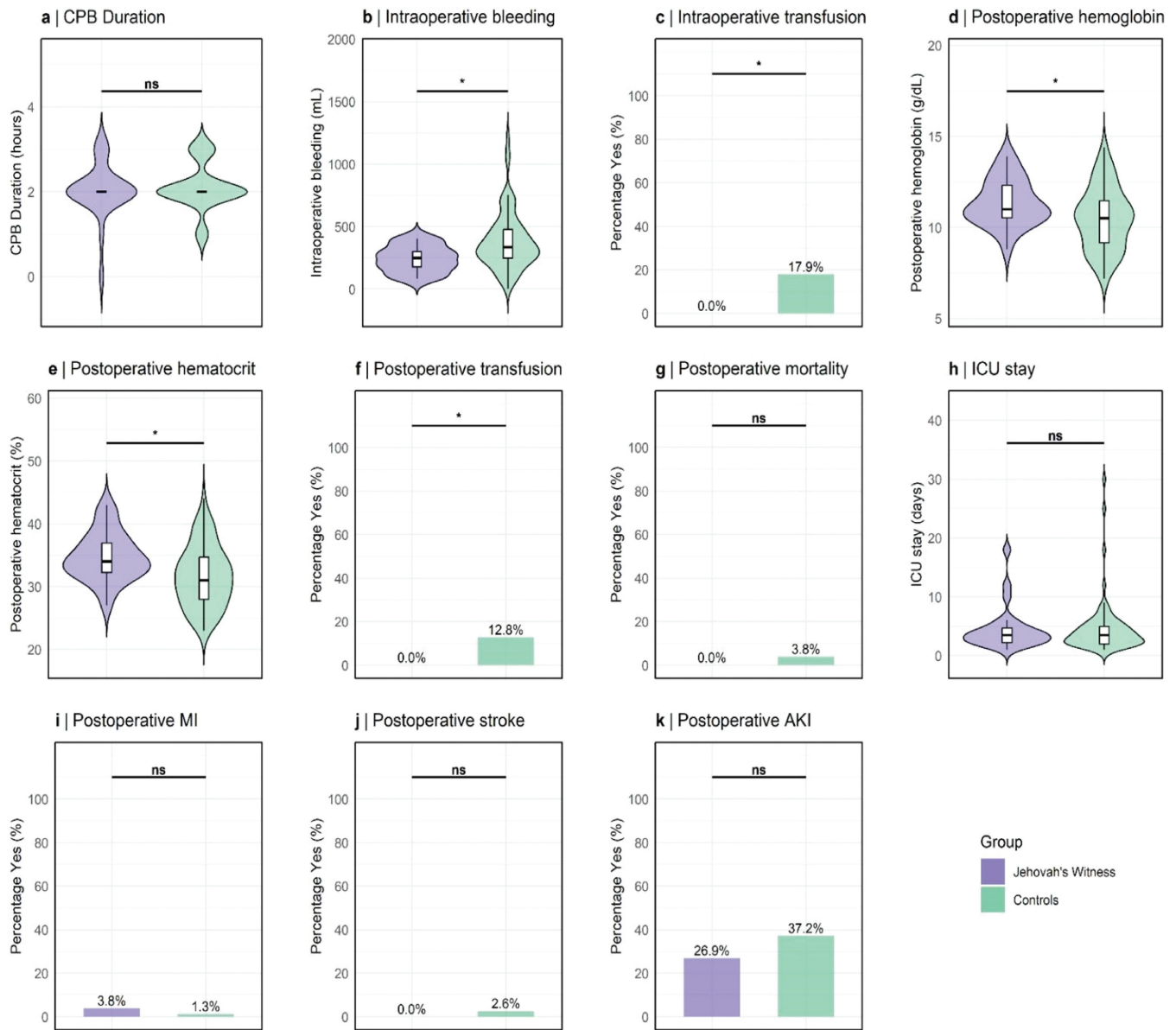


Figure 4 Comparison of intraoperative and postoperative variables between Jehovah's Witnesses and matched controls. Continuous variables (a, b, d, e, and h) are shown as violin plots, while categorical variables (c, f, g, i, j, and k) are shown as bar plots. Statistically significant group differences based on logistic regression models are marked with an asterisk (*), and non-significant differences are labeled as "ns". Jehovah's Witnesses are represented in purple, controls, in green.

depend on multiple patient- and procedure-related factors, including body size, surgical complexity, characteristics of extracorporeal circulation, and anticipated blood loss. Accordingly, Hb optimization in this cohort was individualized rather than protocol-driven, balancing these factors with clinical urgency and institutional constraints. Although Vaislic et al.²¹ proposed a preoperative Hb ≥ 14 g.dL⁻¹, applying this threshold to our cohort would have excluded 42% of JW patients from surgery, despite the absence of clear differences in short-term outcomes compared with those with higher baseline Hb. This observation may suggest that rigid Hb criteria may unnecessarily delay or preclude surgical treatment without clear evidence of benefit.

Consistent with the individualized approach to preoperative hemoglobin optimization, perioperative transfusion

decisions were not based on isolated numeric thresholds. Hemoglobin values were not used as absolute triggers, as acute blood loss with hemodynamic instability may precede measurable hematimetric decline. Likewise, hemodynamic criteria were not defined by rigid cutoffs for vasoactive drug dose, blood pressure, or cardiac output, given interindividual variability in physiologic tolerance. Point-of-care testing was routinely employed intraoperatively and postoperatively to assess hematologic and coagulation parameters when indicated; however, transfusion decisions ultimately relied on clinical judgment and multidisciplinary discussion, reflecting standard practice in cardiac surgery where timely intervention is essential.

This study has important limitations. Although PSM was applied, residual confounding cannot be excluded, particularly

given the small JW cohort and limited overlap of some perioperative exposures. Differences in access to intraoperative blood conservation strategies also represent a key limitation: all JW patients had cell-saver availability, whereas only a subset of non-JW controls did, potentially favoring outcomes in the JW group and limiting causal interpretation. Stroke diagnosis was based on clinically triggered neuroimaging rather than systematic screening, and peri-procedural myocardial infarction was defined according to contemporary CABG-specific criteria. These factors may have led to under-ascertainment of asymptomatic events. The single-center, single-surgeon design reduces procedural variability but limits generalizability to other settings. Finally, multiple secondary and exploratory outcomes were analyzed without adjustment for multiplicity; given the limited sample size and imprecision for rare events, these findings should be interpreted as hypothesis-generating. Larger multicenter studies are needed to provide more precise estimates of the impact of blood management strategies on infrequent but clinically relevant outcomes.

Taken together, the observed findings suggest that structured preoperative preparation, meticulous intraoperative blood conservation, and restrictive transfusion practices may contribute to the generally similar short-term outcomes observed between JW and non-JW patients undergoing CABG. Rather than implying superiority or causality, these results highlight the potential value of applying PBM principles more broadly in contemporary cardiac surgery. Future multicenter, prospective studies are warranted to confirm these observations, better account for residual confounding, and further define the role of blood management strategies across diverse clinical and healthcare settings.

Conclusion

In this retrospective observational study conducted in a public tertiary hospital, CABG in Jehovah's Witness patients was feasible and associated with acceptable short-term outcomes. Despite system constraints, outcomes were generally comparable to those observed in non-JW patients, acknowledging the imprecision related to the small sample size. Rather than implying superiority, these findings highlight the relevance of structured PBM strategies in contemporary cardiac surgery. Further multicenter, prospective studies are warranted to confirm these observations and better define the role of PBM in diverse public healthcare settings.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI, San Francisco, CA, USA) to assist with language

editing, improvement of clarity and organization of the manuscript, and refinement of responses to peer reviewers. After using this tool, the authors carefully reviewed and edited all content as needed and take full responsibility for the accuracy and integrity of the final manuscript.

Authors' contributions

Henrique Ryu Yamanaka Nakano: Conceptualization; methodology; data analysis; investigation; writing of the original draft and manuscript review & editing.

Guilherme Mota Filho: Data acquisition; investigation, and manuscript review & editing.

Bruno Camargo Calabria Araújo: data acquisition; investigation, and manuscript review & editing.

Heloísa Zogheib: Data acquisition; investigation, and manuscript review & editing.

Enzo Camargo Calabria Araújo: Data acquisition; investigation, and manuscript review & editing.

Leonardo Cardoso Saraiva: Data analysis and manuscript review & editing.

Jessica Silva Nicolau: Data acquisition; investigation, and manuscript review & editing.

Suely Pereira Zeferino: Data acquisition; investigation, and manuscript review & editing.

José Otávio Costa Auler Junior: Supervision; project administration, and critical revision of the manuscript.

Funding

The authors declare no funding support.

Conflicts of interest

The authors declare that they have no conflicts of interest and that this study received no external funding.

Acknowledgements

The authors thank the anesthesiology, cardiac surgery, and ICU teams at InCor-HCFMUSP for their support and dedication, as well as the clinical and administrative staff for their assistance during the study.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjane.2026.844765.

Editor

Liana Azi

References

1. Kimmaliardjuk DM, Toeg H, Glineur D, Sohmer B, Ruel M. Operative mortality with coronary artery bypass graft: where do we stand in 2015? *Curr Opin Cardiol*. 2015;30:611–8.
2. Lawson T, Ralph C. Perioperative Jehovah's Witnesses: a review. *Br J Anaesth*. 2015;115:676–87. Erratum in: *Br J Anaesth*. 2017;119:1071. Dosage error in article text.
3. Langstraat M, Megens-Bastiaanse CM, Rettig TCD, Scohy TV, Gerritse BM. Cardiac surgery in Jehovah's Witnesses: 329 consecutive cases. *J Cardiothorac Vasc Anesth*. 2023;37:1601–5.
4. Gebhardt K, Bartsch C. Tod nach Therapieverweigerung. Eine Kausalitätsprüfung am Beispiel eines Zeugen Jehovas [Death after refusing treatment – causality assessment using the example of a Jehovah's Witness]. *Arch Kriminol*. 2009;224:192–7. German.
5. Biskupski A, Waligórski S, Mokrzycki K, Brykczyński M. Postponed surgery of an acute aortic dissection (type A) in a Jehovah's Witness with significant hemostatic disorders. *Kardiochir Torakochirurgia Pol*. 2016;13:251–3.
6. Vitolo M, Mei DA, Cimato P, et al. Cardiac surgery in Jehovah's Witnesses patients and association with peri-operative outcomes: a systematic review and meta-analysis. *Curr Probl Cardiol*. 2023;48:101789.
7. Spence RK, Carson JA, Poses R, et al. Elective surgery without transfusion: influence of preoperative hemoglobin level and blood loss on mortality. *Am J Surg*. 1990;159:320–4.
8. Carson JL, Stanworth SJ, Roubinian N, et al. Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion. *Cochrane Database Syst Rev*. 2016;10:CD002042.
9. Mazer CD, Whitlock RP, Fergusson DA, et al. Restrictive or liberal red-cell transfusion for cardiac surgery. *N Engl J Med*. 2017;377:2133–44.
10. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
11. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39:33–8.
12. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38:2074–102.
13. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med*. 2006;25:4279–92.
14. Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci*. 2010;25:1–21.
15. Ho DE, Imai K, King G, Stuart EA. MatchIt: Nonparametric pre-processing for parametric causal inference. *J Statistical Software*. 2011;42:1–28.
16. Heinze G, Ploner M, Jiricka L, Steiner G. *logistf: Firth's Bias-Reduced Logistic Regression*. R package version 1.26.1. 2025. doi:10.32614/CRAN.package.logistf.
17. Chambault AL, Brown LJ, Mellor S, Harky A. Outcomes of cardiac surgery in Jehovah's Witness patients: a review. *Perfusion*. 2021;36:661–71.
18. Hebert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, et al. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. *N Engl J Med*. 1999;340:409–17.
19. Shander A, Javidroozi M, Perelman S, Puzio T, Lobel G. From bloodless surgery to patient blood management. *Mt Sinai J Med*. 2012;79:56–65.
20. Posluszny Jr JA, Napolitano LM. How do we treat life-threatening anemia in a Jehovah's Witness patient? *Transfusion*. 2014;54:3026–34.
21. Vaislic CD, Dalibon N, Ponzio O, et al. Outcomes in cardiac surgery in 500 consecutive Jehovah's Witness patients: 21-year experience. *J Cardiothorac Surg*. 2012;7:95.

ORIGINAL INVESTIGATION

Evaluation of two large language models for intensive care unit discharge decisions: a prospective observational cohort study



Engin İhsan Turan ^{a,*}, Abdurrahman Engin Baydemir ^b, Ebru Kaya ^a,
Zehra Polat Turan ^c, Ayça Sultan Şahin ^a

^a Istanbul Health Science University Kanuni Sultan Süleyman Education and Training Hospital, Department of Anesthesiology, Istanbul, Turkey

^b Basakşehir Cam ve Sakura City Hospital, Department of Anesthesiology, Istanbul, Turkey

^c Istanbul Health Science University Sisli Hamidiye Etfal Education and Training Hospital, Department of Anesthesiology, Istanbul, Turkey

Received 1 August 2025; accepted 16 May 2026

Available online 29 May 2026

KEYWORDS

Artificial intelligence;
Decision making;
Decision support systems, Clinical;
Intensive care units;
Natural language processing;
Patient discharge

Abstract

Background: The aim of this study was to evaluate the effectiveness of two general-purpose Large Language Models (LLMs), ChatGPT and Gemini, in predicting Intensive Care Unit (ICU) discharge decisions (discharge vs. non-discharge). By comparing their outputs with decisions made by ICU physicians, we sought to determine the alignment of AI-generated recommendations with expert clinical judgment and assess their potential as decision-support tools in critical care.

Methods: This prospective observational cohort study was conducted in a tertiary ICU between September 2024 and May 2025. Adult patients (≥ 18 years) requiring ICU discharge decisions were included. Standardized clinical prompts were generated from electronic health records and input into ChatGPT and Gemini. The models' binary discharge decisions were compared to those of ICU physicians. Model performance was assessed using accuracy, sensitivity, specificity, F1 score, Cohen's kappa, and McNemar's test. Discharge was defined as the positive class for all diagnostic performance analyses.

Results: A total of 398 patients were analyzed. ChatGPT demonstrated higher accuracy than Gemini (87.2% vs. 66.3%), with higher sensitivity (85.9% vs. 46.9%) and F1 score (0.890 vs. 0.628), whereas Gemini showed higher specificity (96.2% vs. 89.2%). Agreement with clinician decisions was substantial for ChatGPT ($\kappa = 0.737$, $p = 0.024$) and fair for Gemini ($\kappa = 0.379$, $p < 0.001$). Laboratory markers such as lactate, hemoglobin, and procalcitonin significantly differed between discharged and non-discharged patients.

* Corresponding author.

E-mail: enginihsan@hotmail.com (E.İ. Turan).

Conclusion: Large language models may support ICU discharge decisions when guided by structured, guideline-informed prompting. ChatGPT achieved higher overall accuracy, sensitivity, and F1 score, whereas Gemini demonstrated higher specificity.

Trial registration: Extention (Discharge) of ICU, NCT06584890, registered 03 September 2024, prospectively registered, <https://register.clinicaltrials.gov/prs/beta/studies/S000EVXZ00000029/recordSummary>.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

The Intensive Care Unit (ICU) plays a central role in the management of critically ill patients, yet the decision to discharge patients from the ICU remains one of the most complex and high-stake challenges in clinical practice.¹ Premature discharge may result in readmissions and increased mortality, whereas delayed discharge can lead to inefficient resource utilization, bed shortages, and unnecessary costs.^{1,2} Clinical guidelines such as those from the Society of Critical Care Medicine (SCCM)³ and expert recommendations emphasize the importance of both physiological stability and organizational factors relevant to ICU discharge decisions.⁴ However, the application of these criteria often varies depending on clinician experience and institutional practices, highlighting the need for standardized, objective decision-support mechanisms.

In recent years, Large Language Models (LLMs) like ChatGPT and Gemini have demonstrated promising capabilities in interpreting clinical data and providing recommendations across various domains of medicine.⁵⁻⁸ Despite their general-purpose design, Large Language Models (LLMs) may assist clinicians by synthesizing complex clinical data and providing structured guidance, particularly for decisions like ICU discharge. However, their accuracy and reliability in this context remain insufficiently studied, and real-world integration requires further evaluation. This study aimed to evaluate the effectiveness of two general-purpose AI models, ChatGPT and Gemini, in predicting ICU discharge status (discharge vs. non-discharge) by comparing their outputs with ICU physicians' decisions, and assessing whether structured, guideline-based prompts improve clinical alignment.

Methods

Study design

This was a prospective observational cohort study conducted to evaluate the effectiveness of general-purpose Artificial Intelligence (AI) models, specifically ChatGPT and Gemini, in predicting Intensive Care Unit (ICU) discharge decisions. The study was designed in accordance with the principles of the Declaration of Helsinki. Approval was obtained from the institutional ethics committee prior to data collection (KAEK-2024.08.175), and the study was registered on ClinicalTrials.gov (NCT06584890). Written informed consent was obtained from all patients or their legal representatives before enrollment.

The study was conducted in the Intensive Care Unit (ICU) of a tertiary care hospital. Data collection was performed

between September 15, 2024, and May 26, 2025. During this period, all eligible ICU admissions were prospectively screened and evaluated on a daily basis for inclusion in the study. The study was strictly non-interventional; AI-generated outputs were not disclosed to the clinical teams and had no impact on patient management or discharge decisions.

During the preparation of this manuscript, the authors used ChatGPT-4o only to improve language readability. The authors reviewed and edited the content as needed and take full responsibility for the scientific content of the manuscript.

Participants

All adult patients (aged ≥ 18 years) admitted to the ICU during the study period were screened for eligibility. Inclusion required the availability of sufficient clinical and demographic data in the Electronic Health Records (EHR). Two patient groups were defined: 1) Those discharged from the ICU, for whom data from the day of discharge were used; and 2) Those who remained in the ICU and who died during ICU stay, for whom data from the day prior to death were analyzed. This approach allowed the models to be exposed to both successful and unsuccessful ICU courses.

For patients considered appropriate for ICU discharge, all predefined clinical, physiological, and laboratory variables were systematically recorded from the day of discharge. Likewise, for patients who subsequently died in the ICU, corresponding data were collected in the same standardized manner from the day prior to death. Because data collection was integrated into daily ICU visits and extracted directly from the EHR system, all included patients had complete data for the predefined study variables. ICU readmission was defined as any return to the ICU within 24 hours of ICU discharge and was ascertained using electronic bed-transfer records. Patients were excluded if key clinical data were missing, if they were admitted for palliative or end-of-life care, or if informed consent was not obtained.

Variables

The primary outcome was ICU discharge status (discharge vs. non-discharge). Model performance was evaluated by comparing AI-generated discharge decisions with physician decisions. The non-discharge group included both patients who remained in the ICU and those who subsequently died, for whom clinical data from the day prior to death were analyzed. The study was designed to evaluate whether contemporary LLM-based systems, when implemented using clinically applicable prompting strategies, can meaningfully

align with physician ICU discharge decisions. For all diagnostic performance analyses, discharge was defined as the positive class. Accordingly, sensitivity, specificity, accuracy, and F1 score were calculated based on this definition.

Predictor variables included demographic data (age, gender, comorbidities), clinical indicators (diagnosis, reason for ICU admission), physiological variables (vital signs, Glasgow Coma Scale, ventilator use, inotropic support), laboratory findings (hemoglobin, lactate, creatinine, procalcitonin), and imaging results.

Subgroup analysis (clinically borderline patients)

A prespecified subgroup analysis was conducted to evaluate model performance in clinically borderline patients, representing cases in which ICU discharge decisions are more challenging and less clinically obvious. Borderline patients were defined as those who were not mechanically ventilated, not intubated, not receiving vasopressor or inotropic support, not on high-flow nasal oxygen therapy, and who were conscious and able to maintain spontaneous respiration.

To avoid inclusion of clinically obvious low-risk cases, patients who were breathing comfortably on room air with a respiratory rate < 20 lt; 20 breaths/min, had stable hemodynamics (mean arterial pressure $\geq$ 65 mmHg and systolic blood pressure < 180 mmHg), and showed no signs of respiratory distress or neurological impairment were excluded from the borderline subgroup.

Data sources and measurement

Data were extracted from the hospital's EHR system and anonymized prior to analysis. All clinical measurements, including laboratory results, vital signs, and physical examination findings, were obtained using standardized hospital protocols to ensure consistency. To evaluate ICU discharge decisions, both models received identical standardized case scenarios derived from the EHR. ChatGPT was accessed through a preconfigured custom GPT that incorporated guideline-based structured instructions and a fixed output framework, such that only the patient scenario was entered for each case. In contrast, Gemini was used in its default interface, and the same standardized prompt text was manually entered together with the patient scenario for each evaluation, without any model-level customization or enforced output structure. These prompts were designed to simulate real-time decision-making and included the following components (Supplementary Material 1, Supplementary Material 2, Supplementary Material 3):

- Demographics (age, sex, comorbidities)
- Reason for ICU admission
- Vital signs (heart rate, blood pressure, respiratory rate, SpO₂)
- Neurological status (Glasgow Coma Scale)
- Laboratory values (hemoglobin, creatinine, lactate, procalcitonin)
- Mechanical support (mechanical ventilation, inotropic use)
- Physical exam findings (general appearance, respiratory effort, consciousness, edema, signs of hemodynamic instability, abdominal findings).

GPT model

To evaluate ICU discharge decisions (discharge vs. non-discharge), a custom GPT-based model was developed using structured prompts derived from two key guidelines: the Society of Critical Care Medicine (SCCM) guidelines on ICU admission, discharge, and triage,³ and the Plotnikoff et al.⁹ framework on discharge planning and communication. The model was instructed to evaluate patient cases using a standardized five-section output format: 1) Clinical Stability (based on SCCM criteria such as vital signs, GCS, and absence of ICU-specific interventions), 2) Organizational Readiness (per Plotnikoff, focusing on care coordination and communication), 3) Risk Factors for readmission, 4) A binary discharge decision, and 5) A rationale explicitly citing guideline-based justifications. This structured approach enabled consistent, guideline-compliant assessments by the AI.

The GPT-based ICU discharge evaluator used in this study is accessible at the following link: <https://chatgpt.com/g/g-682c737d88048191a950d0b7710107af-icu-discharge-evaluator>.

Unlike ChatGPT, Gemini does not currently support custom model creation or structured output configuration. Therefore, Gemini 2.5 Flash was used in its default form, responding to standardized clinical prompts without any user-defined formatting or task-specific adaptation.

ChatGPT was accessed via a custom GPT interface based on the GPT-4 architecture, while Gemini 2.5 Flash was used through its default web interface. All model evaluations were conducted using the same institutional network environment. Each patient case was entered and evaluated in a separate session without prior conversational context to prevent memory carryover between cases. No iterative prompting, regeneration, or repeated runs were performed; only a single output per case was used for analysis. Default model parameters were used for both systems, and no manual adjustments to temperature or determinism settings were applied.

Data governance and privacy

All patient data used in this study were fully de-identified prior to analysis and before being entered into any LLM platform. No directly identifiable information (including names, national identification numbers, hospital numbers, exact dates, or contact details) was used. Only anonymized clinical variables, physiological parameters, and laboratory values were included in structured case templates.

Bias

To minimize potential bias, clinicians making discharge decisions were blinded to AI predictions. AI models were not updated during the study period to avoid data leakage. Data preprocessing and AI output interpretation were conducted by independent investigators not involved in patient care.

Power analysis

A priori sample size calculation was performed based on the results of a pilot study involving 50 patients, in which the proportion of discordant pairs between the AI model and physicians was 50%, and the odds ratio was estimated at

1.685. Using G*Power version 3.1.9.4 for a two-tailed McNemar test with a significance level (α) of 0.05 and a desired power ($1-\beta$) of 0.95, the required sample size was calculated as 398 patients.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics version 21 (IBM Corp., Armonk, NY). Categorical variables were presented as frequencies and percentages, while continuous variables were summarized as medians with interquartile ranges due to non-normal distributions. Normality of continuous variables was assessed using the Shapiro-Wilk test. Group comparisons for categorical variables were conducted using the Chi-Square test or Fisher's exact test, as appropriate. The Mann-Whitney *U*-test was used to compare continuous variables between groups.

Agreement between ChatGPT and intensive care physicians was assessed using Cohen's kappa coefficient. Classification performance metrics, including sensitivity, specificity, accuracy, and F1 score, were calculated based on confusion matrices. Differences between AI model predictions and physician decisions were evaluated using McNemar's test. To evaluate model performance under more clinically ambiguous conditions, a prespecified subgroup analysis was conducted including patients who were not mechanically ventilated and not receiving vasopressor or inotropic support at the time of evaluation. This subgroup was intended to represent more borderline ICU discharge scenarios. A two-sided *p*-value of less than 0.05 was considered statistically significant.

Results

A total of 398 adult patients admitted to the ICU were included in the study (Fig. 1). Of these, 147 patients (36.9%) were admitted for postoperative monitoring following major surgical procedures, including abdominal, orthopedic, and neurosurgical interventions. The remaining patients were admitted due to medical conditions such as sepsis, respiratory failure, or cardiac events. All eligible patients were prospectively screened during daily ICU rounds, and all included

patients had complete data for the predefined study variables required for the primary analyses. Notably, among all patients included in the study, no ICU readmissions occurred within 24 hours of ICU discharge, supporting the overall safety of the discharge decisions made by clinicians.

Table 1 presents the demographic and clinical characteristics of patients according to their discharge status. The median age was 68 (54–77.5) in the non-discharge group and 65 (52–75) in the discharge group ($p = 0.085$). Gender distribution was similar between groups, with 83 females and 74 males in the non-discharge group and 137 females and 104 males in the discharge group ($p = 0.435$). The presence of diabetes mellitus ($p = 0.491$), coronary artery disease ($p = 0.474$), heart failure ($p = 0.128$), stroke ($p = 0.444$), and COPD ($p = 0.790$) did not significantly differ between groups. However, the rates of mechanical ventilation and inotropic support were significantly higher in the non-discharge group ($p < 0.001$ for both).

Table 2 compares ChatGPT's and Gemini's predictions with the decisions made by intensive care doctors regarding patient discharge. ChatGPT predicted 140 out of 157 non-discharge cases and 207 out of 241 discharge cases correctly. There were 17 false positives and 34 false negatives. The agreement between ChatGPT and physicians was statistically significant ($p = 0.024$), with a Cohen's Kappa value of 0.737, indicating substantial agreement. Overall accuracy was 87.2%, sensitivity for predicting discharge was 85.9%, specificity for correctly identifying non-discharge cases was 89.2%, and the F1 score for discharge predictions was 89.0%.

Among the 398 cases, Gemini correctly identified 151 patients as requiring continued intensive care and 113 patients as appropriate for discharge. The overall accuracy of the model was 0.663. The sensitivity, representing the model's ability to correctly identify patients to be discharged, was 0.469, while the specificity, indicating correct identification of non-discharge cases, was 0.962. The F1 score, which reflects the balance between precision and recall, was calculated as 0.628. The agreement between Gemini and clinicians, as measured by Cohen's Kappa, was 0.379.

A prespecified subgroup analysis was performed in clinically borderline patients, defined as those who were not mechanically ventilated and not receiving vasopressor or

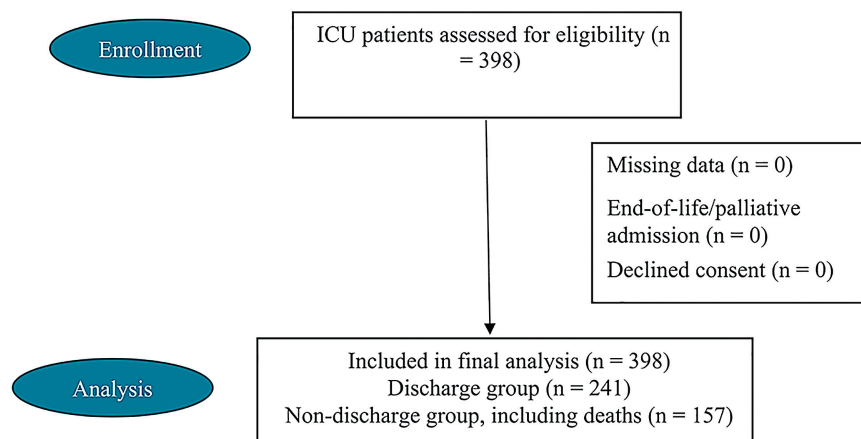


Figure 1 Flowchart of the patients.

Table 1 Demographic data of the patients.

	Non-Discharge (n = 157)	Discharge (n = 241)	p-value
Gender (female/male)	83/74	137/104	p = 0.435 ^a
Age	68 (54–77.5)	65 (52–75)	0.085
Diabetes Mellitus (no/yes)	99/58	159/82	p = 0.491 ^a
Coronary Artery Disease (no/yes)	129/28	191/50	p = 0.474 ^a
Heart Failure (no/yes)	129/28	213/27	p = 0.128 ^a
Stroke (no/yes)	140/17	215/24	p = 0.444 ^a
COPD (no/yes)	128/29	199/42	p = 0.790 ^a
Mechanical ventilation (no/yes)	61/96	241/0	p < 0.001 ^b
Inotropic Support (no/yes)	81/76	240/1	p < 0.001 ^b

^a Chi-Square test was applied.^b Fisher's Exact test was applied.

inotropic support at the time of evaluation. In this subgroup, ChatGPT demonstrated an accuracy of 77.6%, with a sensitivity of 82.4%, specificity of 56.5%, F1-score of 0.857, and fair agreement with clinicians ($\kappa = 0.343$). Gemini showed substantially lower overall performance, with an accuracy of 43.2%, sensitivity of 35.3%, specificity of 78.3%, F1-score of 0.504, and slight agreement with clinicians ($\kappa = 0.067$) (Table 3).

Table 4 presents the comparison of laboratory findings between patients in the non-discharge group and those who were discharged. Median hemoglobin levels were lower in the non-discharge group (8.6 [7.8–10.5]) compared to the discharge group (10.1 [8.9–12.1]) (p < 0.001). Similarly, lactate (3.3 [2.0–10.1] vs. 1.3 [1.0–1.7]), creatinine (1.43 [0.73–2.26] vs. 0.84 [0.64–1.09]), heart rate (90 [83–105.5] vs. 84 [79.5–94]), and procalcitonin levels (1.91 [0.23–6.97] vs. 0.13 [0.06–0.33]) were significantly higher in the non-discharge group, all with p < 0.001. Conversely, SpO₂ values were slightly lower in the non-discharge group (96 [95–98]) compared to the discharge group (97 [96–98]) (p = 0.048).

Discussion

In this prospective cohort, both LLM-based systems demonstrated measurable agreement with clinicians in ICU discharge decisions, with the prompt-guided configuration showing higher concordance than the non-customized configuration. However, performance differences were most evident across clinically stable versus unstable presentations, suggesting that structured input and standardized clinical prompting may help LLMs integrate routine ICU

variables more consistently. A plausible explanation is that ICU discharge decisions are strongly driven by recognizable patterns of physiological stability and laboratory abnormalities; thus, models may differ in how reliably they extract, prioritize, and weigh these signals when structured prompts and output constraints are applied.

ICU discharge decisions frequently rely on a combination of objective clinical parameters and subjective clinical judgment. Previous work has shown substantial inter-clinician variability in ICU discharge practices, largely attributable to the absence of universally standardized discharge criteria.¹⁰ Against this background, the relatively weaker alignment observed with the non-customized configuration may reflect not only model limitations but also the inherent inconsistency of the clinical reference standard itself. However, recent efforts to formalize ICU discharge processes are increasingly emphasizing structured, evidence-based frameworks. In this context, the introduction of guideline-informed, standardized prompting in our study was associated with improved alignment between AI outputs and physician decisions, supporting the concept that structured clinical representations may help LLM-based systems capture the implicit heuristics clinicians apply when synthesizing ICU data.¹¹ Importantly, during the pilot phase of this study, the same model operating without structured prompts demonstrated substantially lower agreement, suggesting that domain-specific adaptation and clinically grounded prompt design may be at least as critical as the underlying model architecture for effective AI deployment in critical care.

Ethical considerations are central to ICU discharge decision-making, as these decisions not only determine resource utilization but also directly affect patient safety and

Table 2 Comparison of ChatGPT's and Gemini's predictions with Intensive care doctor decisions.

		ChatGPT		p ^a	Gemini		p ^a
		Non-Discharge	Discharge		Non-Discharge	Discharge	
Intensive care doctors	Non-Discharge	140	17	0.024	151	6	< 0.001
	Discharge	34	207		128	113	

For both models, discharge was treated as the positive class. ChatGPT: accuracy 87.2% (95% CI 83.5–90.2), sensitivity 85.9% (95% CI 80.8–89.9), specificity 89.2% (95% CI 83.3–93.5), F1 score 0.890, $\kappa = 0.737$ (95% CI 0.669–0.805), McNemar p = 0.024. Gemini: accuracy 66.3% (95% CI 61.5–70.8), sensitivity 46.9% (95% CI 40.7–53.2), specificity 96.2% (95% CI 91.9–98.4), F1 score 0.628, $\kappa = 0.379$ (95% CI 0.310–0.452), McNemar p < 0.001.

^a McNemar test was applied.

Table 3 Subgroup analysis of borderline patients.

	Accuracy	Sensitivity	Specificity	F1 score	Kappa	95% CI
ChatGPT	77.6%	82.4%	56.5%	0.857	0.343	0.208 to 0.482
Gemini	43.2%	35.3%	78.3%	0.504	0.067	-0.006 to 0.140

Table 4 Comparison of laboratory findings of the patients.

	Non-Discharge (n = 157) Median (Q1–Q3)	Discharge (n = 241) Median (Q1–Q3)	p-value ^a
Hemoglobin	8.6 (7.8–10.5)	10.1 (8.9–12.1)	< 0.001
Lactate	3.3 (2.0–10.1)	1.3 (1.0–1.7)	< 0.001
Creatinine	1.43 (0.73–2.26)	0.84 (0.64–1.09)	< 0.001
Heart Rate	90 (83–105.5)	84 (79.5–94)	< 0.001
SpO ₂	96 (95–98)	97 (96–98)	0.048
Procalcitonin	1.91 (0.23–6.97)	0.13 (0.06–0.33)	< 0.001

^a Mann-Whitney *U*-test was applied.

autonomy. The absence of transparent and standardized discharge policies has been shown to generate ethical tension, inter-clinician conflict, and uncertainty in practice.¹² From this perspective, structured AI-supported reasoning may offer value not by replacing clinician judgment, but by promoting greater transparency, reproducibility, and ethical consistency in how discharge decisions are formulated.

However, ICU discharge is not determined solely by patient physiology. System-level factors such as protocol gaps, communication failures, and workflow inefficiencies also play a substantial role in discharge decisions and patient safety.¹³ This highlights that AI-based decision support should not be viewed as a standalone solution, but rather as a component of broader, system-oriented strategies aimed at standardizing transitions of care. In this context, our findings align with prior work demonstrating the feasibility of machine learning-based discharge support. For example, Wu et al. reported high agreement between a LightGBM model and physician decisions, with the added benefit of earlier identification of potential discharge candidates.¹⁴ While their approach relied on structured tabular modeling, our results suggest that LLM-based systems, when clinically guided through structured prompting, may achieve comparable alignment while operating in a more flexible, narrative-compatible format. Previous studies have demonstrated that machine-learning models can support post-ICU risk stratification and discharge-related decision-making.¹⁵ Together, these observations support the potential role of LLMs as integrative clinical support tools within ethically grounded and system-aware ICU discharge frameworks.

Recent evidence indicates that ICU discharge decisions are influenced by factors extending beyond snapshot physiological measurements. The integration of narrative clinical documentation through natural language processing has been shown to improve discharge modeling and, importantly, to uncover patient subgroups that are poorly captured by conventional structured predictors, including those with psychosocial or surgical complexity.¹⁶ This suggests that discharge decisions are shaped not only by measurable physiology but also by contextual, organizational, and

psychosocial dimensions that are often embedded in daily clinical narratives.

Previous studies have suggested that admission-day variables alone may allow effective early risk stratification for ICU readmission and post-ICU deterioration.¹⁷ These findings support the concept that clinically meaningful prognostic signals are often present well before discharge and may complement or, in some contexts, outperform discharge-day assessments.

Error pattern analysis revealed distinct behavioral tendencies between the two LLM-based systems. The prompt-guided configuration occasionally favored discharge in cases involving recent seizures or patients receiving high-flow nasal cannula therapy at ≥ 40 L.min⁻¹, suggesting a greater tolerance for borderline physiological findings. In contrast, the non-customized configuration demonstrated a predominantly conservative pattern, frequently recommending continued ICU care even in clinically low-risk scenarios, which was reflected in its high specificity but limited sensitivity.

These divergent error profiles indicate that model outputs are shaped not only by the underlying architecture but also by how clinical information is framed and constrained. The conservative bias observed in the non-customized configuration suggests prioritization of isolated physiological abnormalities, whereas the structured prompting strategy appeared to promote a more integrative assessment, yielding a balance between sensitivity and specificity that more closely resembled clinician decision-making. This observation reinforces the concept that prompt design is not merely a technical detail but a central determinant of how LLM-based systems operationalize clinical reasoning.

Importantly, the association between ICU non-discharge and markers such as anemia, hyperlactatemia, renal dysfunction, elevated procalcitonin, mechanical ventilation, and inotropic support holds the biological plausibility of the decision patterns observed. However, while the absence of these features in discharged patients validates their inclusion as model inputs, it also underscores the limitations of single-time-point assessments. The integration of longitudinal parameter trends may allow future AI-

supported frameworks to better capture evolving risk profiles and improve performance in more ambiguous discharge scenarios.

Limitations

This study has several limitations. First, it was conducted in a single tertiary care center, which may limit the diversity of patient populations and care protocols represented. Second, although the AI systems were evaluated using real-time clinical data, they were not embedded within the clinical workflow, and their outputs did not influence actual discharge decisions, thereby limiting assessment of real-world clinical impact. Third, while both systems were tested using structured clinical prompts, the reliance on manually curated input data may not fully reflect the variability, incompleteness, and noise inherent to routine ICU documentation. In addition, model performance may have been influenced by the specific phrasing and formatting of prompts, which were standardized but not externally validated. Moreover, long-term patient-centered outcomes, including post-discharge mortality and readmission beyond the index hospitalization, were not assessed, restricting conclusions regarding downstream safety.

A major methodological limitation is the non-equivalence of AI configurations. ChatGPT was evaluated using a structured, guideline-based custom configuration, whereas Gemini was tested in its default, non-customized form. This design choice reflected the primary aim of determining whether LLM-based systems can meaningfully perform this task under clinically usable conditions rather than comparing intrinsic model capabilities. Consequently, observed differences cannot be attributed to inherent model superiority and may primarily reflect the impact of structured prompting and configuration. A methodologically fair model-to-model comparison would require applying equivalent structured prompts to both systems or evaluating both under default conditions, which should be addressed in future work.

In addition, clinical data were derived from the day of ICU discharge or the day prior to death. Although this reflects real-world decision timing, this approach may have included a substantial proportion of clinically straightforward cases, thereby facilitating discrimination and potentially inflating apparent performance. Importantly, not all patients who subsequently died were uniformly unstable at the time of evaluation, and clinical deterioration occurred after the assessed day in a subset of cases. Future investigations should incorporate longitudinal designs capturing multi-day clinical trajectories, dynamic laboratory trends, and validated severity indices to more rigorously evaluate AI performance under evolving and uncertain discharge conditions.

Finally, although LLMs offer powerful capabilities for clinical reasoning, they remain susceptible to hallucinations, plausible-sounding but incorrect outputs, particularly in the absence of rigorously constrained input structures. While structured prompting likely mitigated this risk, the potential for inaccurate or overconfident responses represents a critical limitation to real-world deployment in safety-critical environments such as intensive care.

Generalizability

The findings of this study should be interpreted with caution in broader contexts. While the use of structured prompts and standardized criteria provides a foundation for reproducibility, the results may not directly generalize to ICUs with different patient populations, clinical practices, or electronic health record systems. Moreover, the performance of ChatGPT and Gemini may vary across institutions with differing levels of staffing, documentation quality, and discharge policies. Further validation studies in diverse healthcare settings are needed to evaluate whether similar levels of agreement and accuracy can be maintained when these models are integrated into real-time, multidisciplinary ICU decision-making.

Conclusion

In conclusion, this study supports the feasibility of using general-purpose large language models as decision-support tools for ICU discharge status (discharge vs non-discharge) when implemented with structured, guideline-informed prompting. The prompt-guided configuration demonstrated higher concordance with physician decisions than a non-customized configuration, suggesting that standardization of input structure and task framing may be a key determinant of clinical alignment. Nevertheless, these findings should be interpreted cautiously given the single-center design, the use of a single-time-point assessment, and the non-equivalence of model configurations. Future multicenter studies using standardized, reproducible prompting strategies and longitudinal clinical trajectories are needed to evaluate real-world safety, generalizability, and workflow integration.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request, after de-identification and subject to institutional and ethical approval.

Compliance with author instructions

We confirm that this manuscript complies with all instructions to authors as outlined by *Brazilian Journal of Anesthesiology*. The study is submitted as an original research article and adheres to the structural and ethical standards required by the journal.

Authorship compliance

All listed authors meet the authorship criteria defined by the International Committee of Medical Journal Editors (ICMJE). Each author has made substantial contributions to the work,

approved the submitted version, and agreed to be accountable for all aspects of the manuscript.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Health Science University Kanuni Sultan Süleyman Training and Research Hospital (Approval Number: KAEK-2024.08.175). Written informed consent was obtained from all participants or their legal surrogates prior to enrollment. The study complies with the principles of the Declaration of Helsinki.

Use of reporting checklist

This manuscript adheres to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for observational studies. A completed STROBE checklist is included as a Supplementary File.

Originality statement

This manuscript has not been published previously and is not under consideration for publication elsewhere. All content is original, and appropriate credit is given for all referenced works. No text or images generated by AI were used in the writing or preparation of the manuscript. ChatGPT 4o was used only to assist in improving the readability of the language.

Authors' contributions

Engin İhsan Turan: Study conception and design; AI model development; data interpretation; manuscript drafting, and revision.

Abdurrahman Engin Baydemir: Statistical supervision; methodological refinement, and critical revision.

Ebru Kaya: Patient data collection; clinical interpretation, and coordination of AI model application.

Zehra Polat Turan: Study conception and design; data quality control, and writing.

Ayça Sultan Şahin: Oversight of ICU protocols; ethics documentation, and manuscript revision for intellectual content.

All authors read and approved the final manuscript.

Funding

The authors did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors for this research.

The authors declare that they have not received any financial support for the conduct of this research or the preparation of this manuscript. Furthermore, there are no conflicts of interest to disclose by any of the authors. This study has not been presented in any congress.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

In this study, ChatGPT 4o has been utilized to enhance the readability of the paper, ensuring that the findings and discussions are accessible and comprehensible to a wider audience.

Conflicts of interest

The authors declare no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjane.2026.844768.

Associate Editor

Célio Gomes de Amorim

References

1. Forster GM, Bihari S, Tiruvoipati R, Bailey M, Pilcher D. The Association between Discharge Delay from Intensive Care and Patient Outcomes. *Am J Respir Crit Care Med*. 2020;202:1399–406.
2. Vollam S, Gustafson O, Morgan L, Pattison N, Thomas H, Watkinson P. Patient Harm and Institutional Avoidability of Out-of-Hours Discharge From Intensive Care: An Analysis Using Mixed Methods*. *Crit Care Med*. 2022;50:1083–92.
3. Nates JL, Nunnally M, Kleinpell R, et al. ICU Admission, Discharge, and Triage Guidelines: A Framework to Enhance Clinical Operations, Development of Institutional Policies, and Further Research. *Crit Care Med*. 2016;44:1553–602.
4. Ruppert MM, Loftus TJ, Small C, et al. Predictive Modeling for Readmission to Intensive Care: A Systematic Review. *Crit Care Explor*. 2023;5:e0848.
5. Turan E, Baydemir AE, Özcan FG, Şahin AS. Evaluating the accuracy of ChatGPT-4 in predicting ASA scores: A prospective multicentric study ChatGPT-4 in ASA score prediction. *J Clin Anesth*. 2024;96:111475.
6. Turan E, Baydemir AE, Balıztatlı AB, Şahin AS. Assessing the accuracy of ChatGPT in interpreting blood gas analysis results ChatGPT-4 in blood gas analysis. *J Clin Anesth*. 2025;102:111787.
7. Turan EI, Baydemir AE, Şahin AS, Özcan FG. Effectiveness of ChatGPT-4 in predicting the human decision to send patients to the postoperative intensive care unit: a prospective multicentric study. *Minerva Anesthesiol*. 2025;91:259–67.
8. Dost B, Turan E, Aydın ME, et al. Artificial Intelligence in Anaesthesiology: Current Applications, Challenges, and Future Directions. *Türk J Anaesthesiol Reanim*. 2025;53:282–92.
9. Plotnikoff KM, Krewulak KD, Hernández L, et al. Patient discharge from intensive care: an updated scoping review to identify tools and practices to inform high-quality care. *Crit Care*. 2021;25:438.
10. Hiller M, Burisch C, Wittmann M, Bracht H, Kaltwasser A, Bakker J. The current state of intensive care unit discharge practices -

Results of an international survey study. *Front Med (Lausanne)*. 2024;11:1377902.

11. Hiller M, Wittmann M, Bracht H, Bakker J. Delphi study to derive expert consensus on a set of criteria to evaluate discharge readiness for adult ICU patients to be discharged to a general ward – European perspective. *BMC Health Serv Res*. 2022;22:773.
12. You SB, Ulrich CM. Ethical considerations in evaluating discharge readiness from the intensive care unit. *Nurs Ethics*. 2024;31:896–906.
13. van Sluisveld N, Oerlemans A, Westert G, van der Hoeven JG, Wollersheim H, Zegers M. Barriers and facilitators to improve safety and efficiency of the ICU discharge process: a mixed methods study. *BMC Health Serv Res*. 2017;17:251.
14. Wu CP, Shirley RB, Milinovich A, et al. Exploring timely and safe discharge from ICU: a comparative study of machine learning predictions and clinical practices. *Intensive Care Med Exp*. 2025;13:10.
15. Thorat PJ, Fornasa M, de Bruin DP, et al. Explainable Machine Learning on AmsterdamUMCdb for ICU Discharge Decision Support: Uniting Intensivists and Data Scientists. *Critical Care Explorations*. 2021;3:e0529.
16. Temple MW, Lehmann CU, Fabbri D. Natural Language Processing for Cohort Discovery in a Discharge Prediction Model for the Neonatal ICU. *Appl Clin Inform*. 2016;7:101–15.
17. Loreto M, Lisboa T, Moreira VP. Early prediction of ICU readmissions using classification algorithms. *Comput Biol Med*. 2020;118:103636.

ORIGINAL INVESTIGATION

Prevalence and factors associated with burnout syndrome among anesthesiology residents in Rio Grande do Sul, Brazil: a multicenter cross-sectional study



Airton Bagatini ^{a,*}, Felipe de Sousa Junqueira ^a, Eduarda Druck Magadan ^a,
Sílvia Ramos Hecktheuer ^b, Regis Borges Aquino ^c, Florentino Fernandes Mendes ^d,
Eduardo Francisco Mafassoli Corrêa ^e, Cleiton da Silva Pando ^f,
Daniela Benzano Bumaguin ^g, Cassiana Gil Prates ^a

^a Sociedade de Anestesiologia (SANE), Centro de Ensino e Treinamento, Porto Alegre, RS, Brazil

^b Associação Hospitalar Beneficente São Vicente de Paulo, Centro de Ensino e Treinamento, Passo Fundo, RS, Brazil

^c Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS), Hospital São Lucas, Centro de Ensino e Treinamento, Porto Alegre, RS, Brazil

^d Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Irmandade Santa Casa de Misericórdia de Porto Alegre, Centro de Ensino e Treinamento, Porto Alegre, RS, Brazil

^e Hospital Universitário, Universidade de Santa Maria (UFSM), Centro de Ensino e Treinamento Professor Manoel Alvarez, Santa Maria, RS, Brazil

^f Hospital de Clínicas de Porto Alegre (HCPA), Serviço de Anestesia e Medicina Perioperatória (SAMPE), Centro de Ensino e Treinamento, Porto Alegre, RS, Brazil

^g Universidade Federal do Rio Grande do Sul (UFRGS), Centro de Pesquisa em Álcool e Drogas, Porto Alegre, RS, Brazil

Received 9 January 2026; accepted 20 May 2026

Available online 1 June 2026

KEYWORDS

Anesthesiology;
Burnout;
Professional;
Education;
Medical;
Graduate;
Internship and
Residency;
Mental Health;
Occupational Stress

Abstract

Background: Burnout syndrome is characterized by emotional exhaustion, depersonalization, and low personal accomplishment. Anesthesiology residents are particularly vulnerable due to prolonged working hours, cognitive overload, and exposure to stressful clinical situations. This study aimed to estimate the prevalence of burnout and explore associated occupational and psychosocial factors among anesthesiology residents in Rio Grande do Sul, Brazil.

Methods: This multicenter cross-sectional study included residents from 6 training centers accredited by the Brazilian Society of Anesthesiology. Data were collected using an anonymous online questionnaire and the Portuguese-validated Maslach Burnout Inventory-Human Services Survey (MBI-HSS). Burnout was defined by high emotional exhaustion (≥ 25), high depersonalization (≥ 10), and low personal accomplishment (≤ 32). Associations were analyzed using Fisher's exact test ($p < 0.05$).

* Corresponding author.

E-mail: бага2010@gmail.com (A. Bagatini).

Results: Ninety-two residents participated (response rate, 84.4%). The overall prevalence of burnout was 26.1% (95% CI 17.5%–36.3%). High emotional exhaustion occurred in 72.8%, depersonalization in 42.4%, and low personal accomplishment in 43.5%. Burnout was significantly associated with living alone ($p = 0.036$) and psychotropic medication use ($p = 0.017$). A weekly workload of 60 hours or more ($p = 0.068$) was not significantly associated with burnout. Female residents showed higher emotional exhaustion ($p = 0.039$).

Conclusion: Burnout prevalence among anesthesiology residents in Rio Grande do Sul was high, predominantly characterized by emotional exhaustion. Institutional strategies focusing on psychosocial support, workload management, and mental health promotion may improve residents' well-being and professional performance.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Burnout syndrome was first described by Freudenberger in 1974 as a state of work-related physical and Emotional Exhaustion (EE) among healthcare professionals, characterized by a loss of energy and motivation as well as reduced effectiveness at work.¹ Subsequently, Maslach and Jackson conceptualized burnout as a multidimensional construct comprising EE, Depersonalization (DP), and reduced Personal Accomplishment (PA), and developed the Maslach Burnout Inventory (MBI), which has since become the most widely used instrument for assessing this phenomenon.^{2,3} In 2019, the World Health Organization included burnout in the 11th Revision of the International Classification of Diseases as an occupational phenomenon, recognizing it as the result of chronic, unmanaged workplace stress.⁴

Burnout prevalence estimates vary substantially depending on the definition and cutoff criteria used, particularly regarding whether one, two, or all three MBI dimensions are required for diagnosis. Studies adopting less stringent definitions, such as high scores in a single dimension, tend to report higher prevalence rates, whereas more conservative tridimensional definitions provide greater specificity but lower estimates. In the present study, we adopted a conservative tridimensional definition (high emotional exhaustion, high depersonalization, and low personal accomplishment) to improve the specificity and comparability of our findings.

Among healthcare professionals, burnout has emerged as a global threat to quality of care and patient safety. Evidence from systematic reviews indicates prevalence rates exceeding 50% in certain medical specialties, with consistent associations with medical errors, reduced productivity, and adverse outcomes.^{5,6} Its impact extends beyond the individual, adversely affecting healthcare teams, institutions, and the sustainability of health systems.⁷

Resident physicians represent a particularly vulnerable group. Multicenter studies report prevalence rates ranging from 30% to 60%, depending on specialty, year of training, and assessment methodology.^{8,9} Frequently cited risk factors include prolonged working hours, sleep deprivation, cognitive overload, repeated exposure to human suffering, institutional pressures, and insufficient social support.^{9,10} Consequences include increased risk of depression, suicidal ideation, attrition from the profession, and impaired clinical performance.^{11,12}

In anesthesiology, these risks are further intensified by the need for continuous vigilance, rapid decision-making in

critical situations, and management of highly complex clinical scenarios. Such characteristics heighten vulnerability to EE, placing anesthesiology residents at particularly high risk.^{9,11} Studies conducted in different countries consistently demonstrate high rates of burnout in this population, often exceeding those observed in other medical specialties.^{12,13}

In Brazil, although burnout has been investigated in some medical specialties, data on anesthesiology residents remain limited. A national multicenter study among intensivists reported high prevalence rates,¹⁴ and research involving multiprofessional residents also revealed concerning levels of burnout.¹⁵ However, specific data on anesthesiology residents in Rio Grande do Sul, the southernmost state of Brazil – an area with a strong tradition in medical education and a substantial number of training centers accredited by the Brazilian Society of Anesthesiology (SBA) – are lacking. This gap hinders the development of institutional strategies tailored to the local context and the implementation of effective prevention and support policies.

We hypothesized that anesthesiology residents training in SBA-accredited centers in Rio Grande do Sul, Brazil, have a high prevalence of burnout syndrome and that its occurrence is associated with occupational and psychosocial factors, including workload, sleep deprivation, and perceived institutional support. Therefore, this study aimed to estimate the prevalence of burnout and explore associated occupational and psychosocial factors among anesthesiology residents training in SBA-accredited centers in Rio Grande do Sul, Brazil.

Methods

Study design

This multicenter cross-sectional study included all 6 SBA-accredited training centers in the state of Rio Grande do Sul, Brazil. The project was approved by the Institutional Review Board, and research and methods adhered to the provisions of the Declaration of Helsinki and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

Ethical aspects

The study protocol was approved by the Research Ethics Committee of Hospital Ernesto Dornelles, Porto Alegre, RS, Brazil.

(approval number: 7.337.285; CAAE 85752324.2.0000.5304), in accordance with Resolution Number 466/2012 of the Brazilian National Health Council. Data collection was conducted anonymously using an electronic questionnaire (Google Forms). Access to the survey questions was preceded by mandatory reading and electronic acceptance of the Informed Consent Form (ICF). Acceptance of the ICF (by clicking "I agree") enabled the resident to proceed with the questionnaire, constituting formal informed consent prior to participation. Data confidentiality and participant anonymity were strictly preserved throughout all study phases.

Study population and eligibility

The target population comprised 109 anesthesiology residents eligible across the 6 SBA-accredited training centers in Rio Grande do Sul. Participant recruitment occurred between July and August 2025, and all residents were invited via electronic message containing a link to the online questionnaire.

Inclusion criteria were enrollment in the first, second, or third year of anesthesiology residency at an SBA-accredited center in Rio Grande do Sul and electronic acceptance of the ICF, which was required to access and complete the questionnaire. Exclusion criteria included residents who were on leave at the time of data collection (medical leave, maternity leave, vacation, or other reasons), as well as those who declined to provide electronic informed consent despite invitation and clarification of the study protocol.

Electronic questionnaires with incomplete data, defined as missing responses in more than 10% of the MBI-HSS items or in key sociodemographic variables, would be excluded to preserve the validity of burnout assessment. A total of 109 residents were invited to participate, of whom 92 responded. No participants met exclusion criteria, and all respondents were included in the final analysis.

Data collection and instruments

Data were collected using an anonymous, self-administered, and self-explanatory electronic questionnaire developed on the Google Forms platform. The instrument was divided into 2 main sections: 1) Sociodemographic and occupational variables, and 2) Assessment of burnout syndrome.

The first section consisted of a structured questionnaire designed to capture potential risk factors for burnout. Variables included resident characteristics such as sex, age group, marital status, living arrangement (living alone), year of residency, weekly workload, engagement in external professional activities, motivation for choosing anesthesiology, lifestyle habits (physical activity and alcohol consumption), and mental health-related factors (use of psychotropic medications and engagement in psychotherapy).

The second section assessed burnout using the MBI-Human Services Survey (MBI-HSS), in its Portuguese-language validated version.¹⁶ The MBI-HSS is considered the gold standard for burnout assessment and consists of 22 items distributed across 3 core dimensions: EE (9 items), DP (5 items), and PA (8 items). Each item is rated on a 7-point Likert scale ranging from 0 ("never") to 6 ("every day"), reflecting the frequency of work-related feelings. Each MBI-HSS dimension (emotional

exhaustion, depersonalization, and personal accomplishment) was analyzed separately, with scores calculated according to standard procedures for each subscale.

Items related to EE (items 1, 2, 3, 6, 8, 13, 14, 16, and 20) and DP (items 5, 10, 11, 15, and 22) are negatively oriented, with higher scores indicating greater burnout severity. In contrast, items assessing PA (items 4, 7, 9, 12, 17, 18, 19, and 21) are positively oriented, with higher scores reflecting lower burnout levels. Burnout dimensions were analyzed separately according to established cutoff points. Overall burnout syndrome was defined by the simultaneous presence of high EE, high DP, and low PA, in accordance with validated tridimensional criteria.

Definition of burnout syndrome

Burnout syndrome classification and prevalence were determined using tridimensional cutoff points established in national and international validation studies and systematic reviews based on MBI-HSS scores. Subscale scores were categorized as high (presence) or low (absence) burnout according to the following thresholds: high EE (≥ 25), high DP (≥ 10), and low PA (≤ 32). The simultaneous presence of high EE, high DP, and low PA was used to define overall burnout syndrome.^{5,11}

These cutoff points are based on validation studies of the Portuguese version of the MBI-HSS and have been widely adopted in Brazilian studies involving healthcare professionals. Their use enhances the cultural and linguistic validity of the assessment in this population and allows for meaningful comparison with national and international literature.

Sample size calculation

Sample size was calculated using WINPEPI, version 11.65. The calculation was based on the total number of anesthesiology residents accredited in SBA training centers as of June 30, 2025 ($n = 109$), assuming an expected burnout prevalence of 51% (based on prior literature),⁸ a 95% confidence level, and a margin of error of 10 percentage points. Accounting for an estimated 10% rate of losses and refusals, the minimum required sample size was 59 respondents to ensure the minimum sample size required to achieve the desired precision for prevalence estimation.

Statistical analysis

Participant responses, initially recorded in Google Sheets, were compiled and exported to IBM SPSS, version 20.0, for statistical analysis. Burnout prevalence was reported with corresponding 95% Confidence Intervals (95% CI). Categorical variables were described using absolute (n) and relative (%) frequencies. Associations between burnout syndrome (and its dimensions) and sociodemographic and occupational variables were assessed using Fisher's exact test. A two-tailed $p < 0.05$ was considered statistically significant for all analyses. Participants were nested within six training centers. Although center-level characteristics may influence both exposures and burnout, the analyses were conducted assuming independence between observations, given the similar structure and training conditions across centers.

Given the exploratory nature of the study, no formal adjustment for multiple comparisons was performed. The analyses of factors associated with burnout were secondary and exploratory. Considering the sample size and the limited number of outcome events, multivariable modeling was not performed due to the risk of overfitting and unstable estimates. In addition, considering the rule-of-thumb of at least 10 events per variable, the number of burnout cases would not support a stable multivariable model including multiple covariates. Therefore, only bivariate analyses were conducted, and the findings should be interpreted as exploratory associations rather than independent effects.

Results

A total of 92 anesthesiology residents from the 6 SBA-accredited training centers in Rio Grande do Sul participated in the study, corresponding to a response rate of 84.4% (92/109). The sample consisted predominantly of men (56.5%), with most participants aged 25 to 30 years (54.3%) and a median age of 29 years. The majority were in a stable relationship or dating (66.3%), and nearly half lived alone (47.8%). Residency year distribution was balanced: 31.5% were first-year, 32.6% second year, and 35.9% third-year residents.

Weekly workload was high, with 65.2% of residents reporting ≥ 60 working hours per week and 33.7% reporting 40–60 hours. Nearly half (48.9%) reported engaging in professional activities outside the residency program, and 45.7% stated that anesthesiology was their first-choice specialty. Regarding lifestyle factors, 52.2% had a pet, 64.2% engaged in regular physical activity (up to 4 times per week), and 41.3% reported psychotropic medication use. Approximately one-third (31.5%) were undergoing psychotherapy, and 48.9% reported alcohol consumption once or twice per week (Table 1).

The overall prevalence of burnout was 26.1% (95% CI 17.5%–36.3%). When analyzed separately, the MBI-HSS dimensions showed high EE in 72.8% (95% CI 62.6%–81.6%), high DP in 42.4% (95% CI 32.2%–53.1%), and low PA in 43.5% (95% CI 33.2%–54.2%). These findings indicate a high degree of emotional strain among residents, with substantial impairment across the other burnout dimensions (Table 2).

Table 3 presents the analysis of factors potentially associated with burnout syndrome and its subdimensions. The category of 20–40 weekly hours was not included in comparative analyses due to insufficient sample size. Living alone was significantly associated with a higher overall prevalence of burnout (36.4% vs. 16.7%; $p = 0.036$; Prevalence Ratio: 2.18; 95% CI 1.04–4.59). Among the burnout dimensions, female sex was associated with a higher frequency of EE (85.0% vs. 63.5%; $p = 0.039$; Prevalence Ratio = 1.34; 95% CI 1.05–1.71), suggesting greater vulnerability of women to the emotional component of burnout.

Weekly workload was significantly associated with both high EE ($p = 0.011$; Prevalence Ratio = 1.45; 95% CI 1.05–2.02) and low PA ($p = 0.046$; Prevalence Ratio = 1.84; 95% CI 1.00–3.37). However, no statistically significant association was observed between workload ≥ 60 hours and overall burnout ($p = 0.068$).

Psychotropic medication use was significantly associated with overall burnout syndrome ($p = 0.017$; Prevalence Ratio = 2.37; 95% CI 1.16–4.84), EE ($p = 0.016$; Prevalence

Table 1 Sample characteristics (n = 92).

Characteristic	n (%)
Sex	
Male	52 (56.5)
Female	40 (43.5)
Age group	
20–25 years	7 (7.6)
25–30 years	50 (54.3)
30–35 years	31 (33.7)
35–40 years	4 (4.3)
Living alone	44 (47.8)
Marital status	
Married/stable relationship/dating	61 (66.3)
Single	31 (33.7)
Year of residency	
First year	29 (31.5)
Second year	30 (32.6)
Third year	33 (35.9)
Weekly workload	
20–40 hours	1 (1.1)
40–60 hours	31 (33.7)
60 or more hours	60 (65.2)
Hold jobs outside the residency program	45 (48.9)
Anesthesiology as first-choice specialty	42 (45.7)
Pet ownership	48 (52.2)
Engagement in psychotherapy	29 (31.5)
Psychotropic medication use	38 (41.3)
Alcohol consumption	
No	40 (43.5)
Once or twice per week	45 (48.9)
More than twice per week	7 (7.6)
Physical activity	
No	26 (28.2)
1 to 4 times per week	59 (64.2)
More than 4 times per week	7 (7.6)

Ratio = 1.38; 95% CI 1.09–1.75), and DP ($p = 0.005$; Prevalence Ratio = 2.04; 95% CI 1.26–3.32). These findings should be interpreted as markers of psychological distress rather than causal factors. Similarly, residents undergoing psychotherapy exhibited a higher prevalence of DP ($p = 0.013$; Prevalence Ratio = 1.86; 95% CI 1.19–2.92), which may reflect treatment-seeking behavior among individuals experiencing greater emotional burden.

No statistically significant associations were observed between burnout and age, marital status, year of residency, physical activity, alcohol consumption, pet ownership, or anesthesiology as first-choice specialty.

Table 2 Prevalence of burnout (n = 92).

Variable	% [95% CI]
Overall burnout	26.1 [17.5 – 36.3]
Emotional exhaustion	72.8 [62.6 – 81.6]
Depersonalization	42.4 [32.2 – 53.1]
Low personal accomplishment	43.5 [33.2 – 54.2]

CI, Confidence Interval.

Table 3 Factors potentially associated with overall burnout, Emotional Exhaustion (EE), Depersonalization (DP), and low Personal Accomplishment (PA).

^a Characteristic	Overall Burnout	p	EE	p	DP	p	PA	p
Sex		0.160		0.039		0.296		0.638
Male	17 (32.7%)		33 (63.5%)		25 (48.1%)		21 (40.4%)	
Female	7 (17.5%)		34 (85.0%)		14 (35.0%)		19 (47.5%)	
Age group		0.792		0.875		0.082		0.216
20–25 years	2 (28.6%)		5 (71.4%)		4 (57.1%)		3 (42.9%)	
25–30 years	15 (30.0%)		38 (76.0%)		26 (52.0%)		18 (36.0%)	
30–35 years	6 (19.4%)		21 (67.7%)		8 (25.8%)		18 (58.1%)	
35–40 years	1 (25.0%)		3 (75.0%)		1 (25.0%)		1 (25.0%)	
Living alone		0.036		0.999		0.206		0.141
Yes	16 (36.4%)		32 (72.7%)		22 (50.0%)		23 (52.3%)	
No	8 (16.7%)		35 (72.9%)		17 (35.4%)		17 (35.4%)	
Marital status		0.802		0.322		0.824		0.514
Married	15 (24.6%)		42 (68.9%)		25 (41.0%)		25 (41.0%)	
Single	9 (29.0%)		25 (80.6%)		14 (45.2%)		15 (48.4%)	
Year of residency		0.383		0.999		0.560		0.254
First year	5 (17.2%)		21 (72.4%)		12 (41.4%)		9 (31.0%)	
Second year	10 (33.3%)		22 (73.3%)		15 (50.0%)		14 (46.7%)	
Third year	9 (27.3%)		24 (72.7%)		12 (36.4%)		17 (51.5%)	
Weekly workload		0.068		0.011		0.092		0.046
20–40 hours ^a								
40–60 hours	4 (12.9%)		18 (58.1%)		9 (29.0%)		9 (29.0%)	
≥ 60 hours	20 (33.3%)		49 (81.7%)		30 (50.0%)		31 (51.7%)	
Jobs outside program		0.202		0.485		0.833		0.674
Yes	14 (31.1%)		31 (68.9%)		20 (44.4%)		21 (46.7%)	
No	10 (21.3%)		36 (76.6%)		19 (40.4%)		19 (40.4%)	
First-choice specialty		0.641		0.347		0.207		0.294
Yes	12 (28.6%)		33 (78.6%)		21 (50.0%)		21 (50.0%)	
No	12 (24.0%)		34 (68.0%)		18 (36.0%)		19 (38.0%)	
Pet ownership		0.999		0.359		0.531		0.834
Yes	13 (27.1%)		37 (77.1%)		22 (45.8%)		20 (41.7%)	
No	11 (25.0%)		30 (68.2%)		17 (38.6%)		20 (45.5%)	
Psychotherapy		0.306		0.208		0.013		0.651
Yes	10 (34.5%)		24 (82.8%)		18 (62.1%)		14 (48.3%)	
No	14 (22.2%)		43 (68.3%)		21 (33.3%)		26 (41.3%)	
Psychotropic medication		0.017		0.016		0.005		0.087
Yes	15 (39.5%)		33 (86.8%)		23 (60.5%)		21 (55.3%)	
No	9 (16.7%)		34 (63.0%)		16 (29.6%)		19 (35.2%)	
Alcohol consumption		0.578		0.630		0.164		0.398
No	10 (25.0%)		27 (67.5%)		14 (35.0%)		14 (35.0%)	
1–2 week	11 (24.4%)		34 (75.6%)		20 (44.4%)		22 (48.9%)	
> 2 week	3 (42.9%)		6 (85.7%)		5 (71.4%)		4 (57.1%)	
Physical activity		0.593		0.698		0.999		0.207
No	6 (23.1%)		20 (76.9%)		11 (42.3%)		15 (57.7%)	
1–4 week	15 (25.4%)		41 (69.5%)		25 (42.4%)		22 (37.3%)	
> 4 week	3 (42.9%)		6 (85.7%)		3 (42.9%)		3 (42.9%)	

Data are presented as n (%) and compared using Fisher's exact test.

^a Categories with insufficient sample size were not included in statistical comparisons.

Discussion

This multicenter study, involving a high-response multicenter sample of anesthesiology residents in Rio Grande do Sul, Brazil, identified an overall prevalence of burnout of 26.1% using the strict diagnostic criteria requiring all 3 MBI-HSS subscales. This finding underscores the occupational vulnerability of this

population and highlights the need for immediate attention from residency training programs.

The observed burnout prevalence of 26.1% is consistent with the average reported in studies of residents in high-demand medical specialties in Brazil, although it remains substantially higher than what is observed in the general population. Importantly, because this estimate was derived

using the strictest definition (requiring the simultaneous presence of high EE, high DP, and low PA), it likely underestimates the broader extent of psychological distress. Indeed, high EE was present in 72.8% of residents, emerging as the dominant feature of burnout in this cohort.

The high prevalence of EE among anesthesiology residents reflects the intrinsic demands of training in this specialty, which requires continuous vigilance, rapid decision-making in critical situations, and sustained exposure to stressful environments. EE, as the primary dimension of burnout, is a sensitive marker of chronic psychological overload and has been consistently identified as the most prevalent component in meta-analyses of Brazilian and international resident populations.^{5,16}

Associations observed with variables such as psychotropic medication use, psychotherapy, and living alone likely reflect underlying psychological distress and contextual vulnerability rather than causal relationships. These findings should be interpreted cautiously, as no multivariable adjustment was performed, and causal inferences cannot be established. In addition, reverse causality and bidirectional relationships cannot be excluded, particularly for variables related to mental health care utilization.

In this context, the sample consisted predominantly of young adults (aged 25–35 years) with a balanced distribution across residency years, situating these findings within the period of greatest training intensity, during which the literature consistently reports heightened vulnerability to chronic occupational stressors.^{8–12,17–19}

The overall burnout prevalence observed in this study (26.1%) lies within the lower-to-mid range of estimates reported for residents and anesthesiologists, ranging widely from 20% to 60% depending on definitions, instruments, and cutoff points used.^{5,9,17–21} In large cohorts of Brazilian anesthesiologists assessed during the COVID-19 pandemic, overall burnout prevalence was reported at 19.6%, whereas more than 50% were classified as being “at high risk” when EE or DP were applied alone, illustrating how less stringent definitions can substantially inflate prevalence estimates.²² This discrepancy was addressed in a recent review documenting more than 140 different definitions of burnout, even among studies using the MBI to assess burnout, resulting in prevalence estimates ranging from 0% to 80.5%. Conversely, studies requiring the simultaneous presence of high EE, high DP, and low PA report a substantially lower prevalence of burnout than those based on isolated dimensions.²¹

Among residents, systematic reviews and multicenter studies report burnout prevalence ranging from 30% to 50%, with heterogeneity according to specialty, workload, and institutional context.^{17–19,22,23} Our findings align with a recent Thai study of medical residents demonstrating strong associations between working conditions, mental health, burnout, and medical errors, reinforcing the clinical and operational relevance of this outcome.¹⁹ In Brazil, although using heterogeneous instruments, studies of multiprofessional residents have also reported concerning levels of burnout and suggested organizational determinants (workload, resource availability, and supervision) similar to those identified in medical residency programs.¹⁵

Regarding the MBI dimensions, the predominance of EE (72.8%) is consistent with studies identifying EE as the most

stable and central component of the burnout construct, whereas DP and PA exhibit lower internal consistency and greater cultural variability.^{2,3,24} A recent review by Guille and Sen.²¹ highlights the weak correlation among the 3 domains and questions the validity of rigid cutoffs or composite scores for defining “cases”, a perspective that contextualizes both the magnitude and the precision of our estimates.

The association between working ≥ 60 hours per week and both EE and PA reflects one of the most consistently reported factors associated with burnout in the literature – excessive workload and sleep deprivation – consistently associated with higher levels of exhaustion, impaired performance, and safety concerns.^{8–12,19,21} The finding of higher EE among women (85.0% vs. 63.5%) is consistent with evidence suggesting greater female vulnerability to organizational stressors in environments characterized by high responsibility and limited autonomy, mediated by additional non-work-related burdens and institutional biases.^{21,25,26}

The significant association between living alone and overall burnout highlights the critical role of social support networks and interpersonal connectedness as modulators of resilience and coping strategies in high-stress settings such as residency training. This finding aligns with contemporary emphasis on interventions that strengthen peer support and promote accessible, destigmatized mental health care within educational and training institutions.^{6,7,20,21}

The strong association between psychotropic medication use and overall burnout (as well as EE and DP) likely reflects greater severity of psychological distress and/or the presence of comorbid anxiety and depressive disorders. This relationship is well documented in the literature, which recognizes substantial conceptual and dimensional overlap between burnout and clinical conditions such as anxiety and depression. In this context, psychotropic medication use functions as a marker of distress severity, reinforcing the urgency of providing accessible mental health care.^{11,12,21} Similarly, the higher prevalence of DP among residents undergoing psychotherapy likely reflects treatment-seeking in more severe cases (indication bias) and should not be interpreted as evidence of therapeutic ineffectiveness.

This study has limitations that warrant consideration. We used the Portuguese-validated MBI-HSS and a conservative definition of overall burnout (high EE + high DP + low PA), an approach that is more specific and aligned with recommendations aimed at reducing prevalence overestimation.^{2,3,5,17,24} Nevertheless, the historical heterogeneity of the construct and ongoing debate regarding cutoffs and subscale combinations remain epistemological limitations that affect comparability across studies.²¹ Also, the cross-sectional design precludes causal inferences, and self-reported data may be subject to recall and social desirability bias. Additionally, no multivariable analysis was performed, which limits the ability to control for potential confounding factors and to estimate independent associations. Considering the rule-of-thumb of at least 10 events per variable, the number of burnout cases ($n \approx 24$) would not support a stable multivariable model including the proposed covariates. Therefore, the findings should be interpreted as exploratory associations rather than independent effects.

Participants were clustered within training centers, and center-level characteristics may have influenced both

exposures and outcomes. Due to the limited number of centers, more advanced multilevel approaches were not feasible. Given the multiple comparisons performed across several outcomes and exposure variables, there is an increased risk of type I error (false-positive findings). Therefore, the observed associations should be interpreted cautiously, with greater emphasis placed on effect sizes and confidence intervals rather than isolated p-values.

Non-response bias should also be considered, as residents who did not participate may differ systematically from respondents. However, the high response rate (84.4%) reduces the likelihood of substantial bias.

Despite these limitations, this study has important strengths, including its multicenter design, high response rate, and use of a validated instrument with a conservative definition of burnout. These characteristics enhance the internal consistency and regional relevance of the findings.

Our results highlight 3 priority axes for action within anesthesiology residency programs: 1) Workload management (monitoring duty hours, implementing sustainable schedules, and mitigating sleep deprivation); 2) Structured institutional support (mentorship, environments free from harassment or mistreatment, and safe psychological support channels); and 3) Streamlined access to evidence-based mental health care and well-being strategies.^{6,7,19-21} A recent study of Brazilian anesthesiologists suggests that leisure time exceeding 5 hours per week may be a protective factor, offering an operationally actionable insight for local policy development.²²

Longitudinal studies are needed to elucidate trajectories of EE, DP, and PA across first, second, and third years of residency and during the transition to independent practice, as well as to evaluate multilevel interventions (systemic and individual) using clinical and educational outcomes such as errors, performance, retention, quality, and safety.^{6,9,19,21} Efforts to standardize measurement (by reporting domains separately and avoiding non-validated composite scores) may further enhance international comparability and support the development of more precise policies.^{21,24}

Conclusion

The overall prevalence of burnout (26.1%) among anesthesiology residents in Rio Grande do Sul was high, with emotional exhaustion emerging as the predominant component. Living alone and psychotropic medication use were associated with burnout, the latter likely representing a marker of psychological distress rather than a causal factor. Given the cross-sectional design and exploratory analyses, these findings should be interpreted with caution. These results highlight the need for institutional policies focused on strengthening psychosocial support and promoting mental health throughout medical training.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare no conflicts of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (version 5.2) in the writing process to improve readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ethical approval

The study was approved by the Institutional Review Board of Hospital Ernesto Dornelles (HED), Av. Ipiranga 1801, 11° andar, Azenha, 90160-093, Porto Alegre, RS, Brazil (approval number: CAAE 85752324.2.0000.5304), and all research and methods adhered to the provisions of the Declaration of Helsinki and the STROBE guidelines.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Editor

Liana Azi

References

- Freudenberger HJ. Staff burn-out. *J Soc Issues*. 1974;30:159–65.
- Maslach C, Jackson SE. The measurement of experienced burnout. *J Organ Behav*. 1981;2:99–113.
- Maslach C, Schaufeli WB, Leiter MP. Job burnout. *Annu Rev Psychol*. 2001;52:397–422.
- World Health Organization. Burn-out an “occupational phenomenon”: International Classification of Diseases [Internet] [cited 2025 Dec 16]. Available from <https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases>; 2019.
- Rotenstein LS, Torre M, Ramos MA, et al. Prevalence of burnout among physicians: a systematic review. *JAMA*. 2018;320:1131–50.
- Panagioti M, Panagopoulou E, Bower P, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Intern Med*. 2017;177:195–205.
- Dzau VJ, Kirch D, Nasca T. Preventing a parallel pandemic – A national strategy to protect clinicians’ well-being. *N Engl J Med*. 2020;383:513–5.
- Patel RS, Bachu R, Adikey A, Malik M, Shah M. Factors related to physician burnout and its consequences: a review. *Behav Sci (Basel)*. 2018;8:98.
- Sun H, Warner DO, Macario A, Zhou Y, Culley DJ, Keegan MT. Repeated cross-sectional surveys of burnout, distress, and depression among anesthesiology residents and first-year graduates. *Anesthesiology*. 2019;131:668–77.

10. De Hert S. Burnout in healthcare workers: prevalence, impact and preventative strategies. *Local Reg Anesth.* 2020;13: 171–83.
11. Bagatini A, Bumaguin DB, Bagatini LZ, Ardaís AP. The relationship between burnout syndrome and emotional intelligence in anesthesiologists during the COVID-19 pandemic: a cross-sectional study. *J Patient Saf Risk Manag.* 2025;30:149–56.
12. Dyrbye LN, Shanafelt TD, Sinsky CA, et al. Burnout among health care professionals: a call to explore and address this underrecognized threat to safe, high-quality care. *NAM Perspect.* 2017;2017.
13. Eckleberry-Hunt J, Kirkpatrick H, Barbera T. The problems with burnout research. *Acad Med.* 2018;93:367–70.
14. Tironi MOS, Teles JMM, Barros DDS, et al. Prevalence of burnout syndrome in intensivist doctors in five Brazilian capitals. *Rev Bras Ter Intensiva.* 2016;28:270–7.
15. Mesquita V, Malagris L. Burnout syndrome in multiprofessional health residents of a university hospital. *Rev SBPH.* 2020;23: 65–76.
16. Tamayo MR, Tróccoli BT. Construction and factorial validation of the Burnout Characterization Scale (ECB). *Estud Psicol (Natal).* 2009;14:213–21.
17. Chong MYF, Lin SHX, Lim WY, Ong J, Kam PCA, Ong SGK. Burnout in anaesthesiology residents: a systematic review of its prevalence and stressors. *Eur J Anaesthesiol.* 2022;39:368–77.
18. Bui D, Winegarner A, Kendall MC, Almeida M, Apruzzese P, De Oliveira GS. Burnout and depression among anesthesiology trainees in the United States: an updated national survey. *J Clin Anesth.* 2023;84:110990.
19. Surawattanasakul V, Siviroj P, Kiratipaisarl W. Resident physician burnout and association with working conditions, psychiatric determinants, and medical errors: a cross-sectional study. *PLoS One.* 2024;19:e0312839.
20. World Health Organization. Mental health and psychosocial considerations for health workers [Internet]. 2020 [cited 2025 Dec 16]. Available from: <https://www.who.int/docs/default-source/coronaviruse/mental-health-considerations.pdf>.
21. Guille C, Sen S. Burnout, depression, and diminished well-being among physicians. *N Engl J Med.* 2024;391:1519–27.
22. Azi LMTA, Ferreira TS, Cerqueira-Silva T, Diego LAS, Albuquerque MAC, Azi ML. Prevalence of burnout syndrome in Brazilian anesthesiologists during the COVID-19 pandemic: a cross-sectional survey. *PLoS One.* 2025;20:e0313538.
23. Rodrigues H, Cobucci R, Oliveira A, et al. Burnout syndrome among medical residents: a systematic review and meta-analysis. *PLoS One.* 2018;13:e0206840.
24. Almeida GDC, Souza HRD, Almeida PCD, Almeida BDC, Almeida GH. The prevalence of burnout syndrome in medical students. *Arch Clin Psychiatry (São Paulo).* 2016;43:6–10.
25. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory manual. 4th ed. Menlo Park. Mind Garden; 2018.
26. Hu YY, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med.* 2019;381:1741–52.

ORIGINAL INVESTIGATION

Preoperative angiotensin-converting enzyme inhibitor or angiotensin receptor blocker use and acute kidney injury after high-risk cardiac surgery: a multicenter prospective cohort study¹



Iñigo Rubio ^a, Angel Candela ^b, Gemma Echarri ^a, Raquel Callejas ^a,
Eduardo Tamayo ^c, David Nagore ^d, Jorge M. Núñez-Córdoba ^{e,f,*}, Marc Vives ^{a,g}

^a Clínica Universidad de Navarra, Department of Anesthesiology & Critical Care Medicine, Pamplona, Spain

^b Hospital Universitario Ramón y Cajal, Department of Anesthesiology & Perioperative Medicine, Madrid, Spain

^c Hospital Clínico Universitario de Valladolid, Department of Anesthesiology & Perioperative Medicine, Valladolid, Spain

^d St Bartholomew's Hospital, Barts Heart Centre, Department of Anaesthesia & Perioperative Medicine, London, United Kingdom

^e Clínica Universidad de Navarra, Central Clinical Trials Unit, Research Support Service, Pamplona, Spain

^f Public University of Navarra, Department of Health Sciences, Pamplona, Spain

^g Instituto de Investigación Sanitaria de Navarra, IdiSNA, Pamplona, Spain

Received 2 January 2026; accepted 14 June 2026

Available online 1 July 2026

KEYWORDS

Acute kidney injury;
Angiotensin-converting
enzyme inhibitors;
Angiotensin receptor
antagonists;
Cardiac surgical
procedures;
Kidney diseases;
Risk factors

Abstract

Background: The impact of preoperative Angiotensin-Converting Enzyme Inhibitors or Angiotensin Receptor Blockers (ACEI/ARB) on cardiac surgery-associated Acute Kidney Injury (AKI) remains elusive. We sought to evaluate the association of ACEI/ARB use with stage 2 or 3 AKI and in-hospital mortality, while evaluating whether preoperative withholding versus continuation strategies modulate clinical outcomes.

Methods: Multicenter prospective cohort study involving 14 Spanish and British hospitals. The study population comprised high-risk cardiac surgery patients (Cleveland Clinic Score ≥ 4).

Results: Among 249 patients (mean age: 69.5 years; 39% women; 53.8% ACEI/ARB users), the overall incidence of stage 2 or 3 AKI was 32.9%. ACEI/ARB use was associated with a significantly higher likelihood of stage 2 to 3 AKI (38.8% vs. 26.1%; adjusted OR = 2.79; 95% CI 1.47–5.30; $p = 0.002$). The adjusted risk difference was 0.188 (95% CI 0.079–0.298; $p = 0.001$), yielding a number needed to harm of 5.3. Sensitivity analyses confirmed the robustness of these findings, with an E-value of 2.73 (lower limit: 1.72) and consistent results using augmented inverse probability weighting (average treatment effect: 0.191; 95% CI 0.081–0.300; $p = 0.001$). No statistically significant associations were observed in in-hospital mortality or the impact of preoperative withholding versus continuation strategies on clinical outcomes.

¹And on behalf of the Spanish Perioperative Cardiac Surgery Research Group.

* Corresponding author.

E-mail: jnunezco@unav.es (J.M. Núñez-Córdoba).

Conclusion: This study suggests preoperative ACEI/ARB use may be associated with an increased incidence of stage 2 or 3 AKI. While differences in in-hospital mortality and preoperative withholding versus continuation strategies did not reach statistical significance, these findings reveal clinical trends that warrant further investigation.

© 2026 Published by Elsevier España, S.L.U. on behalf of Sociedade Brasileira de Anestesiologia. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Background

Patients undergoing cardiac surgery frequently develop cardiac surgery-associated Acute Kidney Injury (AKI), with an incidence that is up to 30%.¹ This is associated to a higher risk of morbidity, renal replacement therapy requirement, mortality, and delays of hospital discharge.¹ There have been previous attempts to calculate cardiac surgery-associated AKI risk in these patients. The Cleveland score was validated to identify patients undergoing cardiac surgery who may be at risk for AKI.²⁻⁴ The etiology of this cardiac surgery-associated AKI is multifactorial, systemic inflammatory response syndrome secondary to surgery-induced stress,^{5,6} and shifting fluid volumes that can lead to fluid overload.⁷

Moreover, these patients usually take drugs that may worsen renal function, such as Angiotensin-Converting Enzyme Inhibitors or Angiotensin Receptor Blockers (ACEI/ARB). Whether the use of ACEI/ARB is independently associated with an increased risk of cardiac surgery-associated AKI needs to be clarified.⁸ 2024 EACTS guidelines on perioperative medication in adult cardiac surgery suggested that continuing ACEI and ARBs should be considered until open-heart surgery,⁹ based on observational retrospective data on low-risk patients undergoing cardiac surgery.^{10,11}

Prior studies on the risk/benefit of ACEI/ARB before cardiac surgery reported conflicting results.^{8,12-25}

Some have recommended maintaining both ACEI or ARB because, during Cardiopulmonary Bypass (CPB), the Renin-Angiotensin-Aldosterone System (RAAS) is activated, potentially leading to renal hypoperfusion. Therefore, continued administration of ACEI and ARB might attenuate this RAAS activation and prevent renal medullary ischemia induced by angiotensin II.^{8,12-17} Furthermore, withdrawing ACEI/ARB may also precipitate acute heart failure and increase the risk of postoperative hypertension.^{19,20}

On the other hand, others have found that ACEI/ARB may be associated with an increased risk of cardiac surgery-associated AKI secondary to a higher risk of hypotension and vasogenic shock,¹⁸ leading to renal hypoperfusion and a reduced Glomerular Filtration Rate (GFR) due to dilatation of the efferent glomerular arteriole.²¹⁻²⁵

We hypothesized that preoperative ACEI/ARB use confers an increased risk of postoperative AKI. Accordingly, we sought to evaluate the association between the preoperative ACEI/ARB use and the incidence of stage 2 or 3 AKI as our primary endpoint. Secondary objectives included an assessment of in-hospital mortality, alongside exploratory analyses within the user cohort to elucidate whether withholding versus continuing therapy is associated with the incidence of stage 2 or 3 AKI and in-hospital mortality.

Methods

Study design, setting and participants

This prospective multicenter cohort study included high-risk cardiac surgery patients with Cleveland Clinic Score² (CCS) ≥ 4 and aged ≥ 18 years from 14 centers in Spain and the United Kingdom, recruited between July and December 2017. Patients undergoing heart transplantation and pericardiectomy were excluded, as were those with severe preoperative renal insufficiency, defined as need for Renal Replacement Therapy (RRT) or estimated GFR $< 15 \text{ mL} \cdot \text{min}^{-1} \cdot 1.73 \text{ m}^{-2}$. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (IRB) (Reference #EPA027/15). The informed consent was obtained from all patients or their proxies prior to surgery. The reporting of this study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

Exposure, covariates and outcomes

Patients were classified into groups based on ACEI/ARB use. The baseline severity of the patients was estimated using the CCS. The CCS ranges between 0 (lowest risk) and 17 (highest risk) points. This CCS incorporates the following major risk factors: sex, congestive heart failure, left ventricular ejection fraction $< 35\%$ (LVEF $< 35\%$), preoperative use of Intra-Aortic Balloon Pump (IABP), Chronic Obstructive Pulmonary Disease (COPD), insulin-requiring diabetes, previous cardiac surgery, emergency surgery, surgery type (only coronary artery bypass graft, only valve, and other cardiac surgeries), and preoperative creatinine.⁴

The primary outcome was the development of stage 2 or 3 AKI according to KDIGO criteria.²⁶ To apply these criteria, both creatinine and urine output were measured at Intensive Care Unit (ICU) admission and at 24h, 48h, 72h, 96h, and 7 days postoperatively. The secondary outcome was in-hospital mortality.

The inter-center variability was addressed using a standardized Case Report Form (CRF) shared by all participating institutions. The standardized CRF was designed to ensure that all clinically significant variables related to AKI in cardiac surgery were strictly defined and captured uniformly across all sites.

Statistical analysis

A post-hoc power sensitivity analysis was conducted to determine the study's capacity to detect varying effect sizes for the primary event. With a total sample size of 249 and a baseline incidence of 26.1% for stage 2 or 3 AKI (reference

group: $n = 115$), the study achieved a statistical power of 72.1% to detect an OR of 2.0. For larger effect sizes, the power increased to 85.3% (OR = 2.25), 92.9% (OR = 2.5), 96.8% (OR = 2.75), and 98.6% for an OR of 3.0. These results suggest that the study was sufficiently powered to detect moderate-to-large clinical effects in the primary analysis. All other analyses were considered exploratory in nature.

The primary hypothesis of this study was that preoperative ACEI/ARB therapy was associated with an increased risk of AKI. Consequently, the primary analysis evaluated the association between ACEI/ARB use and the incidence of stage 2 or 3 AKI. Secondary analyses included the evaluation of in-hospital mortality between ACEI/ARB users and non-users. Furthermore, exploratory secondary analyses were conducted within the ACEI/ARB user group to compare the impact of preoperative withholding versus continuation strategies on both stage 2 or 3 AKI and in-hospital mortality.

Continuous variables were expressed as mean and Standard Deviation (SD), with comparisons performed using the Student's *t*-test. Categorical variables were presented as frequencies (%) and analyzed using the Chi-Squared test. Baseline balance between groups was assessed using Absolute Standardized Mean Differences (ASMD). The association between treatment exposure and clinical outcomes was first evaluated using a multivariable logistic regression model. Results were reported as Odds Ratios (OR) with corresponding 95% Confidence Intervals (95% CI). The adjusted Risk Difference (RD) and 95% CI was calculated. For clinical interpretation of significant effects, the Number Needed to Harm (NNH) was determined. To address potential confounding by indication and provide a robust framework for causal inference, we further used Augmented Inverse-Probability Weighting (AIPW). This doubly robust estimator was used to calculate the Average Treatment Effect (ATE). To assess the robustness of the observed point estimates against unmeasured confounding in statistically significant associations, the E-value was calculated for both the estimate and the limit of the confidence interval closest to the null. All tests were two-sided, and p -values < 0.05 were considered to indicate statistical significance. All statistical analyses were performed using Stata 19 (StataCorp. 2025. Stata Statistical Software: Release 19. College Station, TX: StataCorp LLC).

Results

The patient selection process is illustrated in Figure 1. From an initial screening of 2,775 cardiac surgery patients, 261 met the eligibility criteria. Data completeness was generally high. Missing data for the variables of interest were minimal, limited to ACEI/ARB use (3.5%) and LVEF $< 35\%$ (1.2%), the latter being the only severity-defining variable with missing values. Out of 261 eligible patients, 12 (4.6%) were excluded due to missing data, leaving 249 patients for the final analysis.

The baseline demographic and clinical characteristics of the study population are summarized in Table 1, categorized by ACEI/ARB use. A total of 134 patients (53.8%) were ACEI/ARB users. The cohort had an overall mean age of 69.5 years (SD = 11.1), with a predominant representation of white patients (98.4%) and a moderate proportion of women (39%). On average, patients in the ACEI/ARB group were slightly older (71.2, SD: 12.4 vs. 67.5, SD = 12.4 years; AMSD = 0.341; $p = 0.009$) and showed a higher prevalence of hypertension (90.3% vs. 73.9%; AMSD = 0.438; $p = 0.001$), LVEF $< 35\%$ (14.9% vs. 6.1%; AMSD = 0.291, $p = 0.025$), and COPD (35.8% vs. 22.6%; AMSD = 0.294; $p = 0.023$) compared to non-users. The prevalence of CKD (stage III–V) was high and balanced between groups (77.1%). The patients receiving ACEI/ARB were more likely to undergo elective surgery (73.1% vs. 59.1%; AMSD = 0.299; $p = 0.019$) and had shorter mean CBP time (124.7, SD: 67.9 vs. 144.8, SD = 71.5 minutes; AMSD = 0.288; $p = 0.024$). The ACEI/ARB group presented a lower mean CCS compared to non-users (6.0, SD: 1.5 vs. 6.4, SD = 1.5 years; AMSD = 0.255; $p = 0.046$). No significant differences were observed in recorded preoperative covariates, including hemoglobin levels and GFR, across both groups. The intraoperative management was comparable between groups, including RBC transfusion rates (61%) and the use of vasopressors or inotropes.

Association of ACEI/ARB use with the risk of cardiac surgery-associated stage 2 or 3 AKI and in-hospital mortality

The overall incidence of stage 2 or 3 AKI was 32.9%. Specifically, 28.1% of patients developed Stage 1, 12.8% Stage 2,

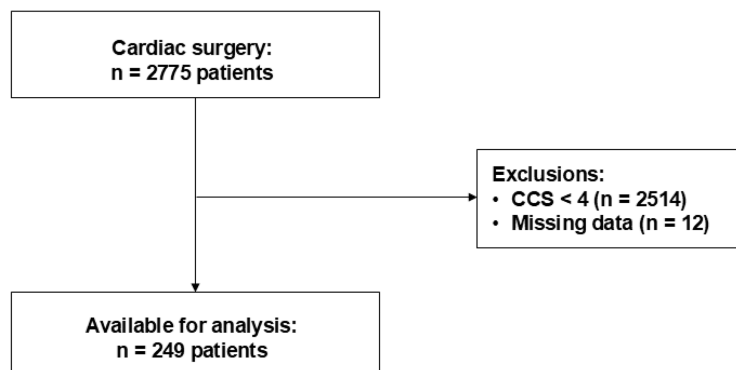


Figure 1 Study flow diagram. CCS, Cleveland Clinic Score. The diagram illustrates the selection process of patients from the initial screening to the final analytical sample. Reasons for exclusion are provided according to STROBE guidelines. The CCS is based on ten major risk factors: sex, congestive heart failure, left ventricular ejection fraction $< 35\%$, preoperative use of intra-aortic balloon pump, chronic obstructive pulmonary disease, insulin-requiring diabetes, previous cardiac surgery, emergency surgery, surgery type, and preoperative creatinine.

Table 1 Characteristics of the study population.

	All patients (n = 249)	No ACEI/ARB use (n = 115)	ACEI/ARB use (n = 134)	ASMD	p-value
Age (years)	69.5 (11.1)	67.5 (12.4)	71.2 (9.6)	0.341	0.009
Women, n (%)	97 (39.0)	40 (34.8)	57 (42.5)	0.160	0.211
Skin color, n (%)					
White	245 (98.4)	114 (99.1)	131 (97.98)	0.111	0.626
Black	4 (1.6)	1 (0.9)	3 (2.2)		
BMI (kg.m ⁻²)	27.9 (4.9)	27.3 (4.5)	28.5 (5.1)	0.265	0.039
Preoperative hemoglobin, g.dL ⁻¹	12.1 (2.0)	11.9 (2.0)	12.3 (2.0)	0.200	0.117
Preoperative comorbid disease, n (%)					
Hypertension	206 (82.7)	85 (73.9)	121 (90.3)	0.438	0.001
Insulin-requiring diabetes	55 (22.1)	26 (22.6)	29 (21.6)	0.023	0.855
CHF	140 (56.2)	66 (57.4)	74 (55.2)	0.044	0.731
LVEF < 35%	27 (10.8)	7 (6.1)	20 (14.9)	0.291	0.025
Previous cardiac surgery	53 (21.3)	28 (24.4)	25 (18.7)	0.138	0.274
COPD	74 (29.7)	26 (22.6)	48 (35.8)	0.294	0.023
Chronic kidney disease stage III–V	192 (77.1)	89 (77.4)	103 (76.9)	0.013	0.922
Serum creatinine, mg.dL ⁻¹	1.5 (0.6)	1.6 (0.7)	1.5 (0.5)	0.153	0.234
eGFR, mL.min ⁻¹ per 1.73 m ²	48.1 (20.9)	48.7 (22.2)	47.5 (19.9)	0.055	0.665
Serum uric acid, mg.dL ⁻¹	7.0 (1.8)	6.9 (1.8)	7.1 (1.8)	0.136	0.286
Preoperative use of IABP, n (%)	18 (7.2)	9 (7.8)	9 (6.7)	0.043	0.736
Type of surgery, n (%)					
CABG	27 (10.8)	11 (9.6)	16 (11.9)	0.043	0.781
Valve	72 (28.9)	35 (30.4)	37 (27.6)		
Other cardiac surgeries	150 (60.2)	69 (60)	81 (60.5)		
Timing, n (%)					
Elective surgery	166 (66.7)	68 (59.1)	98 (73.1)	0.299	0.019
Urgent / Emergency surgery	83 (33.3)	47 (40.9)	36 (26.9)		
CCS	6.2 (1.5)	6.4 (1.5)	6.0 (1.5)	0.255	0.046
Cross clamp time (minutes)	96.9 (52.0)	104.1 (55.5)	90.7 (48.1)	0.257	0.043
CPB time (minutes)	134.0 (70.2)	144.8 (71.5)	124.7 (67.9)	0.288	0.024
Intraoperative RBC transfusion, n (%)	152 (61.0)	70 (60.9)	82 (61.2)	< 0.001	0.958
Intraoperative RBC transfusion (units)	1.7 (1.9)	1.7 (2.0)	1.7 (1.9)	0.018	0.886
Nadir hemoglobin, g.dL ⁻¹	7.9 (1.4)	7.8 (1.4)	8.1 (1.3)	0.205	0.116
Intraoperative dobutamine, n (%)	119 (47.8)	54 (47.0)	65 (48.5)	0.087	0.807
Intraoperative norepinephrine, n (%)	170 (68.3)	81 (70.4)	89 (66.4)	0.031	0.497

Data are presented as mean (standard deviation) unless otherwise stated. ASMD, Absolute Standardized Mean Difference; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers; BMI, Body Mass Index; CABG, Coronary Artery Bypass Graft; CCS, Cleveland Clinic Score; CHF, Congestive Heart Failure; COPD, Chronic Obstructive Pulmonary Disease; CPB, Cardiopulmonary Bypass; eGFR, estimated Glomerular Filtration Rate; IABP, Intra-Aortic Balloon Pump; LVEF < 35%, Left Ventricular Ejection Fraction < 35%; RBC, Red Blood Cell.

and 20.1% Stage 3 AKI, with 15.7% requiring RRT. Among ACEI/ARB users, 52 out of 134 (38.8%) developed stage 2 or 3 AKI, compared to 30 out of 115 (26.1%) in non-users. The multivariable logistic regression model showed a significant association between ACEI/ARB and stage 2 or 3 AKI following cardiac surgery (OR = 2.79; 95% CI 1.47–5.30; $p = 0.002$) after adjusting for age, hypertension, LVEF < 35%, COPD, preoperative hemoglobin, surgical urgency and CPB time. In terms of absolute impact, the adjusted RD was 0.188 (95% CI 0.079–0.298; $p = 0.001$), which corresponds to a NNH of 5.3 (95% CI 3.4–12.7). To assess the potential impact of unmeasured confounding, our sensitivity analysis yielded an E-value of 2.73 for the point estimate (with a lower limit of 1.72). This suggests that while the observed effect is robust to weak confounders, a moderately strong unmeasured factor could potentially render the results non-significant. These findings remained consistent in sensitivity analyses using AIPW. The estimated ATE was 0.191 (95% CI 0.081–0.300; $p = 0.001$), further supporting the initial

multivariable logistic regression findings. After weighting, satisfactory balance was achieved, with all ASMD below 0.1 (Supplementary Fig. 1).

The incidence of in-hospital mortality was 18.7% in the ACEI/ARB group compared with 13.9% in the non-users group, representing a non-statistically significant trend toward higher mortality among ACEI/ARB users (adjusted OR = 1.54; 95% CI 0.71–3.34; $p = 0.272$) (Fig. 2). The ATE for in-hospital mortality estimated using AIPW remained consistent with this initial observation (ATE = 0.054; 95% CI -0.049–0.156; $p = 0.306$).

Clinical outcomes of ACEI/ARB withholding 48 hours prior to the surgery compared with continuation until the day before surgery

The 134 ACEI/ARB users were analyzed to explore the impact of the preoperative ACEI/ARB management strategies (65 patients in the continuing group and 69 patients in

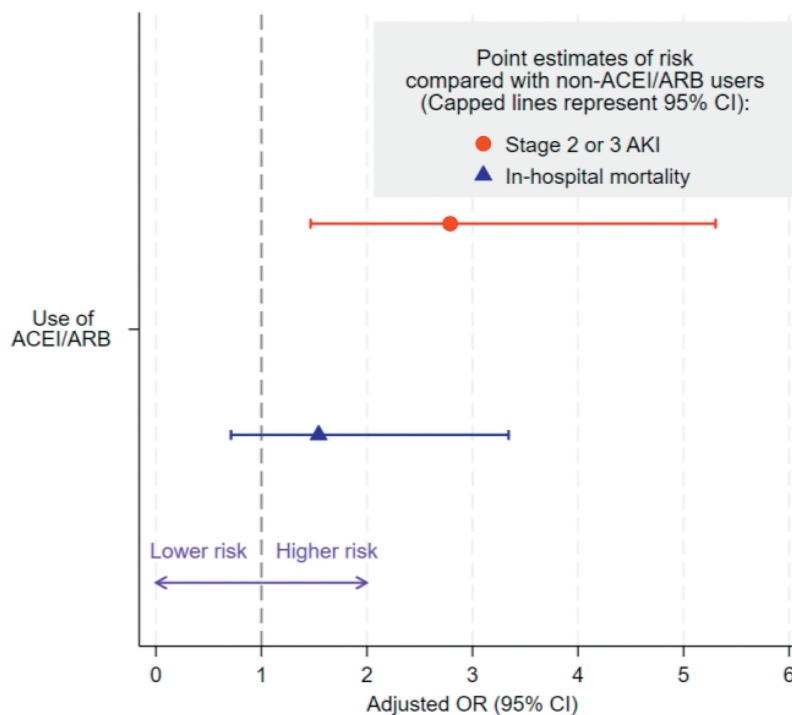


Figure 2 Risk of cardiac surgery-associated stage 2 or 3 AKI and in-hospital mortality according to the use of ACEI/ARB. 95% CI, 95% Confidence Interval; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers; AKI, Acute Kidney Injury; OR, Odds Ratio.

the withholding group). The incidence of Stage 2 or 3 AKI was 40.0% in the continuing group compared to 37.7% in the withholding group. After adjustment for age, preoperative hemoglobin levels, hypertension status, LVEF < 35%, chronic obstructive pulmonary disease, surgical urgency, and cardiopulmonary bypass time, withholding treatment was not associated with a significant reduction in the risk of AKI (adjusted OR = 0.68; 95% CI 0.30–1.51). Furthermore, the ATE showed a non-significant risk difference of -8.3 percentage points (95% CI -23.7% to 7.2%; $p = 0.293$). These results provide a hypothesis-generating signal suggesting a potential benefit of ACEI/ARB discontinuation 48 hours before surgery (Table 2).

Similarly, in-hospital mortality was lower in the withholding group (15.9% vs. 21.5%), although this difference did not reach statistical significance (adjusted OR = 0.78; 95% CI 0.29–2.07). The ATE for mortality was -3.4 percentage points (95% CI -15.3%–8.6%; $p = 0.581$). Consistent with the AKI findings, these results suggest a potential, albeit non-significant, trend toward improved clinical outcomes with preoperative ACEI/ARB discontinuation (Table 3).

Discussion

This study suggests a potential association between ACEI/ARB use and a greater likelihood of cardiac surgery-associated stage 2 or 3 AKI in high-risk cardiac surgery patients. While withholding ACEI/ARB for 48 hours prior to surgery may offer some potential benefits over continuing ACEI/ARB until the day before surgery, the statistical significance was lacking, and so further studies are needed to

confirm this association and establish its clinical relevance. These results are consistent with some existing evidence and contribute to clarifying controversial findings in the literature.

These findings are consistent with previous studies regarding the association between cardiac surgery-associated AKI and the use of ACEI/ARB vs. nonusers.^{13,23,24,27}

A meta-analysis of 69,027 patients from four prospective trials and 25 retrospective observational studies found that preoperative RAS-blocker use was associated with an increased incidence of postoperative AKI (OR = 1.17; 95% CI 1.01–1.36; $p = 0.04$) and mortality (OR = 1.20; 95% CI 1.06–1.35; $p = 0.005$).²³ A prospective cohort study of 1,594 adults undergoing cardiac surgery found that ACEI/ARB use was associated with functional AKI in patients who had no treatment (31%), discontinued treatment (34%), or continued treatment (42%; p for trend 0.03). However, it was not associated with structural AKI (neutrophil gelatinase-associated lipocalin, kidney injury molecule-1, interleukin-18, or liver fatty acid-binding protein).²⁷ Furthermore, data from a meta-analysis of thirteen studies, 3 Randomized Clinical Trials (RCT) and 10 observational studies including 31,390 patients, found that the use of ACEI was associated with an increase risk of hypotension and postoperative renal dysfunction, but had no significant impact on the risk of early mortality in patients undergoing on-pump CABG.²⁴ The hypothesis about the mechanism which might be involved in the increase risk of AKI after cardiac surgery is that ACEI/ARB may cause dilatation of the efferent glomerular arteriole leading to a reduced GFR, as well as, a higher risk of hypotension and vasogenic shock leading to renal hyperperfusion.^{13,18,21–24}

Table 2 Association between preoperative ACEI/ARB management strategies and postoperative stage 2 or 3 acute kidney injury.

	n	Stage 2 or 3 AKI (%)	Crude OR (95% CI)	Adjusted OR (95% CI)	ATE	p-value
Continuing	65	40	Reference	Reference	−0.083	0.293
Withholding	69	37.7	0.91 (0.45 – 1.82)	0.68 (0.30 – 1.51)	(−0.237 – 0.072)	

95% CI, 95% Confidence Interval; AKI, Acute Kidney Injury; ATE, Average Treatment Effect. Continuing was defined as treatment until the day before surgery. Withholding was defined as cessation 48 hours or more before surgery. The OR derived from logistic regression models. The multivariable logistic regression model was adjusted for age, preoperative hemoglobin levels, hypertension status, left ventricular ejection fraction < 35%, chronic obstructive pulmonary disease, surgical urgency, and cardiopulmonary bypass time. The ATE (risk difference) was estimated using an adjusted augmented inverse probability weighting model. The p-value refers to the ATE estimate.

In contrast, several other observational studies showed that the preoperative use of ACEI is not associated to AKI after cardiac surgery.^{8,10,28} A large-scale meta-analysis of 76,321 patients (including 2 RCT plus 21 cohort studies) showed no association between AKI and patients who received perioperative Renin-Angiotensin System Inhibitors (RASi) preoperatively compared to those who did not receive them.⁸ However, the heterogeneity in this study was relatively high. Most of the included studies were cohort studies (only 2 studies were RCT). A large retrospective cohort study also observed that the use of ACEI was not associated with AKI and was associated with less mortality after cardiac surgery.¹⁰ However, the retrospective design of this study involves several limitations. Another recent observational retrospective cohort study, including propensity score match analysis, also observed that the use of preoperative ACEI within 48h of Coronary Artery Bypass Graft (CABG) surgery was associated with a decrease in AKI and early mortality at 30 days after CABG. However, patients in this cohort were significantly lower-risk patients for developing AKI than our cohort (overall AKI 6.6% vs. 58.6%).¹¹

Our findings regarding “withholding vs maintenance” of ACEI before cardiac surgery are also consistent with previous studies.^{29–31} A multicentric observational study on 1,072 patients undergoing cardiac surgery showed that the maintenance of ACEI compared to withholding ACEI 24h before surgery was not associated to AKI after adjusting by propensity matching.²⁹ A RCT of 121 patients under current treatment with ACEI/ARB and undergoing cardiac surgery demonstrated no between-group differences in the incidence of cardiac surgery-associated AKI (1.7% vs. 1.6%, $p = 0.991$) between patients who kept taking their ACEI/ARB until the day of surgery and those who discontinued 48 hours before surgery.³⁰ A meta-analysis of 1,663 patients from three RCTs and three prospective cohort studies found no significant between-group difference between patients who

continued therapy versus those who discontinued their ACEI/ARB 24–48h before cardiac surgery (RR = 1.17, 95% CI 0.99–1.38).³¹ Pigott et al. carried out a RCT with 40 patients undergoing CABG that were assigned to either discontinue or continue their ACEI medication until the day before surgery. Notably, the patients who discontinued ACEI in this study were more likely to have hypertension and required less vasopressors and more vasodilators for blood pressure control.¹⁹ The optimal timing for withholding ACEI/ARB is not clear. Future research could investigate whether longer time (4 days or higher vs. 24–48h) is required for withholding ACEI before cardiac surgery.

Our results can be also interpreted within the framework of the ACC/AHA 2024 guidelines for perioperative cardiovascular management for non-cardiac surgery, which recommend an individualized approach to perioperative ACE inhibitor and ARB management based on the indication for therapy and surgical risk. On one hand, for patients taking ACEIs/ARBs for hypertension, omission 24 hours before surgery may be beneficial to limit intraoperative hypotension (Class 2b, Level B-R). On the other hand, for patients with an indication for Heart Failure with reduced Ejection Fraction (HFrEF), perioperative continuation is reasonable (Class 2a, Level C-EO).³²

An important consideration is that patients with high (SBP > 160 mmHg) or low (SBP < 105 mmHg) blood pressure are excluded from RCTs, and enrollment of high-risk patients, including those with HFrEF, has been limited.³³

The most recent and largest RCT on the use of ACEIs or ARBs in non-cardiac surgery, published after the ACC/AHA 2024 guidelines, showed similar rates of all-cause mortality and major postoperative complications. However, patients randomized to the continuation strategy had more episodes of intraoperative hypotension, and these episodes lasted longer.³³ Furthermore, a posterior 2025 meta-analysis of five RCTs on 10,773 patients confirmed that discontinuation of

Table 3 Association between preoperative ACEI/ARB management strategies and in-hospital mortality.

	n	In-hospital mortality (%)	Crude OR (95% CI)	Adjusted OR (95% CI)	ATE	p-value
Continuing	65	21.5	Reference	Reference	−0.034	0.581
Withholding	69	15.9	0.69 (0.29 – 1.66)	0.78 (0.29 – 2.07)	(−0.153 – 0.086)	

95% CI, 95% Confidence Interval; ATE, Average Treatment Effect. Continuing was defined as treatment until the day before surgery. Withholding was defined as cessation 48 hours or more before surgery. The OR derived from logistic regression models. The multivariable logistic regression model was adjusted for age, preoperative hemoglobin levels, hypertension status, left ventricular ejection fraction < 35%, chronic obstructive pulmonary disease, surgical urgency, and cardiopulmonary bypass time. The ATE (risk difference) was estimated using an adjusted augmented inverse probability weighting model. The p-value refers to the ATE estimate.

ACEIs/ARB likely reduces perioperative hypotension but suggested little to no difference in major clinical outcomes, including AKI, AHF, MI, stroke, and arrhythmias.³⁴

Regarding the timing for resuming these agents postoperatively, some observational data suggest that withholding these medications for more than 48h is associated with increased mortality risk in non-cardiac surgery.³⁵ However, in cardiac surgery, withholding these medications for 48h postoperatively is part of the KDIGO bundle to reduce AKI risk, as shown in several RCTs in cardiac surgery.^{36,37} This patient-centered rationale, which also emphasizes communication with anesthesiologists when favoring continuation, could similarly apply to high-risk cardiac surgery patients, given the shared concerns for hemodynamic instability and the need to balance short-term risks against long-term benefits for underlying conditions.

Our study has some limitations that should be considered. While our findings provide evidence for high-risk patients (CCS ≥ 4), their extrapolation to lower-risk groups warrants cautious interpretation. Significant baseline imbalances persisted despite the selection of a homogeneous cohort of high-risk patients. To account for these differences, a multi-variable adjustment for relevant covariates was performed to minimize residual confounding. Additionally, AIPW was used as a doubly robust estimation method to address potential confounding by indication. As with most observational studies, we cannot exclude the possibility of residual unmeasured confounding. Specifically, the study lacks data on several clinically meaningful variables, such as Major Adverse Kidney Events (MAKE) at 30 and 90 days (including persistent AKI at 30 or 90 days), intraoperative Mean Arterial Pressure (MAP) nadir, fluid balance, glycemic control, and nephrotoxic exposure. Furthermore, information regarding specific ACEI or ARB agents, including dose, duration, and indication, was not available. Finally, the preoperative ACEI/ARB management strategy decision was clinician driven. This reflects ongoing clinical controversy and may also introduce confounding by indication, as unstable patients are more susceptible to treatment interruption, potentially biasing results in favor of the continuation group. Center-level clustering was not accounted for, leaving potential residual confounding from institutional variability.

The study's strengths include its prospective design and the use of pragmatic, real-world data from a high-risk cardiac surgery population. Methodological rigor is further underpinned by the adoption of a consensus AKI definition based on Kidney Disease: Improving Global Outcomes (KDIGO) criteria, while an international multicenter collaboration across 14 cardiothoracic hospitals reinforces broad generalizability.

Future research could investigate the optimal timing for withholding ACEI/ARB in different patient subgroups and explore the potential benefits and risks of this strategy. Randomized controlled trials would be valuable for providing definitive evidence on the impact of ACEI/ARB withholding on cardiac surgery-associated AKI outcomes.

Conclusion

In this multicenter prospective cohort study of high-risk cardiac surgery patients, we observed an association between

preoperative ACEI/ARB use and a higher incidence of stage 2 or 3 AKI. Further research is warranted to provide evidence-based guidelines for the management of these patients.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

AI assistance disclosure

Selected paragraphs throughout the manuscript were revised for linguistic polishing using Google AI Mode (Gemini). All AI-generated suggestions were critically reviewed and verified by the authors, who take full responsibility for the final content.

Funding

This work was funded by departmental resources (Department of Anesthesiology & Critical Care Medicine. Clínica Universidad de Navarra, Pamplona, Spain).

Conflicts of interest

The authors declare no have conflicts of interest.

Acknowledgements

Spanish Perioperative Cardiac Surgery Research Group. Marc VIVES (1), Angel CANDELA (2), Pablo MONEDERO (1), David NAGORE (3), Eduardo TAMAYO (4), Julian Alvarez (5), Esther MENDEZ (6), Alberto PASQUALETTO (6), Tomas MON (6), Rafael PITA (7), Maria Angela VARELA (7), Carlos ESTEVA (7), Miguel Angel PEREIRA (7), Juan SANCHEZ (8), Maria Angeles RODRIGUEZ (8), Alvaro GARCIA (9), Paula CARMONA (10), Marta LÓPEZ (10), Azucena PAJARES (10), Rosario VICENTE (10), Rosa APARICIO (10), Isabel GRAGERA (11), Enrique CALDERON (12), Jose Miguel MARCOS (13), Laura GÓMEZ (14), Jose Maria RODRIGUEZ (15), Adolfo MATILLA (15), Ana MEDINA (16), Alberto HERNÁNDEZ (17), Laura MORALES (18), Luis SANTANA (18), Estefania GARCIA (19), Santiago MONTE-SINOS (20), Pedro MUÑOZ (21), Blanca BRAVO (22), Julian ALVAREZ (23), Verónica BLANCO (24), Nuria GARCIA-FERNÁNDEZ (25), Daniel VILLA (25), Jose Maria MORA (25), Gemma ECHARRI (1), Raquel CALLEJAS (1).

1) Department of Anesthesiology & Critical Care Medicine. Clínica Universidad de Navarra, Pamplona, Spain.

2) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario Ramón y Cajal, Madrid, Spain.

3) Department of Anaesthesia & Perioperative Medicine. Grupo Quirón-Policlinica Guipuzcoa, San Sebastian, Spain.

4) Department of Anesthesiology & Perioperative Medicine. Hospital Clínico Universitario de Valladolid, Spain.

5) Department of Anesthesiology & Perioperative Medicine. Complejo Hospitalario Universitario de Santiago de Compostela, Spain.

6) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario de Bellvitge, Barcelona, Spain.

7) Department of Anesthesiology & Perioperative Medicine. Hospital Alvaro Cunqueiro Vigo, Vigo, Spain.

8) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario Central de Asturias, Oviedo, Spain.

9) Department of Anesthesiology & Perioperative Medicine. Hospital Universitari Vall d'Hebron, Barcelona, Spain.

10) Department of Anesthesiology & Perioperative Medicine. Hospital Universitari i Politècnic La Fe, Valencia, Spain.

11) Department of Anesthesiology & Perioperative Medicine. Hospital Infanta Cristina, Badajoz, Spain.

12) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario Puerta de Mar, Cádiz, Spain.

13) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario de León, Spain.

14) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario Virgen del Rocío de Sevilla, Spain.

15) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario de Salamanca, Salamanca, Spain.

16) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario de Málaga, Spain.

17) Department of Anesthesiology & Perioperative Medicine. Policlínica Ibiza Hospital, Ibiza, Spain.

18) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario Dr Negrín. Las Palmas de Gran Canaria, Spain.

19) Department of Anesthesiology & Perioperative Medicine. Hospital Clínico Universitario de Valencia, Spain.

20) Department of Anesthesiology & Perioperative Medicine. Centro Medico Teknon Barcelona, Spain.

21) Department of Anesthesiology & Perioperative Medicine. Hospital 12 de Octubre, Madrid, Spain.

22) Department of Anesthesiology & Perioperative Medicine. Hospital de Cruces de Bizkaia. Bilbao, Spain.

23) Department of Anesthesiology & Perioperative Medicine. Hospital Clínico Universitario de Santiago, Spain.

24) Department of Anesthesiology & Perioperative Medicine. Hospital Universitario Virgen de la Macarena, Sevilla, Spain.

25) Department of Nephrology. Clínica Universidad de Navarra, Pamplona, Spain.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjane.2026.844786.

Associate Editor

André Prato Schmidt

References

1. Vives M, Hernandez A, Parramon F, et al. Acute kidney injury after cardiac surgery: prevalence, impact and management challenges. *Int J Nephrol Renovasc Dis.* 2019;12:153–66.
2. Thakar CV, Arrigain S, Worley S, Yared JP, Paganini EP. A clinical score to predict acute renal failure after cardiac surgery. *J Am Soc Nephrol.* 2005;16:162–8.
3. Demirjian S, Schold JD, Navia J, et al. Predictive models for acute kidney injury following cardiac surgery. *Am J Kidney Dis.* 2012;59:382–9.
4. Vives M, Monedero P, Perez-Valdivieso JR, et al. External validation and comparison of three scores to predict renal replacement therapy after cardiac surgery: a multicenter cohort. *Int J Artif Organs.* 2011;34:329–38.
5. Pintar T, Collard CD. The systemic inflammatory response to cardiopulmonary bypass. *Anesthesiol Clin North Am.* 2003;21:453–64.
6. Larmann J, Theilmeier G. Inflammatory response to cardiac surgery: cardiopulmonary bypass versus non-cardiopulmonary bypass surgery. *Best Pract Res Clin Anaesthesiol.* 2004;18:425–38.
7. Nagore D, Candela A, Bürge M, et al. Hydroxyethyl starch and acute kidney injury in high-risk patients undergoing cardiac surgery: a prospective multicenter study. *J Clin Anesth.* 2021;73:110367.
8. Zhou H, Xie J, Zheng Z, Ooi OC, Luo H. Effect of Renin-Angiotensin System Inhibitors on Acute Kidney Injury Among Patients Undergoing Cardiac Surgery: A Review and Meta-Analysis. *Semin Thorac Cardiovasc Surg.* 2021;33:1014–22.
9. Jeppsson A, Rocca B, Hansson EC, et al. 2024 EACTS Guidelines on perioperative medication in adult cardiac surgery. *Eur J Cardiothorac Surg.* 2024;67:ezae355.
10. Antoniaki DT, Walters RW, Alla VM. Impact of Renin-Angiotensin System Blockers on Mortality in Veterans Undergoing Cardiac Surgery. *J Am Heart Assoc.* 2021;10:e019731.
11. Seese L, Sultan I, Wang Y, et al. The effect of angiotensin-converting enzyme inhibitor exposure on coronary artery bypass grafting. *J Card Surg.* 2020;35:58–65.
12. Benedetto U, Sciarretta S, Roscitano A, et al. Preoperative angiotensin-converting 12. enzyme inhibitors and acute kidney injury after coronary artery bypass grafting. *Ann Thorac Surg.* 2008;86:1160–5.
13. Zhang Y, Ma L. Effect of preoperative angiotensin-converting enzyme inhibitor on the outcome of coronary artery bypass graft surgery. *Eur J Cardiothorac Surg.* 2015;47:788–95.
14. Dag O, Kaygin MA, Aydin A, et al. Is administration of preoperative angiotensin-converting enzyme inhibitors important for renal protection after cardiac surgery? *Ren Fail.* 2013;35:754–60.
15. Tempe DK, Hasija S. Con: does preoperative discontinuation of angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers reduce postoperative acute kidney injury? *J Cardiothorac Vasc Anesth.* 2020;34:2836–8.
16. Milne B, Gilbey T, Ostermann M, Kunst G. Pro: We Should Stop ACE Inhibitors Early Before Cardiac Surgery to Prevent Postoperative Acute Kidney Injury. *J Cardiothorac Vasc Anesth.* 2020;34:2832–5.
17. Drenger B, Fontes ML, Miao Y, et al. Patterns of use of perioperative angiotensin-converting enzyme inhibitors in coronary artery bypass graft surgery with cardiopulmonary bypass: effects on in-hospital morbidity and mortality. *Circulation.* 2012;126:261–9.
18. Argenziano M, Chen JM, Choudhri AF, et al. Management of vasodilatory shock after cardiac surgery: identification of predisposing factors and use of a novel pressor agent. *J Thorac Cardiovasc Surg.* 1998;116:973–80.
19. Pigott DW, Nagle C, Allman K, Westaby S, Evans RD. Effect of omitting regular ACE inhibitor medication before cardiac

- surgery on haemodynamic variables and vasoactive drug requirements. *Br J Anaesth*. 1999;83:715–20.
20. Tsuyuki RT, McKelvie RS, Arnold JM, et al. Acute precipitants of congestive heart failure exacerbations. *Arch Intern Med*. 2001;161:2337–42.
 21. Radaelli G, Bodanese LC, Guaragna JC, et al. The use of inhibitors of angiotensin-converting enzyme and its relation to events in the postoperative period of CABG. *Rev Bras Cir Cardiovasc*. 2011;26:373–9.
 22. Bakris GL, Weir MR. Angiotensin-converting enzyme inhibitor-associated elevations in serum creatinine: is this a cause for concern? *Arch Intern Med*. 2000;160:685–93.
 23. Yacoub R, Patel N, Lohr JW, Rajagopalan S, Nader N, Arora P. Acute kidney injury and death associated with renin angiotensin system blockade in cardiothoracic surgery: a meta-analysis of observational studies. *Am J Kidney Dis*. 2013;62:1077–86.
 24. Arora P, Rajagopalam S, Ranjan R, et al. Preoperative use of angiotensin-converting enzyme inhibitors/angiotensin receptor blockers is associated with increased risk for acute kidney injury after cardiovascular surgery. *Clin J Am Soc Nephrol*. 2008;3:1266–73.
 25. Chou YH, Huang TM, Wu VC, Chen WS, Wang CH, Chou NK, Chiang WC, Chu TS, Lin SL. National Taiwan University Study Group on Acute Renal F: Associations between preoperative continuation of renin-angiotensin system inhibitor and cardiac surgery-associated acute kidney injury: a propensity score-matching analysis. *J Nephrol*. 2019;32:957–66.
 26. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract*. 2012;120:c179-84.
 27. Coca SG, Garg AX, Swaminathan M, et al. Preoperative angiotensin-converting enzyme inhibitors and angiotensin receptor blocker use and acute kidney injury in patients undergoing cardiac surgery. *Nephrol Dial Transplant*. 2013;28:2787–99.
 28. De Bono JA, Conte SM, Newcomb AE. Effects of Preoperative Angiotensin-Converting Enzyme Inhibitor Therapy on Postoperative Renal Function in Cardiac Surgery. *Heart Lung Circ*. 2020;29:1656–67. <https://doi.org/10.1016/j.hlc.2020.06.003>.
 29. Guilleminot P, Andrei S, Nguyen M, et al. Pre-operative maintenance of angiotensin-converting enzyme inhibitors is not associated with acute kidney injury in cardiac surgery patients with cardio-pulmonary bypass: a propensity-matched multicentric analysis. *Front Pharmacol*. 2024;15:1343647.
 30. van Diepen S, Norris CM, Zheng Y, et al. Comparison of Angiotensin-Converting Enzyme Inhibitor and Angiotensin Receptor Blocker Management Strategies Before Cardiac Surgery: A Pilot Randomized Controlled Registry Trial. *J Am Heart Assoc*. 2018;7:e009917.
 31. Whiting P, Morden A, Tomlinson LA, et al. What are the risks and benefits of temporarily discontinuing medications to prevent acute kidney injury? A systematic review and meta-analysis. *BMJ Open*. 2017;7:e012674.
 32. Writing Committee MembersThompson A, Fleischmann KE, Smilowitz NR, et al. 2024 AHA/ACC/ACS/ASNC/HRS/SCA/SCCT/SCMR/SVM Guideline for Perioperative Cardiovascular Management for Noncardiac Surgery: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2024;84:1869–969.
 33. Legrand M, Falcone J, Cholley B, et al. Continuation vs Discontinuation of Renin-Angiotensin System Inhibitors Before Major Noncardiac Surgery: The Stop-or-Not Randomized Clinical Trial. *JAMA*. 2024;332:970–8.
 34. Bena Nnang JY, Tembi TBT, Fodop SGJ, Mekontso JGK, Esene IN. Discontinuation of renin-angiotensin system inhibitors before non-cardiac surgery: a meta-analysis. *Eur Heart J*. 2025.
 35. Smilowitz NR, Berger JS. Perioperative Cardiovascular Risk Assessment and Management for Noncardiac Surgery: A Review. *JAMA*. 2020;324:279–90.
 36. Meersch M, Schmidt C, Hoffmeier A, et al. Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high-risk patients identified by biomarkers: the PrevAKI randomized controlled trial. *Intensive Care Med*. 2017;43:1551–61.
 37. Zarbock A, Kullmar M, Ostermann M, et al. Prevention of Cardiac Surgery-Associated Acute Kidney Injury by Implementing the KDIGO Guidelines in High-Risk Patients Identified by Biomarkers: The PrevAKI-Multicenter Randomized Controlled Trial. *Anesth Analg*. 2021;133:292–302.

REVIEW ARTICLE

Artificial intelligence in the prediction of intraoperative red blood cell transfusion in cardiac surgery: a systematic review and diagnostic test accuracy meta-analysis



Filipe Giordano Valério ^{a,*}, Amanda Carneiro Rodrigues ^b,
Davi Ricardo Soares Gama de Amorim ^c, Roberta Esterque Cantarino ^a,
Glauro Martins de Araujo ^a, Jimmy Yusuf ^a, Leticia Vianna Leal Soares Bessa ^a,
Millena Mendonça Andrade Paes Leme ^a, Marcella Freire de Campos Euzebio ^a,
Flávio Luiz Seixas ^d, Alexandra Rezende Assad ^e, Luis Antonio dos Santos Diego ^f

^a Universidade Federal Fluminense (UFF), Niterói, RJ, Brazil

^b Faculdade de Ciências Médicas da Santa Casa de São Paulo, (FCMSCSP), São Paulo, SP, Brazil

^c Universidade de Pernambuco (UPE), Recife, PE, Brazil

^d Universidade Federal Fluminense, Institute of Computing, Niterói, RJ, Brazil

^e Universidade Federal do Rio de Janeiro, Department of General and Specialized Surgery, Postgraduate Program in Surgical Sciences, Rio de Janeiro, RJ, Brazil

^f Universidade Federal Fluminense (UFF), Faculdade de Medicina, Department of Surgery, Niterói, RJ, Brazil

Received 10 October 2025; accepted 14 June 2026

Available online 25 June 2026

KEYWORDS

Algorithms;
Artificial intelligence;
Blood transfusion;
Cardiac surgical
procedures;
Erythrocyte
transfusion;
Prediction algorithms

Abstract

Background: Red Blood Cell (RBC) transfusion in cardiac surgery is associated with risks. Conventional prediction scores lack accuracy, conflicting with Patient Blood Management (PBM) principles. Artificial Intelligence (AI) offers a potential avenue for developing precise, data-driven predictive models to enhance clinical decision-making and optimize patient outcomes.

Methods: A systematic review and meta-analysis of diagnostic test accuracy studies was conducted following PRISMA guidelines, searching PubMed, Embase, and Cochrane Library databases until April 2025. Studies developing AI models to predict intraoperative Red Blood Cell (RBC) transfusions in adult cardiac surgery were included. Pooled sensitivity, specificity, and Area Under the receiver operating Characteristic Curve (AUC) were calculated using a bivariate random-effects model. To summarize overall diagnostic performance, Summary Receiver Operating Characteristic curves were generated.

Results: Five studies encompassing 3,063 patients were analyzed in the meta-analysis. AI models demonstrated high pooled specificity, ranging from 86.3% (95% CI 82.8%–89.1%) to 93.6% (95%

* Corresponding author.

E-mail: filipegv@id.uff.br (F.G. Valério).

CI 84.3%–97.6%), whereas pooled sensitivity ranged from 50.3% (95% CI 20.9%–79.6%) to 55.7% (95% CI 33.8%–75.6%). AUC values ranged between 0.793 (95% CI 0.634–0.899) and 0.892 (95% CI 0.740–0.943). Preoperative hemoglobin levels and patient age were consistently identified as clinical predictors for intraoperative RBC transfusion.

Conclusion: This systematic review and meta-analysis suggests that AI models may have potential for predicting intraoperative RBC transfusion in cardiac surgery, with consistently high specificity but only moderate sensitivity. However, given the low certainty of evidence, substantial heterogeneity, and reliance on non-standardized transfusion practices, these findings should be interpreted cautiously.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Intraoperative Red Blood Cell (RBC) transfusion is common in cardiac surgery, a context often complicated by hypovolemia and impaired tissue perfusion. However, transfusions carry risks, including infection, lung injury, transfusion-related immunomodulation, and increased morbidity and mortality.^{1–6} Patient Blood Management (PBM) has emerged as a multimodal, evidence-based approach to optimize transfusion outcomes and enhance patient safety. PBM is structured around three pillars: 1) Optimizing red blood cell mass via screening, diagnosis, and treatment of anemia and iron deficiency; 2) Minimizing surgical and iatrogenic blood loss while managing bleeding and coagulopathies perioperatively; and 3) Optimizing physiological tolerance to anemia by improving oxygen delivery and reducing oxygen demand.^{7–9} This definition also underscores the need for shared decision-making, actively involving patients in the management of their own blood.

Despite increasing adoption of PBM principles, predicting intraoperative RBC transfusion in cardiac surgery remains challenging. Clinicians largely depend on general risk models, such as the STS score and EuroSCORE, which are designed to estimate mortality and major morbidity rather than hematological requirements.^{10,11} Although transfusion-specific tools, including “Transfusion Risk and Clinical Knowledge” (TRACK) and “Transfusion Risk Understanding Scoring Tool” (TRUST), have been developed to address this limitation, their predictive performance remains limited, resulting in reliance on clinical judgment and the team’s prior experience.^{2,12} The absence of reliable prediction methods contributes to blood component overuse and increasing perioperative risk, in conflict with PBM principles.^{2,4,13}

In this scenario, Artificial Intelligence (AI) emerges as a promising strategy to overcome these limitations and improve perioperative blood management. AI enables the analysis of high-dimensional clinical data to develop predictive models, with approaches varying in methodology and complexity.^{1,4,13} These approaches aim to support clinical decision-making, reduce unnecessary transfusions, mitigate patient risk, and optimize the use of blood products and hospital resources. Accordingly, this systematic review and meta-analysis evaluates the statistical robustness of different AI models for predicting intraoperative RBC transfusion in cardiac surgery.

Methods

Study design and protocol registration

This systematic review and meta-analysis was conducted in accordance with the guidelines of the Cochrane Collaboration^{14,15} and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.¹⁶ The protocol was registered on September 16, 2025, before data extraction, in the International Prospective Register of Systematic reviews (PROSPERO) under registration CRD420251146515.

Research question

The research question was elaborated using the PIT framework for diagnostic test accuracy meta-analysis, as recommended by the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (v2.0): Population (P) – patients undergoing cardiac surgery; Index test (I) – use of AI algorithms; Target condition (T) – Occurrence of RBC transfusion in the intraoperative period (described as “intraoperative” in four out of five articles, and “from the initiation of surgery until discharge to the intensive care unit” in Cunha et al.).^{2,14}

Data sources

Searches were systematically conducted in PubMed, Embase, and Cochrane Library, up to April 15, 2025.

Search methods

The first author (F.V.) developed search strategies (Supplementary Table 1). A search of the reference lists of the studies included was conducted to ensure that no other relevant studies were overlooked.

Eligibility criteria

This systematic review included studies evaluating the predictive accuracy of AI models for intraoperative RBC transfusion in adult patients undergoing cardiac surgery. Eligible studies reported the target population, applied AI-based index tests, and assessed intraoperative RBC transfusion prediction accuracy. Studies involving inappropriate populations, case reports, ongoing trials, conference abstracts,

editorials, letters, comments, and studies without final results were also excluded.

Study selection

Retrieved articles were imported into the Rayyan systematic review platform,¹⁷ deduplicated, and independently screened by three reviewers (F.V., L.B., J.Y.). Full texts were subsequently assessed when eligibility was unclear, and studies meeting the criteria were included (Fig. 1). Disagreements were resolved by a supervising reviewer (L.D.).

Outcomes

Outcomes assessed in this review were the pooled sensitivity, specificity, and Area Under the Curve (AUC) for AI algorithms eligible for statistical analysis. All AI models assessed in two or more articles were included in this evaluation. To ensure stable pooled estimates and sufficient statistical power, models evaluated in four or more studies are described

individually. Models assessed in fewer than four studies are summarized under “Other models” in the Results, with their Summary Receiver Operating Characteristic (SROC) curves and supplemental graphs provided in the Supplementary Material. Due to the limited sample sizes, results for these models should be interpreted with caution.

Data extraction

Three blinded reviewers (J.Y., G.A., and A.R.) independently extracted data from each study, including authorship, year of publication, country, and patient characteristics (sample size per group, age, and sex), as well as predictors, AI models, type of surgery, hemoglobin, hematocrit, and diagnostic performance metrics (sensitivity, specificity, accuracy, and precision). Data were collected using Google Sheets and cross-checked for accuracy (Table 1). Corresponding authors were contacted by email to obtain missing data; no responses were received.

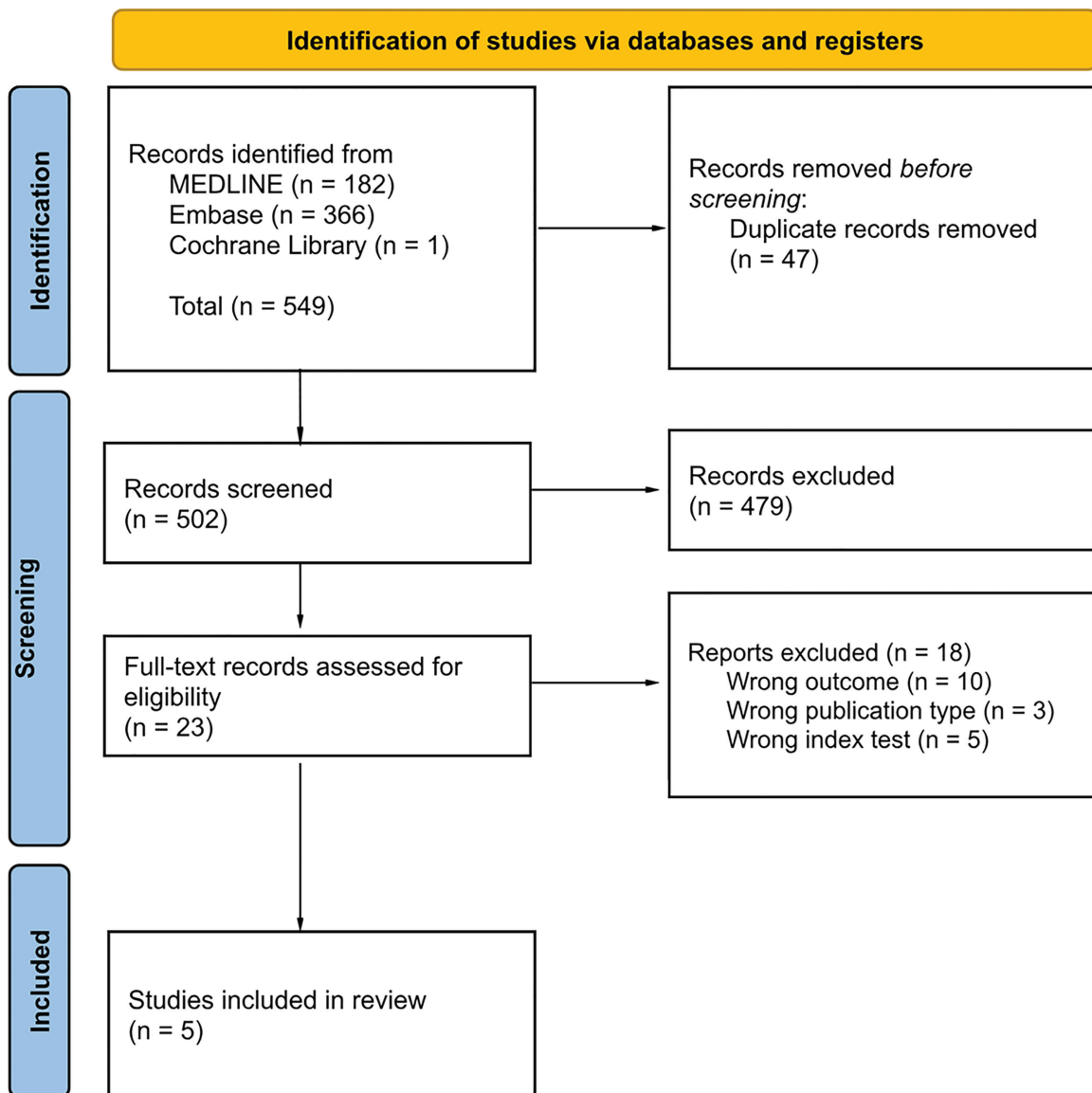


Figure 1 PRISMA flow diagram of study screening and selection.

Table 1 Baseline characteristics of studies.

Study	Index tests	Types of surgery	Number of surgeries		Transfused patients n (%)	Female sex (%)	Age (years)	Hemoglobin (g/dL)		Hematocrit (%)		Red Blood Cells (10 ¹² /L)		Platelets (10 ⁹ /L)		Prothrombin time (seconds)	
			Training set	Testing set				Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
Sun, 2024 ⁶	Extreme Gradient Boosting, Extreme Gradient Boosting, Extra Trees, Logistic Regression, Categorical Boosting, Adaptive Boosting, Linear Discriminant Analysis, Random Forest, Decision Tree, K Nearest Neighbor.	Minimally invasive direct coronary artery bypass (MIDCAB)	613	153	107 (13.97%)	221 (28.85%)	65.2 ± 9.5	13.1 ± 1.72	39.5 ± 5.1	4.28 ± 0.54	215.5 ± 56	12.91 ± 0.92					
			473	204	166 (24.52%)	326 (48.15%)	57.3 ± 13.37	13.12 ± 1.82	39.36 ± 5.08	4.35 ± 0.6	186.50 ± 54.97	11.635 ± 0.97					
Liu, 2021 ⁴	Categorical Boosting, Light Gradient Boosting, Extreme Gradient Boosting, Extra Trees, Logistic Regression, Linear Discriminant Analysis, Random Forest, Adaptive Boosting, Naive Bayes, K Nearest Neighbor, Decision Tree, Quadratic Discriminant Analysis	Mitral and tricuspid valve surgery	491	211	267 (38.03%)	242 (34.47%)	55.76 ± 12.3	13.59 ± 1.81	41 ± 5	4.6 ± 0.6	169.19 ± 61.03	12.1 ± 3.45					
			491	211	267 (38.03%)	242 (34.47%)	55.76 ± 12.3	13.59 ± 1.81	41 ± 5	4.6 ± 0.6	169.19 ± 61.03	12.1 ± 3.45					
Chen, 2024 ³	Logistic Regression, Categorical Boosting, Extra Trees, Naive Bayes (Gaussian and Bernoulli), Multilayer Perceptron, Adaptive Boosting, Random Forest, Light Gradient Boosting, Support Vector Machine, Extreme Gradient Boosting, Decision Tree	Surgical and transcatheter aortic valve replacement and implantation	396	99	284 (57.37%)	196 (39.6%)	56.66 ± 14.17	11.3 ± 2.17	33.9 ± 6.5	-	-	-					
			396	99	284 (57.37%)	196 (39.6%)	56.66 ± 14.17	11.3 ± 2.17	33.9 ± 6.5	-	-	-					
Cunha, 2023 ²	Logistic Regression, Multilayer Perceptron, Random Forest, Support Vector Machine, TRACK, TRUST	Aortic surgery, valve surgery, coronary artery bypass grafting, combined surgery, and other cardiac surgeries (not specified)	338	85	102 (24.11%)	199 (47.04%)	60.1 ± 11.4	13.65 ± 1.39	41.12 ± 4.73	-	172 ± 56.4	13.61 ± 4.68					
			338	85	102 (24.11%)	199 (47.04%)	60.1 ± 11.4	13.65 ± 1.39	41.12 ± 4.73	-	172 ± 56.4	13.61 ± 4.68					
Zhou, 2024 ¹	Extreme Gradient Boosting, Random Forest, Decision Tree, Categorical Boosting, Support Vector Machine, Logistic Regression	Valve replacement surgery	338	85	102 (24.11%)	199 (47.04%)	60.1 ± 11.4	13.65 ± 1.39	41.12 ± 4.73	-	172 ± 56.4	13.61 ± 4.68					
			338	85	102 (24.11%)	199 (47.04%)	60.1 ± 11.4	13.65 ± 1.39	41.12 ± 4.73	-	172 ± 56.4	13.61 ± 4.68					

TRACK, Transfusion Risk and Clinical Knowledge; TRUST, Transfusion Risk Understanding Scoring Tool.

Quality assessment and risk of bias

The methodological quality of each study was verified using the PROBAST-AI¹⁸ tool by two blinded authors (R.C. and G.A.). This tool covers four domains, as follows: patients, predictors, outcome, and analysis. The first author (F.V.) resolved any discrepancies. Two authors (R.C., G.A.) independently assessed studies for quality and certainty of evidence in accordance with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool.¹⁹

Statistical analysis

As per the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (v2.0),¹⁴ the components of the 2×2 contingency table – True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) were calculated using the sensitivity and specificity values reported for the testing set in the original study. The number of TP and TN cases were calculated as:¹ TP = (Total transfused patients) × Sensitivity;² TN = (Total non-transfused patients) × Specificity. Subsequently, the number of False Positives (FP) and False Negatives (FN) were derived:³ FP = (Total non-transfused patients) – TN;⁴ FN = (Total transfused patients) – TP (Supplementary Table 2), rounding results to the nearest integer. When the exact number of transfused patients in the testing set was not reported, it was assumed that the transfused-to-non-transfused ratio matched that of the randomly assigned training set. This assumption applied to two studies, Liu et al.⁴ and Sun et al.⁶ As both studies showed high stability and negligible discrepancies in comparative validation and sensitivity analyses, this assumption likely did not bias the pooled estimates.

For studies reporting medians and interquartile ranges, Luo's method²⁰ was applied to approximate the mean, and Wan's method²¹ to estimate the SD. Statistical analysis was performed in R (version 4.4.1)²² utilizing the Mada²³ and meta packages.²⁴ In accordance with the Cochrane Handbook,¹⁴ a hierarchical approach was employed to synthesize model performance while accounting for the inherent correlation between sensitivity and specificity.

A bivariate random-effects meta-analysis was implemented using the Reitsma function in the Mada R package, using restricted maximum likelihood estimation. As described by Reitsma et al.,²⁵ this model analyzes logit-transformed sensitivity and false positive rates. This hierarchical framework accounts for two levels of variability: within-study sampling error and between-study heterogeneity. This bivariate approach is mathematically equivalent to the Hierarchical Summary Receiver Operating Characteristic (HSROC) model proposed by Rutter and Gatsonis²⁶ in the absence of covariates,²⁷ which is the case of this meta-analysis. The threshold effect was evaluated primarily through bivariate modeling, as recommended by Reitsma et al.²⁵

Heterogeneity was quantified using the Zhou and Dendukuri approach²⁸ for bivariate I^2 , which represents the proportion of total variation attributable to between-study inconsistency. We also examined the between-study variance components (τ) for both sensitivity and the false positive rate. Diagnostic performance was visualized using SROC curves. A prediction region could not be traced due to the low number of included studies.^{14,15} The Area Under the

Curve (AUC) was derived directly from the parameters of the bivariate model as a global measure of accuracy. Small-study effects and publication bias were assessed using Deeks' asymmetry test on the diagnostic odds ratio.¹⁴

To assess potential bias from rounding methods and stratification assumptions, a comparative analysis was conducted (Supplementary Material). Diagnostic parameters were recalculated from reconstructed 2×2 tables using three rounding strategies (floor, ceiling, and nearest integer) and compared with the metrics originally reported by the authors. Although discrepancies remained negligible across all rounding approaches, rounding to the nearest integer consistently provided the closest approximation to the reported values.

Across four out of five included studies, these discrepancies are negligible. However, the study by Zhou et al.¹ exhibited substantial discrepancies despite the availability of explicit transfused and non-transfused patient counts for the testing set. Therefore, a supplementary sensitivity analysis was conducted by excluding this study, which did not significantly change results (Supplementary Fig. 1).

Results

Study selection

The systematic literature search yielded 549 records (Fig. 1). After the removal of 47 duplicates, 502 unique articles were screened, which led to the exclusion of 479 studies. The full texts of the remaining 23 articles were assessed for eligibility. Of these, 18 studies were excluded for reasons including the use of incorrect index tests ($n = 5$), publication as congress abstracts or errata ($n = 3$) or reporting of irrelevant outcomes ($n = 10$). Ultimately, five studies were included in the meta-analysis.^{1-4,6}

Baseline characteristics

This analysis is based on five retrospective studies, encompassing a total cohort of 3,063 patients who underwent various cardiac surgical procedures, including aortic, tricuspid, and mitral valve replacement or implantation, as well as Minimally Invasive Direct Coronary Artery Bypass (MIDCAB).

The pooled weighted mean age of the patient cohort was 59.15 ± 12.67 years, with a sex distribution of 1,879 (61.35%) male and 1,184 (38.65%) female patients. A total of 926 patients (30.23%) received RBC transfusion during the intraoperative period; the proportion of transfused patients ranged from 13.97% to 57.37% among studies. Table 1 presents which AI models were evaluated in each included study and detailed baseline characteristics of patients.

Risk of bias and certainty of evidence

The methodological quality was consistent among the studies, as detailed in the PROBAST-AI¹⁸ assessment (Supplementary Fig. 2). The risk of bias was low for the participants in all studies except for Chen et al.³ Predictors and outcome domains were rated as low risk of bias across every study. The analysis domain presented a high risk of bias for all

studies, due to lack of model calibration and clinical utility assessment.

According to the GRADE analysis¹⁹ (Supplementary Fig. 3), the overall certainty of evidence for the test's diagnostic accuracy was rated as low. This determination resulted from downgrading the evidence due to two critical limitations. A serious risk of bias was identified, stemming from the subjective nature of the reference standard. Additionally, there were serious concerns with inconsistency, driven by significant performance heterogeneity among the AI algorithms evaluated.

Best-performing model included studies and key predictive features

To determine the key predictors for intraoperative RBC transfusion, four of the five studies^{1,3,4,6} employed SHAP (SHapley Additive exPlanations) analysis²⁹ on their best-performing model. The remaining study by Cunha et al.² utilized a permutation importance analysis for the same purpose. This technique evaluates feature importance by randomly permuting its values; the resulting change in model performance reflects the degree of reliance on that feature.³⁰

The resulting top-performing models and their complete sets of identified predictors varied across the analyses. The study by Sun et al.⁶ identified Extreme Gradient Boosting as the most accurate model, listing hemoglobin, prothrombin time, Body Mass Index (BMI), coronary heart disease with additional comorbidities, and a history of percutaneous coronary intervention as its key predictors. In contrast, Chen et al.³ found a Logistic Regression algorithm to be superior, with its most essential predictors being sex, hemoglobin, hematocrit, weight, and age. Categorical Boosting was the optimal model

in two different studies, although each highlighted distinct feature sets. The model by Liu et al.⁴ prioritized hematocrit, age, the surgeon responsible for the procedure, weight, and BMI. Meanwhile, the analysis by Zhou et al.¹ identified sex, hemoglobin, ferritin, Alanine Transaminase (ALT), and total bilirubin as significant factors. Lastly, the statistical approach by Cunha et al.² was limited to five predictors, which were, in order of importance: hemoglobin, age, body surface area, cardiopulmonary bypass, reoperation status, and sex, reporting Logistic Regression as its best-performing AI model. Despite the varied models and predictors, preoperative hemoglobin and patient age consistently emerged as highly significant across multiple studies (Table 2).

Performance of AI models

Logistic Regression model

The meta-analysis showed that the Logistic Regression model (Fig. 2a) achieved a pooled sensitivity of 0.557 (95% CI 0.338–0.756) and a pooled specificity of 0.891 (95% CI 0.758–0.956). The overall diagnostic accuracy was reflected by an AUC of 0.820 (95% CI 0.635–0.905). Threshold-related variation was handled through the bivariate random-effects framework, which preserves the two-dimensional nature of the data by jointly analyzing sensitivity and specificity. Within this joint distribution, the model estimated between-study standard deviations for sensitivity (0.954) and specificity (1.037). In the ROC space, the 95% confidence ellipse illustrates the region containing likely combinations of the mean sensitivity and specificity values. The orientation and spread of this ellipse, along with the concave shape of the curve, are consistent with the presence of a threshold effect. As an exploratory finding within the bivariate

Table 2 The most important predictors for the best-performing model in each study.

Study	Best performing model	Type of analysis	Predictors (in order of importance)
Sun, 2024 ⁶	Extreme Gradient Boosting	SHAP	Hb, PT, BMI, coronary heart disease, history of PCI, weight, heart rate, PLT, fibrinogen, creatinine, TG, CRP, NT-proBNP, multiple CABG, APTT, apo-B, BV, RBC, Tn, LDL-C.
Liu, 2021 ⁴	Categorical Boosting	SHAP	HCT, age, name of surgeon, weight, BMI, height, anemia, operation type, Hb, RBC, tricuspid valve procedure, PLT, sex, ASA, ALT, AST, mitral valve procedure, PT, INR, creatinine.
Chen, 2024 ³	Logistic Regression	SHAP	Sex, Hb, HCT, weight, age, MCH, PT, fibrinogen, stature, operation type, GLB, URIC, TBIL, DBIL, HBDH, RDW-SD, TG, A/G, NYHA, IBIL.
Cunha, 2023 ²	Logistic Regression	PI	Hb, age, BSA, CPB, redo, sex.
Zhou, 2024 ¹	Categorical Boosting	SHAP	Sex, Hb, ALT, ferritin, fibrinogen, TBIL, LVEF, albumin, creatinine, APTT, TAST, PLT, folic acid, vitamin B12, TT, HbA1c.

A/G, albumin/globulin; ALT, alanine transaminase; APTT, activated partial thromboplastin time; ASA, American Society of Anesthesiologists; AST, aspartate transaminase; BMI, body mass index; BSA, body surface area; BV, blood volume; CABG, coronary artery bypass grafting; CPB, cardiopulmonary bypass; CRP, C-reactive protein; DBIL, direct bilirubin; GLB, globulin; Hb, hemoglobin; HbA1c, glycated hemoglobin; HBDH, hydroxybutyrate dehydrogenase; HCT, hematocrit; IBIL, indirect bilirubin; INR, international normalized ratio; LDL-C, low-density lipoprotein cholesterol; LVEF, left ventricular ejection fraction; MCH, mean corpuscular hemoglobin; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Association; PCI, percutaneous coronary intervention; PLT, platelet; PI, permutation importance; PT, prothrombin time; RBC, red blood cell; RDW-SD, red cell distribution width-standard deviation; SHAP, Shapley Additive Explanations; TAST, transferrin saturation; TBIL, total bilirubin; TG, triglycerides; Tn, troponin; TT, thrombin time; URIC, uric acid.

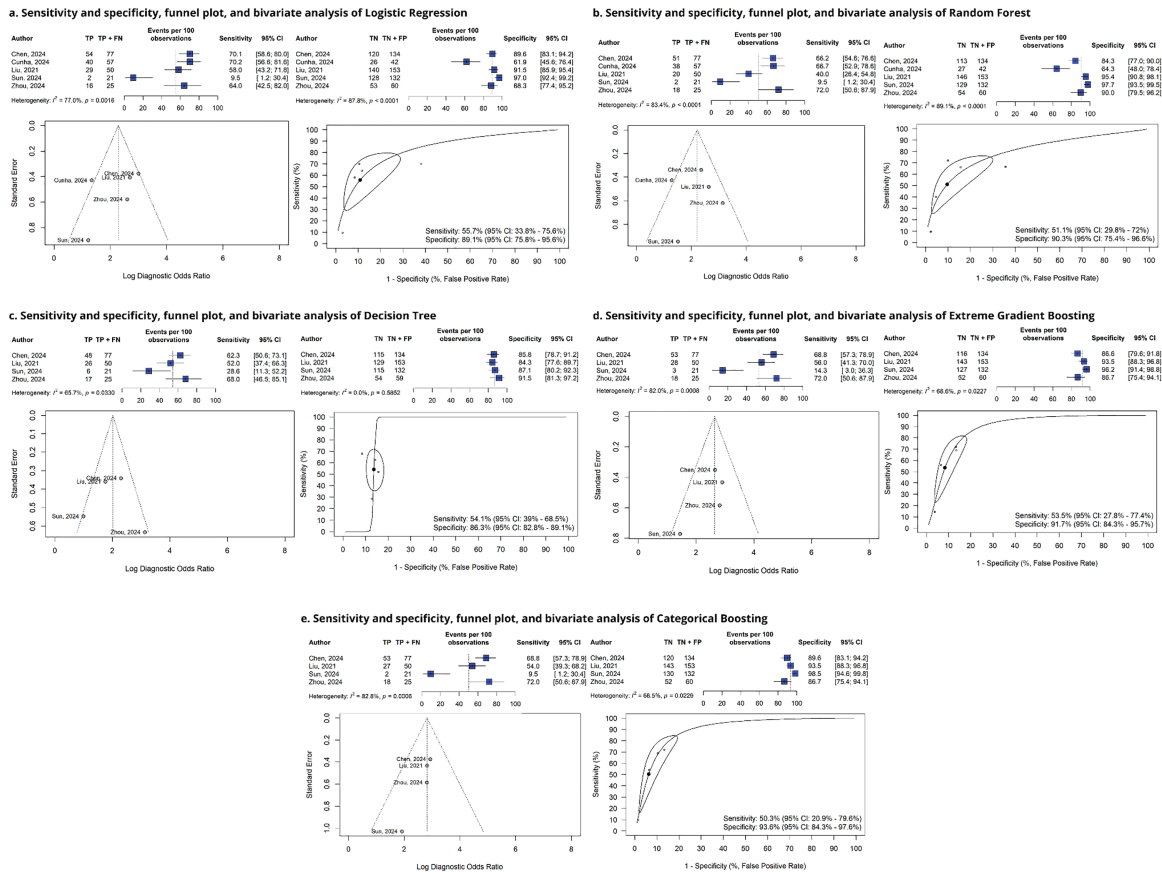


Figure 2 Sensitivity, specificity, funnel plot and bivariate analysis of Logistic Regression (a); Random Forest (b); Decision Tree (c); Gradient Boosting (d); Categorical Boosting (e).

framework, a correlation of 0.833 was noted between the logit-transformed measures. Regarding broader heterogeneity, the Zhou and Dendukuri²⁸ approach estimated a bivariate I^2 of 41.5%, pointing to a moderate level of heterogeneity. Deeks' funnel plot asymmetry test showed no significant asymmetry ($t = -1.36$, $p = 0.268$).

Random Forest

The Random Forest model (Fig. 2b) showed a summary sensitivity of 0.511 (95% CI 0.298–0.720) and specificity of 0.903 (95% CI 0.754–0.966), with an AUC of 0.793 (95% CI 0.634–0.899). Within the bivariate joint distribution, the model estimated between-study standard deviations of 0.954 for sensitivity and 1.208 for specificity. In the ROC space, the orientation and spread of the 95% confidence ellipse are consistent with the presence of a threshold effect. As an exploratory finding, a correlation of 0.928 was estimated between the logit-transformed measures. Bivariate I^2 was 32.7%, reflecting moderate heterogeneity. Deeks' funnel plot asymmetry test indicated no significant publication bias ($t = -0.20$, $p = 0.854$).

Decision Tree model

For the Decision Tree model (Fig. 2c) the meta-analysis yielded a pooled sensitivity of 0.541 (95% CI 0.390–0.685) and a pooled specificity of 0.863 (95% CI 0.828–0.891). The overall diagnostic performance was strong, with an AUC of 0.862 (95% CI 0.595–0.887). The model estimated between-study standard deviations of 0.515 for sensitivity and 0.016

for specificity. In the ROC space, the orientation and spread of the 95% confidence ellipse are not consistent with the presence of a threshold effect. As an exploratory finding, a correlation of -1.000 was estimated between the logit-transformed measures. Between-study heterogeneity was minimal, with the Zhou and Dendukuri approach estimating a bivariate I^2 of 0%. Evaluation of small-study effects using Deeks' funnel plot asymmetry test was non-significant ($t = 0.03$, $p = 0.980$).

Extreme Gradient Boosting model

The Extreme Gradient Boosting model (Fig. 2d) showed a pooled sensitivity of 0.535 (95% CI 0.278–0.774) and specificity of 0.917 (95% CI 0.843–0.957), with an AUC of 0.886 (95% CI 0.716–0.929). The model estimated between-study standard deviations of 1.044 for sensitivity and 0.646 for specificity. In the visual ROC space, the orientation and spread of the 95% confidence ellipse are consistent with the presence of a threshold effect. As an exploratory finding, a correlation of 1.000 was estimated between the logit-transformed measures. Bivariate I^2 was 0%, suggesting minimal residual heterogeneity attributable to sampling error and consistent discriminative performance across studies. Deeks' funnel plot asymmetry test indicated no significant publication bias ($t = -1.37$, $p = 0.303$).

Categorical Boosting model

The Categorical Boosting model showed a pooled sensitivity of 0.503 (95% CI 0.209–0.796) and specificity of 0.936 (95%

CI 0.843–0.976), with an AUC of 0.892 (95% CI 0.740–0.943). Within the bivariate joint distribution, the model estimated between-study standard deviations of 1.308 for sensitivity and 0.953 for specificity. In the ROC space, the orientation and spread of the 95% confidence ellipse are consistent with the presence of a threshold effect. As an exploratory finding, a correlation of 1.000 was estimated between the logit-transformed measures. Bivariate I^2 was 0%, suggesting residual heterogeneity is due to sampling error. Deeks' funnel plot asymmetry test was significant ($t = -5.33$, $p = 0.033$), indicating potential publication bias or small-study effects affecting pooled estimate reliability.

Other models

The diagnostic performance of AI architectures evaluated in fewer than four studies is detailed below, with full outputs available in Supplementary Figures 4–11. The Adaptive Boosting ($n = 3$), Gradient Boosting ($n = 3$), Support Vector Machine ($n = 3$), and Gradient Boosting ($n = 3$) demonstrated varied discriminative capabilities. Adaptive Boosting achieved an AUC of 0.867 (95% CI 0.543–0.897) with a pooled sensitivity of 0.494 (95% CI 0.265–0.725) and specificity of 0.871 (95% CI 0.833–0.902); its negligible heterogeneity ($I^2 = 0\%$), moderate threshold correlation ($r = 0.632$), and 95% confidence ellipse suggest that variability was driven by random sampling error. Support Vector Machine showed negligible heterogeneity ($I^2 = 2.1\%$) and an AUC of 0.698 (95% CI 0.658–0.883). The Gradient Boosting model presented a moderate level of inconsistency ($I^2 = 50.9\%$) despite an AUC of 0.815 (95% CI 0.456–0.924). The Extra Trees model ($n = 3$) yielded an AUC of 0.770 (95% CI 0.451–0.939). Other analyzed models were K Nearest Neighbor ($n = 2$; AUC = 0.786), Light Gradient Boosting ($n = 2$; AUC = 0.699), Linear Discriminant Analysis ($n = 2$; AUC = 0.902), and Multilayer Perceptron ($n = 2$; AUC = 0.776).

Discussion

This systematic review and meta-analysis suggests that AI models may have potential for predicting intraoperative RBC transfusion in cardiac surgery; however, this interpretation is limited by low certainty of evidence, substantial heterogeneity, and restricted generalizability. Across five studies with 3,063 patients, models consistently exhibited high specificity (ranging from 86.2% for Decision Tree to 93.6% for Categorical Boosting) and moderate sensitivity (ranging from 50.3% for Categorical Boosting to 55.7% for Logistic Regression). The consistently high specificity observed across models suggests that positive predictions may help identify patients at higher likelihood of receiving transfusion within the studied contexts. However, the moderate-to-low sensitivity indicates that a negative prediction is not enough to rule out the need for blood.³¹ Importantly, the predicted outcome reflects transfusion events determined by local clinical practice rather than a standardized or objective measure of physiological need, which limits clinical interpretability.

Predictive variables

Predictive features for intraoperative RBC transfusion were heterogeneous, but preoperative hemoglobin and patient

age consistently emerged as key risk factors. Other important predictors include sex, hemoglobin, hematocrit, weight, and mean corpuscular hemoglobin, which reflect blood volume; fibrinogen, which reflects coagulation status; and type of surgery, which influences bleeding. The higher prevalence of iron deficiency anemia among females further highlights sex as a relevant risk factor.³²

Comparison with established scores

Compared with intraoperative RBC transfusion risk scores such as TRACK and TRUST, which report AUCs of 0.76–0.80,^{12,33} the discriminative performance of AI models observed in this review appears numerically higher. However, despite the high specificities and AUCs of AI intraoperative RBC transfusion prediction models, some studies report that in less complex surgeries, such as minimally invasive direct coronary artery bypass, reported AUCs were lower (~ 0.73).¹ This may suggest that model accuracy depends on surgical complexity, institutional practices, and data quality.

Although AI models show numerically higher discrimination than conventional scores, the clinical significance of this difference remains uncertain. In surgical practice, enhanced discrimination is meaningful only if it improves bedside decision-making, which has not yet been demonstrated. AUC differences may reflect the ability of AI models to capture non-linear interactions between risk factors, such as age and preoperative hemoglobin. However, this has not translated into a meaningful change in the sensitivity-specificity trade-off, as sensitivity remains comparable to conventional scores.

Heterogeneity

Heterogeneity varied across AI models, reflecting differences in study populations, model design, surgical complexity, and local RBC transfusion practices. Threshold effects in some models suggest that performance variability was partly driven by non-standardized transfusion criteria and institutional decision-making. A concave shape of the SROC curve, a diagonally distributed and an elongated spread of the 95% confidence ellipse are indicative of the presence of threshold effect. In contrast, the Decision Tree model showed low heterogeneity and a sharper SROC curve shape, indicating that classical threshold effects along a common ROC curve were not the main source of between-study variability. This apparent consistency may reflect more stable, rule-based decision structures but should be interpreted cautiously, as low heterogeneity could also result from limited dataset variability. Overall, heterogeneity in AI-based intraoperative RBC transfusion prediction appears influenced by both algorithmic characteristics and clinical context, highlighting the need for standardized transfusion criteria and external validation.

Implications for PBM and resources

These findings suggest a potential exploratory role for AI in perioperative risk stratification. High specificity supports use as a “rule-in” tool for identifying patients at high transfusion risk with a positive result, while low sensitivity limits

reliable exclusion, leaving many cases undetected. However, current evidence is insufficient to support these AI models as practice-changing tools in cardiac surgery.

In recent years, increasing awareness of transfusion-associated risks has prompted a paradigm shift towards PBM and more restrictive transfusion strategies.^{6,8,34-36} However, a gap remains between guidelines and clinical practice. Clinicians often rely on their individual experience to assess intraoperative RBC transfusion needs. Therefore, robust multicenter studies with standardized RBC transfusion criteria and external validation are needed before attempting clinical implementation of AI. Until then, AI models should be regarded as investigational and are not yet suitable for use as clinical decision-support tools.

Limitations and future prospects

Interpretation of these findings is limited by the available evidence. Given the novelty of this field, few studies met the inclusion criteria, and substantial heterogeneity persisted despite similar predictive features. A major limitation is the use of a subjective reference standard, as the absence of standardized protocols for RBC administration means model outcomes reflect clinician discretion and institutional practices rather than universal physiological thresholds, introducing potential incorporation bias. Furthermore, a relevant methodological limitation involves the reconstruction of 2×2 contingency tables for studies where the exact number of transfused patients in the testing set was not reported. In these instances, it was necessary to assume that the transfused-to-non-transfused ratio matched that of the training set. While comparative validation and sensitivity analyses demonstrated minimal discrepancies, this assumption remains a potential source of imprecision in the pooled estimates that must be explicitly highlighted. The low quality of evidence in GRADE assessment, which primarily derives from the high risk of bias and heterogeneity among studies, also presents a significant limitation.

Future research should prioritize large-scale, standardized, multicenter studies to validate these models in real-world settings, assess their impact on clinical outcomes, intraoperative RBC transfusion rates, and healthcare costs, and ensure generalizability through external validation. Adherence to standardized reporting frameworks, such as TRIPOD,³⁷ is essential to enable meaningful comparisons across studies.^{2,3,6,9}

Conclusion

This systematic review and meta-analysis suggest a potential exploratory role for AI in PBM strategies; however, current evidence is insufficient to support their use as a clinical decision-support tool. This interpretation is limited by the low certainty of evidence, heterogeneity across studies, and the use of non-standardized transfusion outcomes. Further large, standardized, and externally validated studies are required before clinical implementation can be considered.

Data availability statement

This meta-analysis is based on aggregate data extracted from publicly available studies. Requests for additional information should be directed to the corresponding authors of the original publications.

Study registry (PROSPERO)

CRD420251146515 (<https://www.crd.york.ac.uk/PROSPERO/view/CRD420251146515>, published September 16, 2025).

Declaration of generative AI and AI-assisted technologies in the writing process

During the revision of this study, the authors used Gemini 3 in order to reduce the word count and improve readability. The authors reviewed and edited the content as needed, and take full responsibility for the contents of the publication.

Authors' contributions

Filipe Giordano Valério: Conceptualization; methodology; formal analysis; investigation; writing-original draft; writing-review & editing. Amanda Carneiro Rodrigues: Data curation; writing-original draft; writing-review & editing. Davi Ricardo Soares Gama de Amorim: Writing-original draft; writing-review & editing; investigation. Roberta Esterque Cantarino: Writing-original draft; writing-review & editing. Glauco Martins de Araujo: Data curation; writing-review & editing. Jimmy Yusuf: Investigation, Data curation; writing-review & editing. Leticia Vianna Leal Soares Bessa: Investigation; writing-original draft; writing-review & editing. Milena Mendonça Andrade Paes Leme: Writing-original draft; writing-review & editing. Marcella Freire de Campos Euzebio: Writing-original draft; writing-review & editing. Flávio Luiz Seixas: Formal analysis; supervision; writing-review & editing. Alexandra Rezende Assad: Supervision; writing-review & editing. Luis Antonio dos Santos Diego: Supervision; writing-review & editing.

Conflicts of interest

The authors declare no conflicts of interest.

Funding

This research received no funding from public, commercial, or non-profit sources.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjane.2026.844782.

Editor

Liana Azi

References

1. Zhou R, Li Z, Liu J, et al. Prediction of intraoperative red blood cell transfusion in valve replacement surgery: machine learning algorithm development based on non-anemic cohort. *Front Cardiovasc Med*. 2024;11:1344170.
2. Cunha CBC, Lima TA, Ferraz DLM, et al. Predicting the Need for Blood Transfusions in Cardiac Surgery: A Comparison between Machine Learning Algorithms and Established Risk Scores in the Brazilian Population. *Braz J Cardiovasc Surg*. 2024. <https://doi.org/10.21470/1678-9741-2023-0212>.
3. Chen C, Wang Y, Yang X, et al. Machine Learning-Based Prediction of Intraoperative Red Blood Cell Transfusion in Aortic Valve Replacement Surgery. *Clin Lab*. 2024;70:709–17.
4. Liu S, Zhou R, Xia X-Q, et al. Machine learning models to predict red blood cell transfusion in patients undergoing mitral valve surgery. *Ann Transl Med*. 2021;9: 530-530.
5. Hung M, Besser M, Sharples LD, Nair SK, Klein AA. The prevalence and association with transfusion, intensive care unit stay and mortality of pre-operative anaemia in a cohort of cardiac surgery patients*. *Anaesthesia*. 2011;66:812–8.
6. Sun Z, Cui Z, Xie Y, et al. Evaluation of the Factors Influencing Blood Transfusion during Minimally Invasive Direct Coronary Artery Bypass Surgery. *Cardiology*. 2025;150:98–109.
7. Céspedes IC, Figueiredo MS, Hossne Junior NA, et al. Patient Blood Management Program Implementation: Comprehensive Recommendations and Practical Strategies. *Braz J Cardiovasc Surg*. 2024;39:e20240205.
8. Shander A, Hardy JF, Ozawa S, et al. A Global Definition of Patient Blood Management. *Anesth Analg*. 2022;135:476–88.
9. Boer C, Meesters MI, Milojevic M, et al. 2017 EACTS/EACTA Guidelines on patient blood management for adult cardiac surgery. *J Cardiothorac Vasc Anesth*. 2018;32:88–120.
10. Hassan J, Abbas M, Shaikat M, et al. Assessing operative mortality risk in cardiothoracic surgery: analysis of STS scores – a retrospective study. *Cardiothorac Surg*. 2025;33:18.
11. Nashef SAM, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. *Eur J Cardiothorac Surg*. 2012;41(4):734–44. discussion 744-745.
12. Alghamdi AA, Davis A, Brister S, Corey P, Logan A. Development and validation of Transfusion Risk Understanding Scoring Tool (TRUST) to stratify cardiac surgery patients according to their blood transfusion needs. *Transfusion (Paris)*. 2006;46:1120–9.
13. Chen S, Liu LP, Wang YJ, et al. Advancing Prediction of Risk of Intraoperative Massive Blood Transfusion in Liver Transplantation With Machine Learning Models. A Multicenter Retrospective Study. *Front Neuroinformatics*. 2022;16:893452. -893452.
14. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 1st edition Chichester (UK): John Wiley & Sons; 2023.
15. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions* editor(s). 2nd Edition Chichester (UK): John Wiley & Sons; 2019.
16. Salameh JP, Bossuyt PM, McGrath TA, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ*. 2020;370:m2632.
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5:210.
18. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 2019;170:51–8.
19. Schünemann HJ, Higgins JPT, Vist GE, et al. GRADE handbook for grading quality of evidence and strength of recommendations. *Cochrane handbook for systematic reviews of interventions*. p 375, 2nd ed. Chichester (UK): John Wiley & Sons; 2019. p. 375–402.
20. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Stat Methods Med Res*. 2018;27: 1785–805.
21. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14:135.
22. Posit team. *RStudio: Integrated Development Environment for R*. Posit Software. Boston, MA: PBC; 2025. Retrieved from <http://www.posit.co/>.
23. Philipp Doeblner. *mada: Meta-Analysis of Diagnostic Accuracy* Retrieved from <https://cran.r-project.org/package=mada>; 2025.
24. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019;22:153–60.
25. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982–90.
26. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20:2865–84.
27. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007;8:239–51.
28. Zhou Y, Dendukuri N. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: the case of meta-analyses of diagnostic accuracy. *Stat Med*. 2014;33: 2701–17.
29. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clin Transl Sci*. 2024;17:e70056.
30. Kaneko H. Cross-validated permutation feature importance considering correlation between features. *Anal Sci Adv*. 2022;3:278–87.
31. McKenna SLB, Dohoo IR. Using and interpreting diagnostic tests. *Vet Clin North Am Food Anim Pract*. 2006;22:195–205.
32. Mirza FG, Abdul-Kadir R, Breyman C, Fraser IS, Taher A. Impact and management of iron deficiency and iron deficiency anemia in women's health. *Expert Rev Hematol*. 2018;11:727–36.
33. Leff J, Romano CA, Gilbert S, Nair S. Validation Study of the Transfusion Risk and Clinical Knowledge (TRACK) Tool in Cardiac Surgery Patients: A Retrospective Analysis. *J Cardiothorac Vasc Anesth*. 2019;33:2669–75.

34. Ranucci M, Aronson S, Dietrich W, et al. Patient blood management during cardiac surgery: Do we have enough evidence for clinical practice? *J Thorac Cardiovasc Surg.* 2011;142:249.e1–249.e32.
35. Clevenger B, Mallett SV, Klein AA, Richards T. Patient blood management to reduce surgical risk. *Br J Surg.* 2015;102: 1325–37.
36. Butcher A, Richards T. Cornerstones of patient blood management in surgery. *Transfus Med.* 2018;28:150–7.
37. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med.* 2015;13:1–10.

REVIEW ARTICLE

Programmed intermittent epidural bolus versus continuous epidural infusion combined with patient-controlled epidural analgesia for postoperative analgesia: a meta-analysis of randomized controlled trials



Gabriel Lemos González ^a, Vinícius Fernandes Sarmiento ^b,
Bruno Francisco Minetto Wegner ^c, Gustavo Roberto Minetto Wegner ^d,
Alesson Marinho Miranda ^e, Iaci Luisa Lopes de Mattos ^a,
Tatiana Souza do Nascimento ^{f,*}

^a Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro, RJ, Brazil

^b Hospital do Servidor Público Estadual de São Paulo, São Paulo, SP, Brazil

^c Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

^d Universidade Federal da Fronteira Sul, Chapecó, SC, Brazil

^e Universidade Federal de São Paulo, São Paulo, SP, Brazil

^f University of Iowa, Iowa, USA

Received 12 December 2025; accepted 16 May 2026

Available online 23 May 2026

KEYWORDS

Continuous epidural infusion;
Epidural analgesia;
Patient-controlled epidural analgesia;
Postoperative pain;
Programmed intermittent epidural bolus

Abstract

Background: Programmed Intermittent Epidural Bolus (PIEB) and Continuous Epidural Infusion (CEI) are widely used strategies for maintaining postoperative analgesia, typically combined with Patient-Controlled Epidural Analgesia (PCEA). Although PIEB is theorized to enhance epidural drug distribution and analgesic effectiveness through intermittent high-pressure boluses, evidence supporting its superiority over CEI in the postoperative setting remains inconsistent.

Methods: We conducted a systematic review and meta-analysis of Randomized Controlled Trials (RCTs) comparing PIEB and CEI, both with PCEA, in adult postoperative patients (PROSPERO CRD420251046001). PubMed, Embase, and Cochrane were searched until May 1, 2025. Primary outcome: pain at 24 hours. Secondary outcomes: pain at other intervals, anesthetic consumption, and adverse events. Data were synthesized using a random-effects model; Risk of Bias (RoB-2) and small-study effects were assessed.

Results: Ten RCTs comprising 692 patients were included. No significant difference was found in 24-hour pain scores at rest (MD = -0.45; 95% CI -0.99 to 0.10) or movement (MD = -0.88; 95% CI

* Corresponding author.

E-mail: tatiana-souzadonascimento@uiowa.edu (T.S. Nascimento).

-2.07 to 0.32). PIEB showed mildly lower pain at other time points and reduced total epidural volume (MD = -7.31 mL; 95% CI -13.7 to -0.92 mL). However, PIEB was associated with higher hypotension risk (RR = 1.71; 95% CI 1.05 to 2.78). No differences were found in opioid consumption, PCEA demands, or other adverse events.

Conclusion: PIEB offers no significant analgesic advantage and increases hypotension risk compared to CEI in postoperative settings, when both are combined with PCEA. These findings should be interpreted with caution due to the limited number of studies and substantial heterogeneity.

PROSPERO registry: <https://www.crd.york.ac.uk/PROSPERO/view/CRD420251046001>.

© 2026 Sociedade Brasileira de Anestesiologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

The effective management of acute postoperative pain remains a major challenge in anesthesiology. Despite decades of pharmacologic and procedural advances, inadequate pain relief continues to hinder functional recovery, often prolonging hospitalization by limiting oral intake and early mobilization.^{1,2} Importantly, severe pain within the first 24 hours post-surgery is a known risk factor for the development of chronic pain and long-term impairment in quality of life.²⁻⁴

While various regional techniques have emerged, neuraxial analgesia delivered through an epidural catheter remains a widely implemented choice for major postoperative pain management.^{5,6} Its efficacy depends on multiple procedural factors, including the level of injection, the composition and concentration of the anesthetic solution, and the method of drug administration.^{7,8} Although single-bolus injections can be used, continuous infusions generally provide more stable and sustained analgesia for up to 72 hours after surgery compared with single-shot or systemic approaches.⁹

Modern infusion pumps enable two main modalities for maintaining baseline epidural analgesia: Continuous Epidural Infusion (CEI) and Programmed Intermittent Epidural Bolus (PIEB). CEI delivers a constant flow of medication, whereas PIEB administers pre-programmed boluses at fixed intervals. Both methods are frequently supplemented with Patient-Controlled Epidural Analgesia (PCEA) to manage breakthrough pain.⁷

In labor analgesia, combining PIEB or CEI with PCEA has been shown to enhance pain relief compared with traditional techniques, with PIEB demonstrating superior outcomes to CEI in several trials.¹⁰⁻¹² However, postoperative and labor pain represent distinct clinical entities. Labor pain is a unique physiological process modulated by individual perceptions and expectations of childbirth, precluding direct extrapolation of labor findings to surgical populations.¹³ Moreover, although also occurring in the obstetric setting, cesarean delivery constitutes a surgical procedure characterized by tissue injury and postoperative somatic pain, and should be distinguished from the vaginal delivery assessed in labor analgesia studies.¹⁴ As such, in postoperative settings, available Randomized Controlled Trials (RCTs) comparing PIEB and CEI combined with PCEA remain limited and heterogeneous, and their results have not been systematically integrated.

Therefore, we conducted a systematic review and meta-analysis of RCTs to assess whether PIEB provides improved

postoperative analgesia compared with CEI, both combined with PCEA, in adult patients. We evaluated pain at rest and during movement primarily at 24 hours as the primary outcome, and the secondary outcomes included other prespecified postoperative timepoints (12, 36, and 48 hours), anesthetic and opioid consumption, PCEA use, and adverse effects, including Postoperative Nausea and Vomiting (PONV), motor block, pruritus, hypotension, and patient satisfaction.

Methods

This systematic review and meta-analysis adhered to the Cochrane Handbook and PRISMA guidelines (Supplemental Table S1).^{15,16} The study protocol was registered at the International Prospective Register of Systematic Reviews (PROSPERO), under the registration number CRD420251046001 on May 5, 2025.

Eligibility criteria

Eligibility criteria were defined using the PICOS framework. We included studies enrolling adults (≥ 18 years) undergoing surgery (P) who received PIEB (I) or CEI (C), both combined with PCEA. Outcomes of interest (O) included postoperative pain at rest and during movement, anesthetic and opioid consumption, PCEA use, PONV, pruritus, motor block, hypotension, and patient satisfaction. Only peer-reviewed RCTs were eligible for inclusion (S).

We excluded studies with overlapping patient populations or with relevant differences in intraoperative and/or postoperative analgesic protocols beyond the studied interventions (e.g., use of different systemic analgesics between groups and studies presenting intervals with paused epidural infusion), as such differences could confound the comparative assessment of PIEB and CEI.

Data source and search strategy

We systematically searched PubMed, Embase, and Cochrane Library databases until May 1, 2025. The complete search strategies (free-text and indexed terms) for each database are provided in Supplemental Table S2.

After removing duplicates, two authors (G.L.G. and V.F.S.) screened the titles and abstracts and independently assessed the full-text articles for inclusion based on the prespecified criteria. Discrepancies were resolved through a panel discussion with the senior author (T.S.N.). There was no backward

citation tracking or request for data from authors. No automated tool was used, and all screening and selection processes were conducted manually.

Outcomes of interest

Our prespecified primary endpoint was pain score measured at 24 hours post-surgery. Pain outcomes were divided into rest pain and movement-related pain (elicited by movement).

Secondary outcomes included: 1) Rest pain at 12, 36 and 48 hours; 2) Movement-related pain at 12, 36 and 48 hours; 3) Epidural anesthetic solution consumption; 4) Patient-Controlled Epidural Analgesia (PCEA) administration frequency; 5) Opioid consumption; 6) PONV; 7) Motor block occurrence (Bromage Scale ≥ 1); 8) Pruritus; 9) Hypotension; and 10) Patient satisfaction.

Data extraction

Two authors (G.R.M.W. and A.M.M.) independently extracted data for each study using a standardized study form to determine: authors, study publication year, inclusion and exclusion criteria, sample size, follow-up period, baseline patient characteristics, type of surgery, intraoperative anesthesia, epidural analgesia characteristics, and adjuvant rescue analgesia techniques. Outcomes of interest were independently extracted by two authors (G.L.G. and B.F.M.W.). Any discordant assessments were adjudicated by the senior investigator (T.S.N.).

When data were reported as median with Interquartile Range (IQR) or median with range, we estimated the mean and Standard Deviation (SD) using the method developed by Luo, Shi, and Wan.¹⁷⁻¹⁹ When available, outcome data derived from intention-to-treat analyses were preferentially extracted. When trials reported outcomes using per-protocol or as-treated analyses, data were extracted as reported. No attempt was made to reconstruct intention-to-treat analyses or to reassign participants to randomized groups when such analyses were not reported.

All extracted numerical pain scores reported on 0–100 scales were converted to 0–10, and patient satisfaction scores originally reported on 0–10 scales were converted to 0–100 scales. Brief Pain Inventory reporting “pain in the last 24 hours” was considered equivalent to “pain at 24 hours”.

Data on the “total epidural volume” outcome were converted from milligrams to milliliters, if necessary, due to available solution concentration. Likewise, data related to PCEA administrations were converted from volume (mL) to the number of administrations when the bolus volume per administration was provided. Data on rescue opioid analgesia were converted to “Oral Morphine Milligrams Equivalent” (OMME).²⁰

Whenever feasible, heterogeneous outcome definitions (pain, hypotension, and PONV) were standardized to ensure consistency across studies. Detailed descriptions of outcome harmonization and study-specific adjustments are provided in Supplemental Table S3. Except for pain scores, all outcomes were analyzed as cumulative measures from the start of the postoperative period, ensuring no double-counting across time intervals. The follow-up duration applied to each outcome is detailed in Supplemental Table S4.

Subgroup and sensitivity analysis

Subgroup analyses were conducted to explore potential sources of heterogeneity and to assess the influence of clinical and methodological factors on pain outcomes and adverse events. Analyses were stratified by obstetric versus non-obstetric setting to account for physiological influences of labor. Additional subgroup analyses evaluated the impact of higher flow rates (defined as $\geq 5 \text{ mL.h}^{-1}$) and study Risk of Bias (RoB), comparing trials with increased risk to those at low risk according to RoB-2 assessment.

For the primary outcome, further subgroup analyses were performed according to surgical site, epidural solution, and PIEB interval. Secondary outcomes with shorter follow-up durations were stratified according to studies reporting outcomes at ≥ 24 vs. < 24 h.

Sensitivity analyses were performed for rest pain at 24 hours to assess the robustness of the findings. These included leave-one-out analyses and subgroup analyses separating studies with skewed data from those assuming normal distributions.

Quality assessment and risk of bias

Risk of bias and quality assessment of individual studies were analyzed using the Cochrane Collaboration’s tool RoB-2 for assessing the risk of bias in randomized studies and displayed using Robvis.^{21,22} Each trial was rated as “low risk,” “some concerns,” or “high risk” across five domains: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Assessments were conducted with respect to the effect of assignment to the intervention, consistent with an intention-to-treat framework. This assessment was performed independently by two authors (G.L.G. and G.F.M.W.), with disagreements resolved by consensus discussion involving the senior author (T.S.N.) (Fig. 1).

Potential reporting bias was assessed by visual inspection of contour-enhanced funnel plots for all outcomes. Effect estimates were plotted against their standard errors, with contours representing conventional significance thresholds and dashed pseudo-confidence limits centered on the pooled random-effects estimate. Egger’s regression was planned for outcomes including at least 10 RCTs but was not performed because no outcome met this threshold.

Grading of Recommendations Assessment, Development, and Evaluation (GRADE).²³ was also performed by two authors (G.L.G. and I.L.L.M.) on the most relevant clinical outcomes: rest pain at 24 hours; movement-related pain at 24 hours; opioid rescue analgesia consumption; total epidural volume; occurrence of hypotension; PONV; and patient satisfaction.

Statistical analysis

Binary outcomes were summarized as Risk Ratios (RRs), while continuous outcomes were analyzed using Mean Differences (MDs). Numerical rating scales, including the Visual Analog Scale (VAS) and the Numerical Rating Scale (NRS), were pooled using MDs due to their conceptual similarity and comparable clinical interpretability. For studies reporting medians without corresponding measures of dispersion,

	Risk of bias domains					
	D1	D2	D3	D4	D5	Overall
Avinash 2024	+	-	+	-	-	X
Bang 2024	+	+	+	+	+	+
Duncan 1998	+	-	+	+	-	-
Higashi 2019	+	-	+	-	+	-
Po-Yi 2020	+	+	-	+	+	-
Satomi 2018	+	+	+	+	+	+
Su-sung 2022	+	-	+	+	+	-
Wang 2022	+	+	+	+	+	+
Wiesmann 2018	+	+	+	+	+	+
Xiaofei 2022	+	+	+	+	+	+

Study

Domains:
D1: Bias arising from the randomization process.
D2: Bias due to deviations from intended intervention.
D3: Bias due to missing outcome data.
D4: Bias in measurement of the outcome.
D5: Bias in selection of the reported result.

Judgement
X High
- Some concerns
+ Low

Figure 1 Risk of Bias. Graphic representation of the RoB-2 risk of bias assessment informing the risk of every study's individual domain.

standard deviations were not imputed to minimize the risk of introducing bias.

Effect estimates are reported with 95% Confidence Intervals (95% CIs) and 95% Prediction Intervals (95% PIs). For continuous outcomes, MDs were calculated as PIEB minus CEI. For outcomes where lower values reflect benefit (e.g., pain scores and anesthetic consumption), negative MDs favor PIEB. For binary outcomes, RRs greater than 1 indicate a higher risk with PIEB, whereas values less than 1 favor PIEB.

Statistical heterogeneity was evaluated using Cochran's Q-test and the I^2 statistic. A p-value below 0.10 for the Q-test was considered indicative of statistically significant heterogeneity. I^2 values of < 25%, 25–50%, and > 50% were interpreted as low, moderate, and high heterogeneity, respectively. Owing to anticipated clinical variability across studies, a random-effects model was applied to all outcomes, with between-study variance estimated using the Restricted Maximum Likelihood (REML) method. No additional small-sample correction was applied, and statistical inference relied on conventional Wald-type confidence intervals. Statistical significance was defined as $p < 0.05$. All analyses were performed using R software (version 4.5.0; R Foundation for Statistical Computing, Vienna, Austria).

Results

Study selection and characteristics

Our initial search yielded 1528 potential articles. After removing duplicates and screening by title and abstract, 14 articles were retrieved and thoroughly reviewed for eligibility. Finally, 10 RCTs^{24–33} met the inclusion criteria and were included

in the analysis. A total of 692 patients were included, with a mean age ranging from 31.6 to 65 years, and 17.7% of participants were male. A comprehensive description of included studies is provided in Table 1, as well as a description of excluded studies in Supplemental Table S5. The complete study selection process is detailed in the flow diagram presented in Figure 2.

Outcomes

Primary outcomes

The PIEB group showed no statistically significant difference in rest pain at 24 hours (MD = -0.45; 95% CI -0.99 to 0.10; 95% PI = -2.16 to 1.27; $p = 0.109$; $I^2 = 73.3\%$; Fig. 3A) or movement-related pain scores at 24 hours (MD = -0.88; 95% CI -2.07 to 0.32; 95% PI -4.57 to 2.81; $p = 0.15$; $I^2 = 81.6\%$; Fig. 3B).

Secondary outcomes

Rest pain was significantly lower with PIEB at 12 hours (MD = -0.67; 95% CI -1.21 to -0.13; 95% PI -2.23 to 0.89; $p = 0.015$; $I^2 = 62.5\%$; Fig. 4A), at 36 hours (MD = -1.01; 95% CI -1.61 to -0.40; 95% PI -2.94 to 0.92; $p = 0.001$; $I^2 = 70.1\%$; Fig. 4B), and at 48 hours (MD = -0.70; 95% CI -1.27 to -0.13; 95% PI -2.41 to 1.01; $p = 0.016$; $I^2 = 59.2\%$; Fig. 4C). Meanwhile, movement-related pain was lower at 12 hours (MD = -0.82; 95% CI -1.39 to -0.24; 95% PI -2.08 to 0.45; $p = 0.006$; $I^2 = 0.0\%$; Fig. 4D), but not at 36 hours (Fig. S1). Insufficient data precluded the analysis of movement-related pain at 48 hours.

Regarding indirect pain-related outcomes, PIEB was associated with a lower total epidural volume administered (MD = -7.31 mL; 95% CI -13.7 to -0.92 mL; 95% PI -27.53 to

Table 1 Comprehensive description of included studies.

Author	n	Male (%)	Age (years)	ASA	Surgery	Type of anesthesia	Epidural Catheter	Epidural Solution	Loading dose	CEI	PIEB	Rescue analgesia	Follow-up
Avinash 2024	30	55%	60.3 [7.95]	I-II	Lower limb	Spinal	L2-3; L3-4	Ropivacaine 0.1% + Morphine 1 mg	-	6 mL.h ⁻¹	6 mL/ 60 min	PCEA: 4 mL; LT: 30 min; IV tramadol 50 mg PCEA: 2 mL; LT: 15 min; MV: 12 mL; IM Ketoprofen 100 mg PCA: morphine 1 mg; LT: 5 min	24 hours
Bang 2024	74	0%	35.3 [3.38]	II	Cesarean	Spinal	L2-3	Ropivacaine 0.11% + Fentanyl 3.7 µg.mL ⁻¹	-	4 mL.h ⁻¹	4 mL/ 60 min	PCEA: 2 mL; LT: 15 min; MV: 12 mL; IM Ketoprofen 100 mg PCA: morphine 1 mg; LT: 5 min	36 hours
Duncan 1998	40	0%	46.4 [14.83]	I-III	Lower abdominal	GA	T10-11; T11-12; T12-L1; L2-3	Bupivacaine 0.375%	Intraoperative or at end of surgery: 10 mL of bupivacaine 0.5% Postoperative: 5 mL of ropivacaine 0.2% + Fentanyl 2 µg.mL ⁻¹	5 mL.h ⁻¹	5 mL/ 60 min	PCEA: 3 mL; LT: 15 min	24 hours
Higashi 2019	42	-	65 [3.34]	-	Open lung lobectomy or partial lobectomy	GA	T4-6	Ropivacaine 0.2% + Fentanyl 2 µg.mL ⁻¹	5 mL of ropivacaine 0.2% + Fentanyl 2 µg.mL ⁻¹	3.4 mL.h ⁻¹	5.1 mL/ 90 min	PCEA: 3 mL; LT: 15 min	36 hours
Po-Yi 2020	120	29%	56.5 [15.04]	I-III	Major abdominal	GA	T6-8; T9-10; T11-12	Ropivacaine 0.0625% + Fentanyl 2 µg.mL ⁻¹	5 mL ropivacaine 0.0625% + Fentanyl 2 µg.mL ⁻¹	8 mL.h ⁻¹	4 mL/ 30 min	PCEA - CEI 2 mL; LT: 15 min PCEA - PIEB 2 mL; LT: 10 min; Opioids PCEA: 4 mL/h; LT: 1 hour; MV: 12 mL/h	72 hours
Satomi 2018	54	0%	53.5 [11.42]	I-III	Gynecological	GEA	T10-11; T11-12	Ropivacaine 0.2% + Fentanyl 2 µg.mL ⁻¹	At induction: 6 mL of ropivacaine 0.2% + Fentanyl 1.6 mcg.mL ⁻¹ At induction: 7.5 mL of levobupivacaine 0.2% + Fentanyl 6.67 mcg.mL ⁻¹	4 mL.h ⁻¹ *	4 mL/ 60 min**	PCEA: 3 mL; LT: 30 min; Ketorolac 30 mg	48 hours
Su-sung 2022	70	68%	63.1 [9.11]	I-III	Open thoracotomy	GA	T6-7; T7-8	Levobupivacaine 0.2% + Fentanyl 3.3 µg.mL ⁻¹	At induction: 7.5 mL of levobupivacaine 0.2% + Fentanyl 6.67 mcg.mL ⁻¹	1.1 mL.h ⁻¹	3 mL/ 180 min	PCEA: 3 mL; LT: 30 min; Ketorolac 30 mg	48 hours
Wang 2022	58	0%	31.6 [4.29]	I-II	Cesarean	Spinal	L1-2	Ropivacaine 0.2% + Fentanyl 2 µg.mL ⁻¹	Postoperative: 6 mL of ROPI 0.2% + Fentanyl 2 µg.mL ⁻¹	3 mL.h ⁻¹ *	3 mL/ 60 min**	PCEA: 2 mL; LT: 15 min; MV: 12 mL	48 hours
Wiesmann 2018	84	28%	60.6 [14.18]	I-III	Major abdominal & gynecological cancer	GEA	T8-9; T9-10	Ropivacaine 0.2% + Sufentanil 0.75 µg.mL ⁻¹	At induction: 15 mL of ropivacaine 0.375%	6 mL.h ⁻¹	6 mL/ 60 min	PCEA: 4 mL; LT: 30 min	48 hours
Xiaofei 2022	120	0%	33 [4.97]	II-III	Cesarean	CSE	L2-3; L3-4	Ropivacaine 0.1%	Postoperative: 8 mL of ropivacaine 0.125% + hydromorphone 0.6 mg + naloxone 0.04 mg	6 mL.h ⁻¹	6 mL/ 60 min ***	PCEA: 6 mL; LT: 15 min; MV: 24 mL; Rectal diclofenac 100 mg	36 hours

Age expressed in Mean [SD]; ASA, American Society of Anesthesiologists Physical Status; CEI, Continuous epidural infusion; PIEB, Programmed Intermittent Epidural Bolus; GA, General anesthesia; PCA, Patient controlled analgesia; GEA, Combined general-epidural anesthesia; CSE, Patient controlled epidural analgesia; LT, lockout time; MV, Maximum volume; CSE, Combined spinal-epidural; IV, Intravenous.

* Started immediately after loading dose;

** Started 1 hour after loading dose;

*** Started 30 minutes after loading dose.

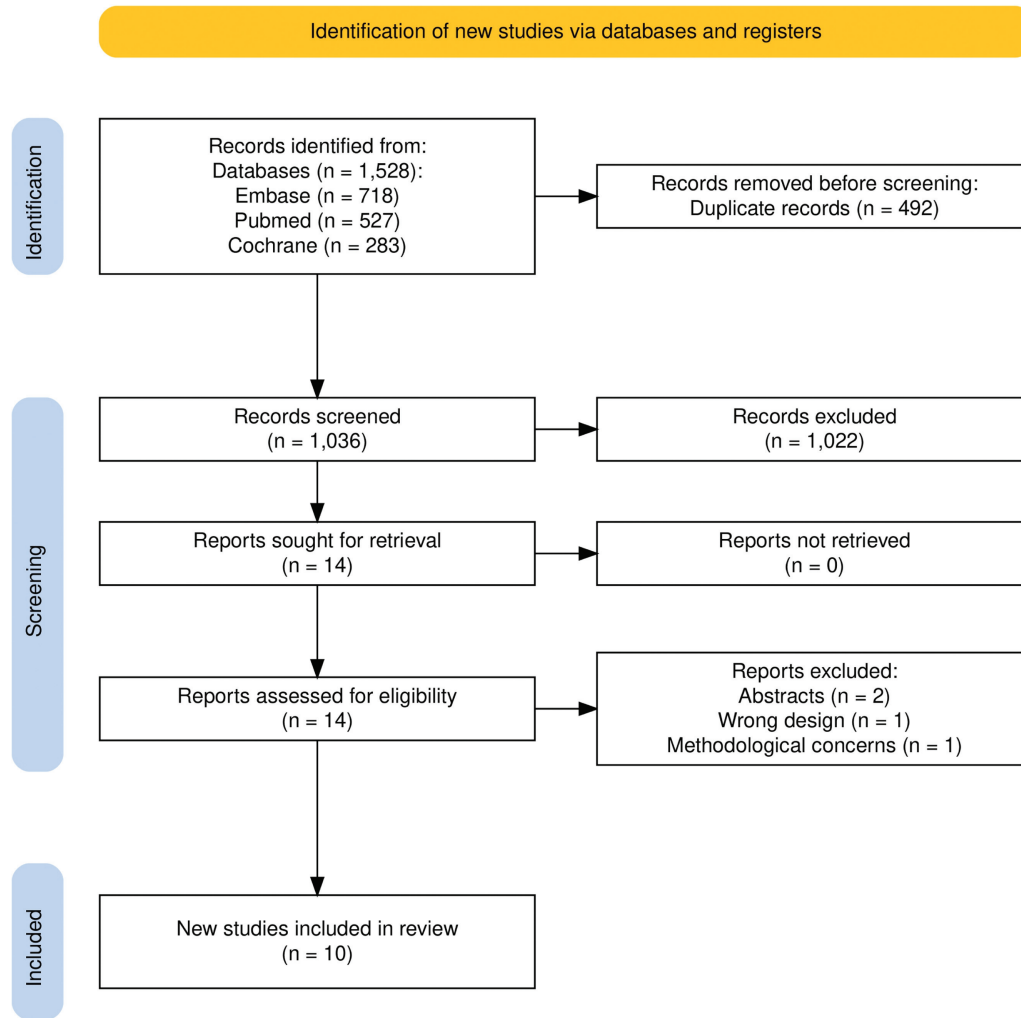


Figure 2 Prisma Flow Diagram. Diagram depicting information pertaining to different steps of the systematic review search process. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

12.90 mL; $p = 0.025$; $I^2 = 79.8\%$; Fig. 5A). For binary secondary outcomes, PIEB was associated with higher incidence of hypotension (RR = 1.71; 95% CI 1.05 to 2.78; 95% PI 0.78 to 3.77; $p = 0.03$; $I^2 = 0.0\%$; Fig. 5B).

Finally, no statistically significant differences were found between groups in PCEA administrations (Fig. S2), incidence of pruritus, urinary retention, PONV, motor block, or paresthesia (Fig. S3).

Subgroups and sensitivity analysis

Subgroup analyses by obstetric setting showed that cesarean patients receiving PIEB had lower movement-related pain at 36 hours (MD = -2.11; 95% CI -3.00 to -1.22; 95% PI -9.18 to 4.95; $I^2 = 12.4\%$) and higher satisfaction (MD = 8.59; 95% CI 1.89 to 15.28; 95% PI -56.04 to 73.22; $I^2 = 57.6\%$), with reduced heterogeneity and narrower 95% CIs for rest pain at 12 and 36 hours. Higher flow rates were associated with lower rest pain at 24 hours and a narrower 95% PI (MD = -0.85; 95% CI -1.23 to -0.47; 95% PI -1.69 to -0.01; $I^2 = 0.0\%$).

Studies at low RoB favored PIEB for satisfaction (MD = 8.59; 95% CI 1.89 to 15.28; 95% PI -56.04 to 73.22; $I^2 = 57.6\%$) and movement-related pain at 36 hours (MD = -2.11; 95%

CI -3.00 to -1.22; 95% PI -9.18 to 4.95; $I^2 = 12.4\%$). Unlike studies with increased RoB, those with low-RoB also demonstrated statistically significant differences favoring PIEB for rest pain at 12, 36, and 48 hours, movement-related pain at 12 hours, and total epidural volume.

Further directed subgroup analyses for rest pain at 24 hours showed significant effects favoring PIEB in epidural solutions containing only local anesthetics (MD = -0.62; 95% CI -1.18 to -0.06; 95% PI -4.25 to 3.02; $I^2 = 0.0\%$), PIEB intervals of 60-min (MD = -0.68; 95% CI -1.11 to -0.25; 95% PI -1.77 to 0.40; $I^2 = 48.2\%$), and studies without skewed data (MD = -0.75; 95% CI -1.24 to -0.26; 95% PI -2.08 to 0.58; $I^2 = 54.4\%$). No surgical site subgroup favored either intervention. Movement-related pain at 24 hours was not subjected to subgroup analysis because of the limited number of studies.

Stratification of secondary outcomes by follow-up duration did not materially alter results; however, hypotension lost statistical significance in both $< 24h$ and $\geq 24h$ subgroups. All subgroup analyses were exploratory and hypothesis-generating and should be interpreted cautiously given multiple testing and limited sample sizes (Supplemental Figs. S4–S9).

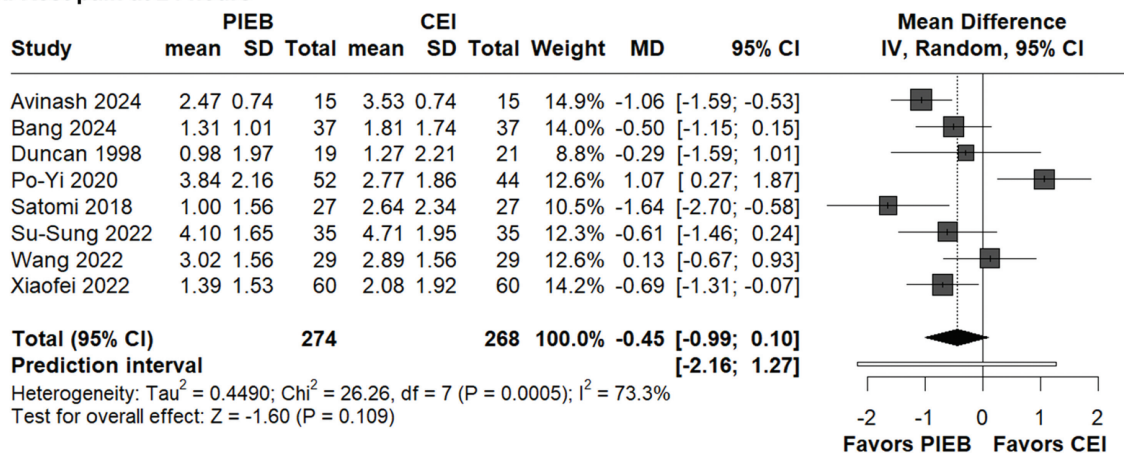
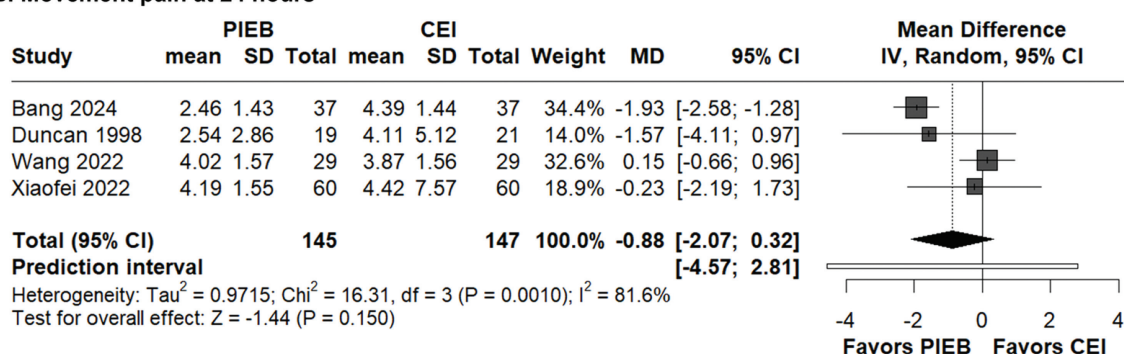
A. Rest pain at 24 hours**B. Movement pain at 24 hours**

Figure 3 Pain at 24 hours. Forest plots comparing PIEB and CEI (both with PCEA) on postoperative rest and movement-related pain (defined as pain elicited by movement), in a scale ranging from 0 to 10, at 24 hours. CEI, Continuous Epidural Infusion; PCEA, Patient-Controlled Epidural Analgesia; PIEB, Programmed Intermittent Epidural Bolus.

Leave-one-out sensitivity analyses reduced heterogeneity and favored PIEB for rest pain at 24 hours after exclusion of Po-Yi et al.²⁸ (MD = -0.68; 95% CI -1.04 to -0.32; $I^2 = 38.1\%$; Fig. S7) and for movement-related pain at 24 hours after exclusion of Wang et al.³² (MD = -1.51; 95% CI -2.56 to -0.47; $I^2 = 24.0\%$). Exclusion of Bang et al.³³ eliminated heterogeneity without altering statistical significance. Leave-one-out sensitivity analyses are presented in Supplemental Figures S10.

Risk of Bias

Based on the RoB-2 assessment, five of the ten included studies had some concerns, mainly related to missing outcome data and loss to follow-up, and one study²⁷ was judged at high risk of bias due to multiple domain-level concerns (Fig. 2).

Contour-enhanced funnel plots did not show clear evidence of reporting bias; studies were generally symmetrically distributed around pooled effects, with any observed asymmetry occurring within statistically significant contours and not altering the direction of effect. Interpretation was limited by imprecision and the small number of contributing trials, and no outcome included ≥ 10 RCTs to permit Egger's regression (Fig. S11).

Quality assessment

In the GRADE assessment, rest pain at 24 hours, total epidural volume, hypotension, and PONV were considered to have low certainty of evidence due to the relevant heterogeneity and high study variability. Meanwhile, movement-related pain at 24 hours, opioid rescue analgesia, and patient satisfaction were considered to have a very low certainty of evidence, as they also presented few events or studies. The evidence profile and summary of findings are presented in Supplemental Tables S6–S7.

Discussion

This systematic review and meta-analysis of 10 RCTs, encompassing 692 patients, compared the clinical outcomes of PIEB versus CEI, both combined with PCEA, for postoperative analgesia. The primary analysis revealed no statistically significant difference in pain scores at 24 hours, a key time point in postoperative recovery. Moreover, although PIEB was associated with mildly lower rest pain at other time points (12, 36, and 48 hours) and reduced total anesthetic volumes, these benefits were offset by a higher incidence of hypotension. No significant differences were observed in

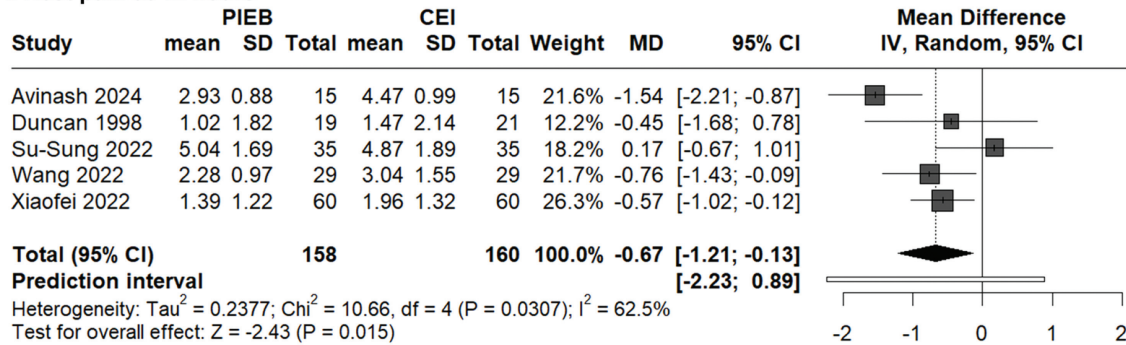
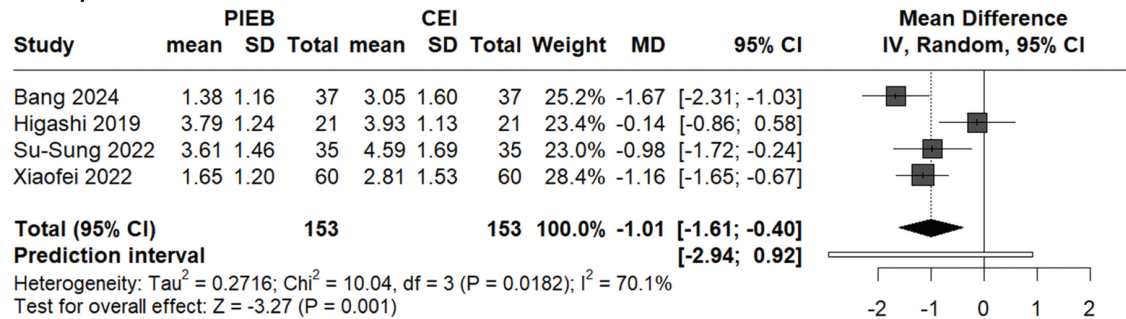
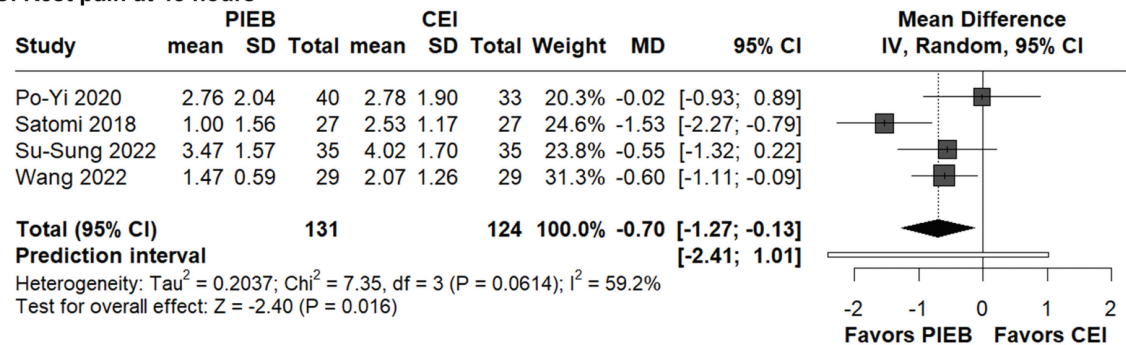
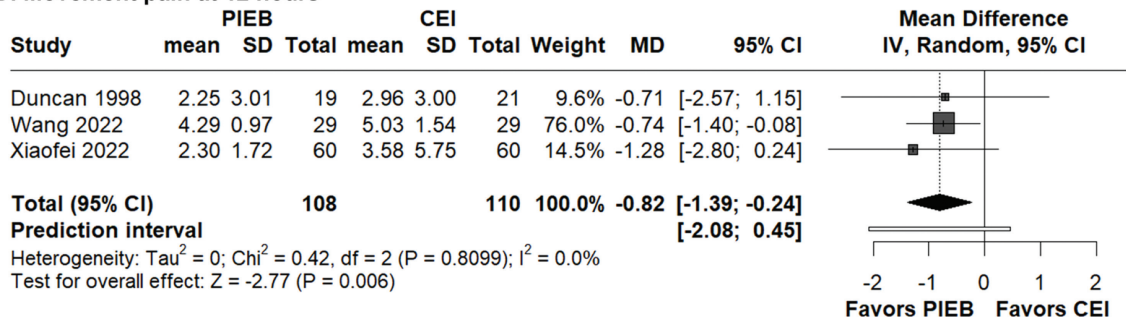
A. Rest pain at 12 hours**B. Rest pain at 36 hours****C. Rest pain at 48 hours****D. Movement pain at 12 hours**

Figure 4 Pain at other intervals. Forest plots comparing PIEB and CEI (both with PCEA) on postoperative rest pain at 12 (A), 36 (B) and 48 hours (C), as well as movement-related pain (defined as pain elicited by movement) at 12 hours (D), both in a scale ranging from 0 to 10. CEI, Continuous Epidural Infusion; PCEA, Patient-Controlled Epidural Analgesia; PIEB, Programmed Intermittent Epidural Bolus.

rescue opioid consumption, PCEA demands, patient satisfaction, or other adverse effects, including motor block and PONV.

The superior analgesic effects attributed to PIEB in labor analgesia^{10-12,34} have been linked to a wider epidural spread

from high-pressure intermittent boluses, a mechanism supported by experimental studies.³⁵⁻³⁸ This broader spread may enhance dermatomal coverage and analgesic efficiency.³⁹⁻⁴¹ However, such a physiological advantage likely

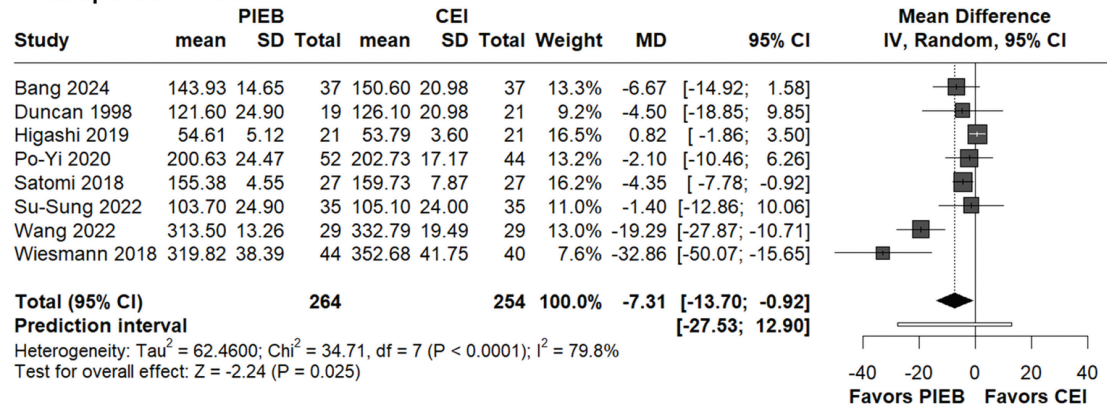
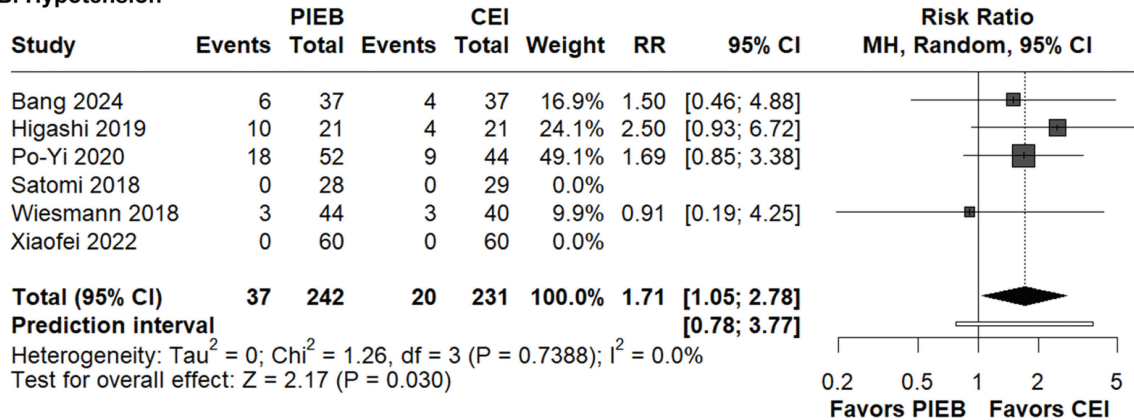
A. Total epidural volume**B. Hypotension**

Figure 5 Total epidural volume and hypotension. Forest plots comparing PIEB and CEI (both with PCEA) on total epidural volume infused (in milliliters) (A) and the occurrence of postoperative hypotension (B). CEI, Continuous Epidural Infusion; PCEA, Patient-Controlled Epidural Analgesia; PIEB, Programmed Intermittent Epidural Bolus.

depends on relatively homogeneous pain patterns in labor, whereas postoperative pain is more heterogeneous, involving variable somatic, visceral, and neuropathic components across surgical contexts. This variability, particularly when combined with PCEA-mediated bolus dosing, may attenuate any incremental benefit of programmed boluses.

Although PIEB was associated with statistically significant pain reductions at some timepoints, most effects did not reach the Minimal Clinically Important Difference (MCID) of one point on a 0–10 scale.⁴² Instances exceeding the MCID, such as rest pain at 36 hours in mixed cohorts and movement-related pain at 36 hours in cesarean patients, or where the 95% CI exceeded the MCID, in movement-related pain at 36 hours in low RoB studies, were supported by limited data. Thus, these isolated findings should be interpreted cautiously and in the context of adjacent timepoints that did not demonstrate clinically meaningful differences.

Furthermore, the combination of a non-significant pooled effect with substantial heterogeneity and wide PIs in the primary outcome suggests clinically meaningful variation across settings, indicating that the pooled estimates may not adequately represent individual surgical contexts. Variations in surgical pain intensity, catheter tip position, and solution viscosity likely influenced epidural spread, contributing to heterogeneity and attenuating the relative impact

of the epidural component. Consistently, the absence of clinically relevant differences in pain scores, opioid consumption, and PCEA use challenges the hypothesis that PIEB provides superior analgesic efficacy over CEI when both are combined with PCEA. It is plausible that the bolus mechanism inherent to PCEA, together with concurrent multimodal analgesia, further diminishes any incremental benefit of programmed boluses.

Exploratory subgroup and sensitivity analyses demonstrated only modest variations in effect estimates when evaluating the potential influence of labor-related physiological changes in postoperative settings – through analyses restricted to patients undergoing cesarean delivery – as well as when stratifying by epidural flow rate, RoB, PIEB interval, surgical site, anesthetic solution, data skewness, and follow-up duration. Despite isolated subgroup differences and multiple circumstances, the overall findings remained consistent with the primary analysis, showing no uniform clinically meaningful reduction in pain scores and a persistently increased risk of hypotension associated with PIEB.

Moreover, the overall consistency observed across subgroup and sensitivity analyses – despite several scenarios in which heterogeneity was reduced, such as analyses limited to cesarean delivery, studies at low RoB, and protocols using local anesthetics alone – suggests that, although residual

confounding related to the inclusion of mixed surgical settings cannot be excluded, the observed effects are likely reflective of the underlying clinical comparison of interest. Given the number of subgroup analyses performed, these findings are subject to an increased risk of false-positive results and should be interpreted cautiously.

Furthermore, exclusion of Po-Yi et al.,²⁸ which uniquely reported pain as “average pain”, or Wang et al.,³² which applied an unclear definition of movement-related pain, shifted pooled estimates in favor of PIEB and markedly reduced heterogeneity. While these findings may indicate outcome-definition-related bias in individual studies, they more importantly underscore the fragility of the available evidence and support the overall conclusion that clinical significance remains uncertain, despite the presence of isolated statistically significant results.

A key finding of this analysis is the increased risk of hypotension associated with PIEB, with an absolute risk increase of approximately 6.2% (range 0.4%–15.4%), a signal inconsistently reported in prior meta-analyses.^{10,11} This effect is biologically plausible, reflecting broader sympathetic blockade from enhanced epidural spread and reduced vascular tone.³⁵ Follow-up-stratified subgroup analyses (< 24 h; 24 vs. ≥ 24 hours), performed to explore potential temporal attenuation of hypotension as PCEA requirements decreased, were limited by reduced statistical power and did not reach statistical significance. Given that postoperative hypotension is an independent risk factor for major complications, including myocardial injury and acute kidney injury,^{43,44} any potential analgesic benefit of PIEB should be weighed against this hemodynamic risk.

These findings should be interpreted within the evolving landscape of postoperative pain management. The role of epidural analgesia itself is progressively being re-evaluated against newer regional techniques,^{5,6} which may carry a lower risk of failure, complications, and contraindications.^{7,8,45} The choice between PIEB and CEI represents refinement within a single technique, and clinicians should also evaluate less invasive alternatives according to patient and surgical factors.

Our analyses indicate that, contrary to theoretical expectations and previous findings in other settings, PIEB does not provide a significant analgesic advantage over CEI when both are combined with PCEA. Clinicians should consider these findings when selecting an epidural modality, balancing PIEB's inconsistent analgesic benefits against its potential for hemodynamic instability. Further high-quality RCTs with standardized protocols and clearly defined outcomes are needed to clarify the role of PIEB in specific surgical populations.

This meta-analysis has several limitations. Considerable heterogeneity existed across trials in surgical procedures, epidural techniques, bolus volumes and intervals, and follow-up duration. The predominance of female participants – largely reflecting gynecologic and obstetric surgery – limits generalizability to male patients. Outcome definitions and assessment time points varied, and several studies reported incomplete or imprecise data. In addition, the small number of trials per outcome precluded formal assessment of small-study effects. To address these issues, we standardized outcomes when feasible, performed prespecified subgroup and sensitivity analyses, and qualitatively evaluated funnel plot asymmetry.

Conclusion

Our findings indicate that PIEB does not provide a clinically meaningful or consistent analgesic advantage over CEI when both are combined with PCEA for postoperative analgesia. However, definitive inferences are limited by substantial heterogeneity across studies. In contrast, PIEB is associated with an increased risk of hypotension, representing a clinically relevant adverse effect that should be considered when selecting an epidural maintenance strategy.

Data availability statement

The dataset and R code used to reproduce all analyses are publicly available at Zenodo (DOI: <https://doi.org/10.5281/zenodo.20102829>). GRADE assessments were performed using the GRADEpro GDT online tool (grade.pro.org), and risk-of-bias visualizations were generated using the Robvis web application; these outputs are not reproduced by the provided R script.

AI assistance disclosure

The AI tool “Manus AI” was used exclusively for language polishing. All outputs were reviewed and edited by the authors. No AI or automated tool was involved in the core intellectual processes of the research. The authors assume full responsibility for the content and originality of this work.

Authors' contributions

Conception and design of the research: G.L.G. Acquisition of data: G.L.G., V.F.S., I.L.L.M., and B.F.M.W. Analysis and interpretation of the data: G.L.G., B.F.M.W., G.R.M.W., and A.M.M. Quality assessment: G.L.G., G.R.M.W., and I.L.L.M. Writing of the manuscript: All authors; Critical revision of the manuscript for intellectual content: All authors. Project supervision: G.L.G. and T.S.N.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The authors declare no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjane.2026.844766.

Associate Editor

Paulo Corrêa da Silva Neto

References

1. Moonesinghe SR, McGuckin D, Martin P, et al. The Perioperative Quality Improvement Programme (PQIP patient study): protocol for a UK multicentre, prospective cohort study to measure quality of care and outcomes after major surgery. *Perioper Med (Lond)*. 2022;11:37.
2. Small C, Laycock H. Acute postoperative pain management. *Br J Surg*. 2020;107:e70–80.
3. Fletcher D, Stamer UM, Pogatzki-Zahn E, et al. Chronic postsurgical pain in Europe: an observational study. *Eur J Anaesthesiol*. 2015;32:725–34.
4. Sinatra R. Causes and consequences of inadequate management of acute pain. *Pain Med*. 2010;11:1859–71.
5. Grass JA. The role of epidural anesthesia and analgesia in postoperative outcome. *Anesthesiol Clin North Am*. 2000;18:407–28. viii.
6. Block BM, Liu SS, Rowlingson AJ, Cowan AR, Cowan JA, Wu CL. Efficacy of postoperative epidural analgesia: a meta-analysis. *JAMA*. 2003;290:2455–63.
7. Avila Hernandez AN, Hendrix JM. Epidural anesthesia. *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2025. [cited 2025 Jul]. Available from <https://www.ncbi.nlm.nih.gov/books/NBK542219/>.
8. Brull R, Macfarlane A, Chan V. Spinal, epidural and caudal anesthesia. In: Miller RD, ed. *Miller's Anesthesia*, 9th ed., Philadelphia: Elsevier; 2019:1413–8.
9. Svensson I, Sjöström B, Haljamäe H. Assessment of pain experiences after elective surgery. *J Pain Symptom Manage*. 2000;20:193–201.
10. Wang L, Huang J, Chang X, Xia F. Effects of different neuraxial analgesia modalities on the need for physician interventions in labour: a network meta-analysis. *Eur J Anaesthesiol*. 2024;41:411–20.
11. Xu J, Zhou J, Xiao H, et al. Programmed intermittent bolus versus continuous infusion as the background infusion for parturient-controlled epidural analgesia: a systematic review and meta-analysis. *Sci Rep*. 2019;9:2583.
12. Wydall S, Zolger D, Owolabi A, Nzekwu B, Onwochei D, Desai N. Comparison of different delivery modalities of epidural analgesia and intravenous analgesia in labour: a systematic review and network meta-analysis. *Can J Anaesth*. 2023;70:406–42.
13. Whitburn LY, Jones LE, Davey MA, McDonald S. The nature of labour pain: an updated review of the literature. *Women Birth*. 2019;32:28–38.
14. Hardy-Fairbanks AJ, Lauria MR, Mackenzie T, McCarthy M. Intensity and unpleasantness of pain following vaginal and cesarean delivery: a prospective evaluation. *Birth*. 2013;40:125–33.
15. Higgins JPT, Thomas J, Chandler J, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.5 (updated 2024 Aug). Cochrane; 2024. Available from www.cochrane.org/handbook.
16. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
17. Shi J, Luo D, Weng H, et al. Optimally estimating the sample standard deviation from the five-number summary. *Res Synth Methods*. 2020;11:641–54.
18. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Stat Methods Med Res*. 2018;27:1785–805.
19. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14:135.
20. Nielsen S, Degenhardt L, Hoban B, Gisev N. A synthesis of oral morphine equivalents for opioid utilisation studies. *Pharmacoepidemiol Drug Saf*. 2016;25:733–7.
21. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.
22. McGuinness LA, Higgins JPT. Risk-of-bias visualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods*. 2021;12:55–61.
23. Guyatt GH, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383–94.
24. Higashi M, Shigematsu K, Nakamori E, Sakurai S, Yamaura K. Efficacy of programmed intermittent bolus epidural analgesia in thoracic surgery: a randomized controlled trial. *BMC Anesthesiol*. 2019;19:107.
25. Wiesmann T, Hoff L, Prien L, et al. Programmed intermittent epidural bolus versus continuous epidural infusion for postoperative analgesia after major abdominal and gynecological cancer surgery: a randomized, triple-blinded clinical trial. *BMC Anesthesiol*. 2018;18:154.
26. Lee SS, Baek JH, Park SJ, Kim HJ, Kim HY, Byeon GJ. Comparison of programmed intermittent epidural bolus injection and continuous epidural injection in controlling nighttime pain and improving sleep quality after thoracotomy. *Medicine (Baltimore)*. 2022;101:e31684.
27. Avinash M, Narayana S, Prameela V, Shyamprasad A. Comparison between epidural ropivacaine plus morphine intermittent boluses versus continuous infusion for postoperative analgesia in lower limb surgery. *J Cardiovasc Dis Res*. 2024;15:1616–22.
28. Su P, Peniche A, Clelland E, et al. Comparison of programmed intermittent epidural bolus and continuous epidural infusion for postoperative analgesia after major abdominal surgery: a randomized controlled trial. *J Clin Anesth*. 2020;64:109850.
29. Duncan L, Fried M, Lee A, Wildsmith J. Comparison of continuous and intermittent administration of extradural bupivacaine for analgesia after lower abdominal surgery. *Br J Anaesth*. 1998;80:7–10.
30. Satomi S, Kakuta N, Murakami C, Sakai Y, Tanaka K, Tsutsumi Y. The efficacy of programmed intermittent epidural bolus for postoperative analgesia after open gynecological surgery: a randomized double-blinded study. *Biomed Res Int*. 2018;2018:6297247.
31. Mo X, Zhao T, Chen J, et al. Programmed intermittent epidural bolus versus continuous epidural infusion for uterine contraction pain relief after cesarean section: a randomized, double-blind clinical trial. *Drug Des Devel Ther*. 2022;16:999–1009.
32. Wang L, Wu Z, Hu L, Wang Y. Programmed intermittent epidural bolus for post-cesarean delivery analgesia: a randomized controlled double-blind trial. *J Anesth*. 2022;36:32–7.
33. Bang Y, Jeong H, Kang R, et al. Comparison of analgesic effects between programmed intermittent epidural boluses and continuous epidural infusion after cesarean section: a randomized controlled study. *Korean J Anesthesiol*. 2024;77:374–83.
34. Huang R, Zhu J, Zhao Z, Wang B. The effect of programmed intermittent epidural bolus compared with continuous epidural infusion in labor analgesia with ropivacaine: a meta-analysis of randomized controlled trials. *Ann Palliat Med*. 2021;10:2408–20.
35. Callahan EC, Lim S, George RB. Neuraxial labor analgesia: maintenance techniques. *Best Pract Res Clin Anaesthesiol*. 2022;36:17–30.
36. Kaynar AM, Shankar KB. Epidural infusion: continuous or bolus? *Anesth Analg*. 1999;89:534.
37. Cole J, Hughey S. Bolus epidural infusion improves spread compared with continuous infusion in a cadaveric porcine spine model. *Reg Anesth Pain Med*. 2019;44:1080–3.
38. Mowat I, Tang R, Vaghadia H, Krebs C, Henderson WR, Sawka A. Epidural distribution of dye administered via an epidural catheter in a porcine model. *Br J Anaesth*. 2016;116:277–81.

39. Ueda K, Ueda W, Manabe M. A comparative study of sequential epidural bolus technique and continuous epidural infusion. *Anesthesiology*. 2005;103:126–9.
40. Wong CA, McCarthy RJ, Hewlett B. The effect of manipulation of the programmed intermittent bolus time interval and injection volume on total drug use for labor epidural analgesia: a randomized controlled trial. *Anesth Analg*. 2011;112:904–11.
41. Ni F, Wu Z, Zhao P. Programmed intermittent epidural bolus in maintenance of epidural labor analgesia: a literature review. *J Anesth*. 2023;37:945–60.
42. Myles PS, Myles DB, Gallagher W, et al. Measuring acute postoperative pain using the visual analog scale: the minimal clinically important difference and patient acceptable symptom state. *Br J Anaesth*. 2017;118:424–9.
43. Hoppe P, Kouz K, Saugel B. Perioperative hypotension: clinical impact, diagnosis, and therapeutic approaches. *J Emerg Crit Care Med*. 2020;4. <https://doi.org/10.21037/jeccm.2019.10.12>.
44. Liem VGB, Hoeks SE, Mol KHJM, et al. Postoperative hypotension after noncardiac surgery and the association with myocardial injury. *Anesthesiology*. 2020;133:510–22.
45. Rawal N. Epidural analgesia for postoperative pain: improving outcomes or adding risks? *Best Pract Res Clin Anaesthesiol*. 2021;35:53–65.

LETTER TO THE EDITOR

Erector spinae plane block vs. paravertebral block for postoperative analgesia in breast surgery: a meta-analysis of randomized trials



Dear Editor,

Effective analgesia after breast surgery is essential because acute postoperative pain may impair recovery and contribute to chronic postsurgical pain.¹ Paravertebral Block (PVB) remains a recommended regional analgesic technique for breast surgery, but its technical complexity and potential complications have prompted interest in simpler alternatives such as the Erector Spinae Plane Block (ESPB).² Although previous meta-analyses have suggested broadly comparable analgesic effects between ESPB and PVB, available evidence has been limited by small trial numbers, heterogeneous populations, and variable methodology.^{3,4} In light of the recently published Cochrane network meta-analysis showing no clinically meaningful difference between ESPB and PVB for 24-hour postoperative pain, we designed our analysis as a focused direct comparison of ESPB versus PVB in adult breast surgery.⁵ Our aim was not to establish equivalence, but to examine whether the available head-to-head randomized evidence demonstrates a clinically meaningful difference in 24-hour opioid consumption, incorporating Trial Sequential Analysis (TSA), exploratory meta-regression, and a prespecified Minimum Clinically Important Difference (MCID) to improve clinical interpretation.

We conducted a prospectively registered systematic review and meta-analysis (PROSPERO CRD420251118777) of randomized trials comparing ESPB with PVB in adult breast surgery. PubMed, Scopus, Embase, CENTRAL, and Web of Science were searched to 4 August 2025. The primary outcome was 24-hour opioid consumption, converted to intravenous Morphine Milligram Equivalents (MME). Random-effects meta-analysis was performed, with GRADE certainty assessment, sensitivity/subgroup analyses, meta-regression, and TSA using a prespecified MCID of 10 mg intravenous MME.⁶ The search strategy, PRISMA flow diagram, study characteristics, risk-of-bias assessment, forest and funnel plots, meta-regression, subgroup and sensitivity analyses, TSA, and certainty-of-evidence assessment are provided in the Supplementary Material for consultation.

27 RCTs published between 2019 and 2025 were included, comprising 2015 patients: 1004 receiving ESPB and 1011 receiving PVB. Risk of bias was judged as low in nine studies, with some concerns in ten, and high in eight. 17 RCTs (including 1021 patients) reported 24-hour opioid consumption. There was no significant difference between ESPB and PVB (MD = 0.62 MME, 95% CI -0.23 to 1.46; $p = 0.152$, $I^2 = 86.8\%$) (Table 1, Fig. 1). Egger's test suggested possible small-study effects ($p = 0.023$). TSA suggested that the cumulative evidence crossed the required information size. However, this result should not be interpreted as definitive, given the very low certainty of evidence, substantial heterogeneity, and possible small-study effects.

Subgroup analyses by surgery type and axillary involvement did not show significant effect modification. Sensitivity analysis excluding studies at high risk of bias produced results consistent with the primary analysis. Post-hoc meta-regression suggested that country, patient age, and local anesthetic choice may contribute to heterogeneity. However, these findings should be considered exploratory because they were based on study-level data and may be confounded by differences in surgical practice, block technique, perioperative analgesic protocols, and local anesthetic regimens.

Pain outcomes were generally comparable between techniques. Although pain at rest at extubation statistically favored PVB, no significant differences were observed at rest at 6, 12, or 24 hours, or during movement at any assessed time point. Similarly, PONV (OR = 0.93, 95% CI 0.61 to 1.42; $I^2 = 0\%$) and rescue analgesic requirement (OR = 1.13, 95% CI 0.63 to 2.05; $I^2 = 42.7\%$) did not differ significantly between groups (Table 1).

Overall, the available head-to-head randomized evidence does not demonstrate a clinically meaningful difference between ESPB and PVB for postoperative analgesia after breast surgery. However, these findings should not be interpreted as evidence of true equivalence. Confidence in the pooled estimate is limited by very low certainty of evidence, substantial heterogeneity, possible small-study effects, and variability in block technique, local anesthetic regimen, perioperative analgesic protocols, and outcome assessment.

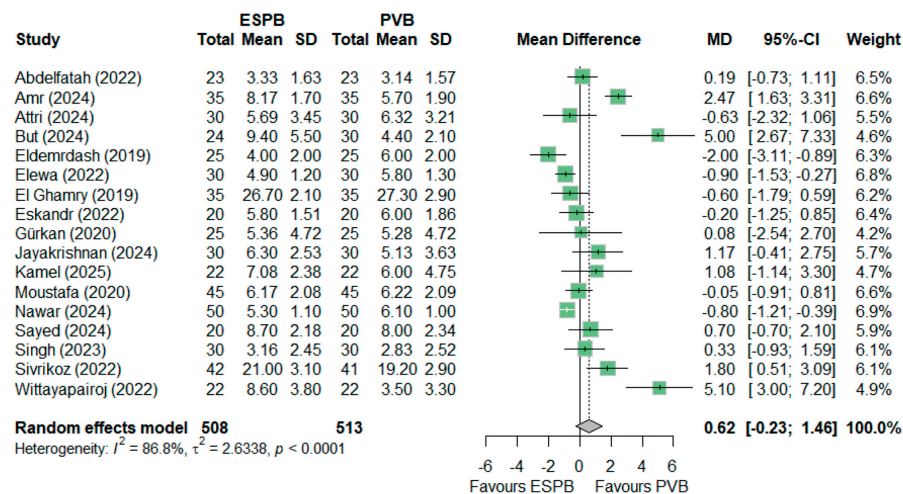
Study registration

PROSPERO: CRD420251118777.

Table 1 Overall analysis.

Outcome	N	ESPB	PVB	Total effect (95% CI)	I^2 (%)	p-value	Publication bias
Main outcome							
MME	17	508	513	MD 0.62 (-0.23; 1.46)	86.8	0.152	0.023
Pain at rest at extubation	16	495	500	MD 0.23 (0.01; 0.45)	81.4	0.039	0.085
Pain at rest at 6h	16	501	500	MD 0.02 (-0.24; 0.28)	81.1	0.905	0.737
Pain at rest at 12h	17	559	558	MD 0.00 (-0.31; 0.32)	91	0.995	0.913
Pain at rest at 24h	16	501	500	MD -0.01 (-0.23; 0.22)	77.1	0.965	0.449
Pain on movement at extubation	8	235	240	MD 0.36 (-0.11; 0.84)	82.1	0.132	Not assessed
Pain on movement at 6h	7	211	210	MD 0.23 (-0.12; 0.58)	80.7	0.204	Not assessed
Pain on movement at 12h	7	211	210	MD 0.34 (-0.03; 0.71)	82.8	0.060	Not assessed
Pain on movement at 24h	7	211	210	MD 0.44 (-0.06; 0.93)	84.6	0.080	Not assessed
PONV	13	409	408	OR 0.93 (0.61; 1.42)	0	0.737	0.448
Rescue Analgesics	6	164	163	OR 1.13 (0.63; 2.05)	42.7	0.675	Not assessed

ESPB, Erector Spinae Plane Block; PVB, Paravertebral Block; CI, Confidence Interval; N, Number of studies; MD, Mean Difference; MME, Morphine Milligram Equivalents; OR, Odds Ratio; PONV, Postoperative Nausea and Vomiting.

**Figure 1** Forest plots of milligrams of morphine equivalents at 24 hours between the groups.

Data availability statement

This study was based on aggregate data extracted from previously published randomized trials. All data generated or analyzed during this study are included in this article and its Supplementary Material. Additional analytic materials are available from the corresponding author upon reasonable request.

Authors' contributions

BD, ET and YEK conceptualized the manuscript. BD, YEK, CK, ET, MB and ADC collected data and wrote the first draft. CK, BD and ADC conducted literature review and assisted in data analysis. YEK, ET, MB, and BD wrote the final version of the manuscript with critical revisions from ADC and CK. All authors contributed significantly to the study's design, conduct, and analysis. All authors reviewed and approved the final version of the manuscript for publication.

Funding

None.

Conflicts of interest

The authors declare no conflicts of interest.

AI assistance disclosure

OpenAI's ChatGPT 5.2 was used to assist with language editing and text clarity. All scientific content and conclusions are the sole responsibility of the authors.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjane.2026.844767.

Associate Editor

Durval Kraychete

References

1. Jacobs A, Lemoine A, Joshi GP, Van de Velde M, Bonnet F. PROSPECT Working Group collaborators. PROSPECT guideline for oncological breast surgery: a systematic review and procedure-specific postoperative pain management recommendations. *Anaesthesia*. 2020;75:664–73.
2. Dost B, Bugada D, Karapinar YE, et al. Paravertebral block is not superior to the interpectoral and pectoserratus plane block for patients undergoing breast surgery: an updated meta-analysis of randomized controlled trials with meta-regression and trial sequential analysis. *Eur J Anaesthesiol*. 2025;42:637–48.
3. Weng WT, Wang CJ, Li CY, Wen HW, Liu YC. Erector spinae plane block similar to paravertebral block for perioperative pain control in breast surgery: a meta-analysis study. *Pain Physician*. 2021;24:203–13.
4. Chen W, Zhou X, Li H, et al. Analgesic effectiveness of preoperative ultrasound-guided erector spinae plane block versus paravertebral nerve block for breast surgery: A systematic review and meta-analysis of four randomized controlled trials. *Pain Pract*. 2023;23:511–22.
5. Clephas PR, Orbach-Zinger S, Gosteli-Peter MA, et al. Regional analgesia techniques for postoperative pain after breast cancer surgery: a network meta-analysis. *Cochrane Database Syst Rev*. 2025;6:CD014818.
6. De Cassai A, Dost B, Tulgar S, Boscolo A. Methodological standards for conducting high-quality systematic reviews. *Biology (Basel)*. 2025;14(8):973.

Burhan Dost ^{a,*}, Yunus Emre Karapinar ^b,
Muzeyyen Beldagli ^c, Esra Turunc ^a, Cengiz Kaya ^a,
Alessandro De Cassai ^{d,e}

^a Ondokuz Mayıs University Faculty of Medicine, Department of Anesthesiology and Reanimation, Samsun, Turkey

^b Istanbul University-Cerrahpasa, Cerrahpasa Faculty of Medicine, Department of Anesthesiology and Reanimation, Istanbul, Turkey

^c Samsun Training and Research Hospital, Department of Anesthesiology and Reanimation, Samsun, Turkey

^d University of Padua, Department of Medicine (DIMED), Padua, Italy

^e University Hospital of Padua, Institute of Anesthesia and Intensive Care Unit, Padua, Italy

* Corresponding author.

E-mail: burhandost@hotmail.com (B. Dost).

Received 13 January 2026; accepted 17 May 2026

Available online 28 May 2026

LETTER TO THE EDITOR

Psychotropic medication use during preanesthetic evaluation at a Brazilian university hospital: a retrospective cross-sectional analysis



Dear Editor,

The global rise in mental health disorders has been accompanied by increased use of psychotropic medications, which are now among the most frequently prescribed drug classes worldwide.¹ In surgical patients, chronic psychotropic use is clinically relevant because these agents may interact with anesthetics through pharmacokinetic and pharmacodynamic mechanisms.² Despite this, contemporary epidemiological data on psychotropic use among surgical patients remain limited.

We conducted a retrospective cross-sectional analysis using routinely collected electronic medical record data from the anesthesiology outpatient clinic of the Onofre Lopes University Hospital, Brazil. Adult patients (≥ 18 years) evaluated between January and December 2024 were consecutively included. Records without medication data were excluded. Duplicate records were filtered by retaining the most complete or most recent entry. Records with missing values for variables such as Body Mass Index (BMI) were retained for prevalence analyses, but excluded from statistical tests requiring those data.

Extracted variables included demographics, clinical characteristics, surgical specialty, and psychotropic medication use. Current psychotropic use – encompassing both scheduled and as-needed use – was determined by patient self-report during preanesthetic consultation and documented in the medical record. Without prior prespecification, reported agents were retrospectively categorized into therapeutic classes (antidepressants, anxiolytics/hypnotics, antipsychotics, mood stabilizers, and psychostimulants). Polypharmacy was defined as the concurrent use of two or more psychotropic agents.

Comparisons involved Chi-Square tests for categorical variables and Student's *t*-test or Mann-Whitney *U*-tests for continuous variables (age and BMI), with normality assessed via the Shapiro-Wilk test. Logistic regression estimated Odds Ratios (OR) with 95% Confidence Intervals. Models included age, sex, and BMI category a priori; additional variables

were excluded due to data heterogeneity and to preserve parsimony.

A total of 1,038 patients were included in the final analysis, of whom 23.0% ($n = 239$) were using at least one psychotropic medication (Table 1). This prevalence is higher than rates reported in Brazilian community-based studies, which generally range from 10% to 18%, suggesting that surgical patients may represent a subgroup with greater psychiatric burden and healthcare utilization.

Psychotropic medication use was significantly more frequent among women, affecting 29.55% of women compared with 14.12% of men ($p < 0.001$). In a multivariable logistic regression adjusted for age and BMI category, female sex remained independently associated with a higher likelihood of psychotropic medication use (OR = 2.74; 95% CI 1.80–4.15; $p < 0.001$) (Table 2), while age was not significantly associated with the outcome ($p = 0.440$). Due to missing BMI data, this multivariable model included 704 patients. Biological factors, including hormonal fluctuations influencing serotonergic and GABAergic neurotransmission, along with psychosocial determinants such as cumulative stress and social role burden, may contribute to this disparity.³

Mean BMI was higher among psychotropic users ($34.49 \pm 10.48 \text{ kg.m}^{-2}$) than among non-users ($32.13 \pm 9.89 \text{ kg.m}^{-2}$; $p = 0.005$). Although no independent association between categorical BMI and psychotropic use remained after adjustment, this finding is consistent with literature showing a bidirectional relationship between obesity and psychiatric disorders.⁴ From a perioperative perspective, obesity combined with psychotropic use may increase the risk of difficult airway management, hypoventilation, and obstructive sleep apnea. Psychotropic medication use also varied across surgical specialties, with higher proportions observed in digestive tract, plastic, and thoracic surgeries.

The pharmacological profile revealed that antidepressants were the most frequently prescribed therapeutic class (71.97%), followed by anxiolytics/hypnotics (30.13%), mood stabilizers (24.69%), and antipsychotics (19.67%) (Table 3). Clonazepam, sertraline, and fluoxetine were the most commonly prescribed individual medications.

Psychotropic polypharmacy was observed in 41.84% of users, representing a relevant point for preoperative medication review and individualized anesthetic planning. The concurrent use of multiple psychotropic medications may increase the potential for pharmacokinetic and pharmacodynamic interactions in the perioperative setting, including

Table 1 Demographic and clinical characteristics of patients.

Variable	Psychotropic use (n = 239) n (%) or mean $\pm$ SD	No psychotropic use (n = 799) n (%) or mean $\pm$ SD	p-value
Sex, n (%)			< 0.001 ^b
Male	62 (14.12)	377 (85.88)	
Female	177 (29.55)	422 (70.45)	
Age (years), mean $\pm$ SD	49.54 $\pm$ 14.42	50.99 $\pm$ 15.44	0.182 ^a
BMI (kg.m⁻²), mean $\pm$ SD	34.49 $\pm$ 10.48	32.13 $\pm$ 9.89	0.005 ^a
BMI category, n (%)			0.066 ^b
Underweight (BMI < 18.5 kg.m ⁻²)	1 (7.14)	13 (92.86)	
Normal weight (BMI 18.5–24.9 kg.m ⁻²)	26 (19.25)	109 (80.74)	
Overweight (BMI 25.0–29.9 kg.m ⁻²)	40 (19.32)	167 (80.67)	
Obesity class I (BMI 30–34.9 kg.m ⁻²)	29 (24.57)	89 (75.42)	
Obesity class II (BMI 35–39.9 kg.m ⁻²)	25 (30.48)	57 (69.51)	
Obesity class III (BMI $\geq$ 40.0 kg.m ⁻²)	43 (29.05)	105 (70.94)	
Surgical specialty, n (%)			< 0.001 ^b
General Surgery	57 (19.86)	230 (80.14)	
Digestive Tract Surgery	84 (32.18)	177 (67.81)	
Urology	20 (12.26)	143 (87.73)	
Vascular Surgery	22 (19.81)	89 (80.18)	
Plastic Surgery	21 (29.57)	50 (70.42)	
Thoracic Surgery	11 (36.67)	19 (63.33)	
Other Specialties	24 (20.87)	91 (79.13)	

SD, Standard Deviation; BMI, Body Mass Index; ^a *t*-test, ^b Chi-Square test.

Note: The total 'n' for BMI variables (n = 704) differs from the total sample due to missing data in medical records. For categorical variables, percentages indicate the proportion within each category/row, not column percentages.

cytochrome P450 modulation, GABA-A receptor adaptation, and QT interval prolongation.^{2,5} While our study was fundamentally descriptive and did not evaluate clinical outcomes, these mechanisms may contribute to altered anesthetic requirements and autonomic instability.

Current consensus favors individualized perioperative management rather than routine discontinuation of psychotropic medications. Abrupt withdrawal of antidepressants and anxiolytics may precipitate psychiatric destabilization, whereas medications such as monoamine oxidase inhibitors and lithium require specific perioperative strategies because of their interaction profiles.²

This study has limitations inherent to its retrospective and cross-sectional design, including reliance on electronic medical records, absence of dosage and treatment duration

data, and inability to correlate psychotropic exposure with perioperative outcomes. Nevertheless, the large sample and standardized extraction methods provide a relevant epidemiological overview of psychotropic medication use in a tertiary surgical population.

In summary, nearly one in four patients presenting for preanesthetic evaluation in our tertiary university hospital was using a psychotropic medication, with frequent polypharmacy and a higher prevalence among women. Antidepressants and benzodiazepines predominated. These findings reinforce the importance of psychotropic medication reconciliation during preanesthetic assessment. Future prospective studies should evaluate the relationship between psychotropic exposure patterns and postoperative outcomes to better guide anesthetic planning.

Table 2 Logistic regression analysis of factors associated with psychotropic drug use.

Variable	Category/Comparison	OR	95% CI	p-value
Sex (unadjusted)	Female vs. Male	2.52	1.83 – 3.49	< 0.001
Age	Per year increase	0.99	0.98 – 1.01	0.44
Sex (adjusted)	Female vs. Male	2.74	1.80 – 4.15	< 0.001
BMI category^a	Underweight	0.30	0.03 – 2.52	0.27
	Overweight	0.95	0.54 – 1.66	0.85
	Obesity class I	1.05	0.56 – 1.95	0.88
	Obesity class II	1.29	0.67 – 2.51	0.44
	Obesity class III	1.17	0.65 – 2.10	0.59

OR, Odds Ratio; CI, Confidence Interval.

^a Reference category: normal weight (BMI 18.5–24.9 kg.m⁻²). The multivariable model included age, sex, and BMI category and was fitted in 704 patients with complete BMI data. The adjusted OR for sex is adjusted for age and BMI category; BMI category estimates are adjusted for age and sex.

Table 3 Distribution of psychotropic medications by therapeutic class, individual agents and polypharmacy.

Therapeutic class	n	%
Antidepressants	172	71.97
Anxiolytics/Hypnotics	72	30.13
Mood stabilizers	59	24.69
Antipsychotics	47	19.67
Psychostimulants	3	1.26
Psychotropic medication	n	%
Clonazepam	54	22.59
Sertraline	46	19.25
Fluoxetine	43	17.99
Amitriptyline	31	12.97
Escitalopram	28	11.72
Quetiapine	26	10.88
Pregabalin	22	9.21
Topiramate	13	5.44
Risperidone	12	5.02
Sodium valproate	10	4.18
Desvenlafaxine	10	4.18
Number of psychotropic medications	n	%
One	139	58.16
Two	61	25.52
Three or more	39	16.32

Note: Percentages were calculated among psychotropic medication users (n = 239). Therapeutic classes and individual medications are not mutually exclusive.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

AI assistance disclosure

Claude AI was used exclusively to assist with language revision and formatting of the English text. No AI system was involved in the study design, data collection, data analysis, interpretation of results, or manuscript content generation. The authors take full responsibility for the content of the publication.

Authorship Criteria (ICMJE)

All authors meet the four International Committee of Medical Journal Editors (ICMJE) criteria for authorship: (1) Substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data; (2) Drafting the work or revising it critically for important intellectual content; (3) Final approval of the version to be published; and (4) Agreement to be accountable for all aspects of the work.

Ethics approval

The study was approved by the local Research Ethics Committee in accordance with national regulations for research involving human subjects (CAAE: 91167425.10000.5292).

Authors' contributions

Matheus Salera Gordim, Vitor Yan Bezerra de Araújo, Luzinelson Muniz da Silva Júnior and Danilo de Oliveira da Câmara Filho contributed to data collection, data analysis, and drafting of the manuscript. Heleno Paiva Oliveira, Emerson Arcoverde Nunes and Wallace Andrino da Silva contributed to study supervision, data analysis, critical revision of the manuscript for important intellectual content, and approval of the final version to be published. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding

The authors received no specific funding for this work.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

Ethics approval and consent to participate: Ethical approval was granted by the Onofre Lopes University Hospital Research Ethics Committee, under the Ethical Appreciation Presentation Certificate number 91167425.10000.5292.

Editor

Liana Azi

References

1. Stahl SM. Stahl's essential psychopharmacology: neuroscientific basis and practical applications. 5th ed. Cambridge: Cambridge University Press; 2021.
2. Oprea AD, Keshock MC, O'Glasser AY, et al. Preoperative Management of Medications for Psychiatric Diseases. *Mayo Clin Proc.* 2022;97:397–416.
3. Estancial Fernandes CS, De Azevedo RCS, Goldbaum M, Barros MBDA. Psychotropic use patterns: Are there differences between men and women? Reddy H, editor. *PLOS ONE.* 2018;13:e0207921.
4. Opel N, Hanssen R, Steinmann LA, et al. Clinical management of major depressive disorder with comorbid obesity. *Lancet Psychiatry.* 2025;12:780–94.
5. Kaye AD, Kline RJ, Thompson ER, et al. Perioperative implications of common and newer psychotropic medications used in clinical practice. *Best Pract Res Clin Anaesthesiol.* 2018;32:187–202.

Matheus Salera Gordim ^a, Vitor Yan Bezerra de Araújo ^a, Danilo de Oliveira da Câmara Filho ^a, Luzinelson Muniz da Silva Júnior ^b, Emerson Arcoverde Nunes ^c, Heleno Paiva Oliveira ^d, Wallace Andrino da Silva ^{d,*}

^a Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brazil

^b *Universidade Federal do Rio Grande do Norte (UFRN),
Hospital Universitário Onofre Lopes, Department of Surgery
(Anesthesiology Residency Program), Natal, RN, Brazil*

^c *Universidade Federal do Rio Grande do Norte (UFRN),
Hospital Universitário Onofre Lopes, Department of Internal
Medicine, Natal, RN, Brazil*

^d *Universidade Federal do Rio Grande do Norte (UFRN),
Hospital Universitário Onofre Lopes,
Department of Surgery, Natal, RN, Brazil*

* Corresponding author.

E-mail: wallaceandrino@yahoo.com.br (W.A. Silva).

Received 26 March 2026; accepted 14 June 2026

Available online 22 June 2026

LETTER TO THE EDITOR

Preoperative fasting and GLP-1 receptor agonists: gastric retention without increased aspiration risk



Dear Editor,

Preoperative fasting is a cornerstone of anesthetic safety, aiming to reduce pulmonary aspiration risk by minimizing gastric volume and acidity at induction.¹ Contemporary guidelines allow clear fluids up to 2 hours before anesthesia, reflecting a shift toward evidence-based, patient-centered fasting strategies. However, the increasing use of Glucagon-Like Peptide-1 Receptor Agonists (GLP-1RAs) has raised concerns due to their well-established effect of delaying gastric emptying, potentially leading to residual gastric contents despite guideline-compliant fasting.

Literature has consistently shown that GLP-1RAs are associated with delayed gastric emptying and increased residual gastric contents. Evidence from a large systematic review and meta-analysis shows a higher prevalence of retained gastric contents in patients receiving GLP-1RAs, particularly when administered in the days preceding surgery.² These findings are supported by physiological data showing reduced antral motility, increased pyloric tone, and modulation of vagal pathways, thereby contributing to delayed gastric emptying.³

Despite these physiological effects, higher-quality clinical evidence does not demonstrate a corresponding increase in aspiration-related outcomes. A recent observational cohort study of more than 300.000 patients reported no significant differences in aspiration pneumonia, respiratory complications, or mortality between GLP-1RAs users and non-users.⁴ This apparent dissociation between gastric retention and aspiration risk is clinically relevant and challenges the assumption that delayed gastric emptying alone translates into increased perioperative risk.

This discrepancy may reflect advances in perioperative airway management, widespread adoption of protective strategies, and improved risk stratification, which together mitigate the clinical impact of residual gastric contents. In this context, delayed gastric emptying does not appear to reliably predict aspiration risk in modern anesthetic practice.

Multidisciplinary consensus guidance reflects this evolving evidence, recommending continuation of GLP-1RAs in most patients and emphasizing individualized perioperative risk

assessment rather than routine drug discontinuation.⁵ Importantly, GLP-1RAs exert beneficial ancillary effects, including improved glycemic stability, modulation of inflammatory pathways, and potential cardiometabolic protection, which may be particularly relevant in the perioperative setting.

In conclusion, current evidence indicates that although GLP-1RAs are associated with delayed gastric emptying and increased residual gastric contents, this does not translate into a measurable increase in perioperative aspiration risk. Routine preoperative discontinuation therefore appears unwarranted. Instead, perioperative management should be guided by patient-specific risk assessment, taking into account fasting duration and, when appropriate, point-of-care gastric ultrasound, thereby avoiding unnecessary interruption of therapies with potential systemic benefits and supporting a tailored approach to the perioperative management of GLP-1RAs.

Data availability statement

No new datasets were generated or analyzed in this study.

Use of artificial intelligence

No AI or AI-assisted technologies were used in the preparation of this manuscript.

PROSPERO registration

Not applicable.

Reporting guidelines

Not applicable.

Funding

This research received no external funding.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Joshi GP, Abdelmalak BB, Weigel WA, et al. 2023 American Society of Anesthesiologists Practice Guidelines for Preoperative Fasting: Carbohydrate-containing Clear Liquids with or without Protein, Chewing Gum, and Pediatric Fasting Duration – A Modular Update of the 2017 American Society of Anesthesiologists Practice Guidelines for Preoperative Fasting*. *Anesthesiology*. 2023;138:132–51.
2. Elkin J, Rele S, Sumithran P, et al. Association between glucagon-like peptide-1 receptor agonist use and peri-operative pulmonary aspiration: a systematic review and meta-analysis. *Anaesthesia*. 2025;80:846–58.
3. Bellavance D, Chua S, Mashimo H. Gastrointestinal Motility Effects of GLP-1 Receptor Agonists. *Curr Gastroenterol Rep*. 2025;27:49.
4. Chen Y-H, Zink T, Chen Y-W, et al. Postoperative Aspiration Pneumonia Among Adults Using GLP-1 Receptor Agonists. *JAMA Netw Open*. 2025;8:e250081.
5. El-Boghdady K, Dhesi J, Fabb P, et al. Elective peri-operative management of adults taking glucagon-like peptide-1 receptor agonists, glucose-dependent insulinotropic peptide agonists and sodium-glucose cotransporter-2 inhibitors: a multidisciplinary consensus statement: A consensus statement from the Association of Anaesthetists, Association of British Clinical Diabetologists, British Obesity and Metabolic Surgery Society, Centre for Perioperative Care, Joint British Diabetes Societies for Inpatient Care, Royal College of Anaesthetists, Society for Obesity and Bariatric Anaesthesia and UK Clinical Pharmacy Association. *Anaesthesia*. 2025;80:412–24.

Filomena Di Vezza ^a, Marco Sanvitti ^b, Federico Bilotta ^{a,*}

^a *University of Tor Vergata, Department of Anesthesiology and Intensive Care, Rome, Italy*

^b *“Sapienza” University of Rome, Department of Pediatric Surgery, Rome, Italy*

* Corresponding author.

E-mail: federico.bilotta@uniroma2.it (F. Bilotta).

Received 27 February 2026; accepted 15 April 2026

Available online 1 May 2026

LETTER TO THE EDITOR

Erector spinae plane block versus caudal block in children: similar analgesia, different stories beneath the surface



Dear Editor,

We read with interest the systematic review and meta-analysis comparing Erector Spinae Plane Block (ESPB) with Caudal Epidural Block (CEB) in pediatric surgery.¹ In a specialty where practice often reflects institutional preference rather than robust evidence, such attempts to systematically synthesize available data are especially valuable. We would like to offer a few comments that may help readers interpret the findings and guide future research.

First, there is some inconsistency in the reporting of evidence certainty. The abstract states that all outcomes were supported by high-quality evidence, whereas the main text and GRADE summary indicate very low certainty for the primary outcome (time to first rescue analgesia) and moderate certainty for FLACC scores at 24 hours, with high certainty applying solely to Postoperative Urinary Retention (POUR) and Postoperative Nausea and Vomiting (PONV). This discrepancy may mislead readers who rely primarily on the abstract. Clarifying it and emphasizing that the primary outcome is supported by low-certainty evidence due to marked heterogeneity and imprecision, would provide a more balanced representation of how confidently equivalence between ESPB and CEB can be inferred.

Second, the extremely high heterogeneity for the primary analgesic outcome ($I^2 = 99\%$) substantially limits the interpretability of a single pooled estimate. Persistence of heterogeneity suggests that multiple clinically meaningful effect modifiers are contributing: variations in patients' age, inclusion of diverse surgical procedures (inguinal hernia, hypospadias, circumcision, hip and proximal femur surgery, renal surgery, lower limb surgery), variation in ESPB levels (sacral, lumbar, thoracic), and marked differences in local anesthetic concentrations and volumes. These factors may not only generate statistical heterogeneity but also question the clinical validity of aggregating all trials into a single summary measure. While subgroup analyses by surgical category or ESPB level could offer useful insights, the relatively small number of available trials would likely limit statistical power and precision. Such analyses should therefore be viewed as

hypothesis-generating rather than definitive, but they may still help frame future research questions.

Third, ESPB is treated as a uniform technique, although sacral, lumbar, and thoracic ESPB represent anatomically and mechanistically distinct blocks.² Sacral ESPB performed for procedures such as hypospadias or circumcision differs fundamentally from lumbar or lower thoracic ESPB used for hip or femur surgery, and these techniques may not be clinically interchangeable in terms of injectate spread, neural targets or analgesic profile. Therefore, pooling these under a single label risks masking level-specific outcomes. Future reviews may benefit from evaluating sacral ESPB separately or classifying it more precisely as a sacral multifidus plane block,³ particularly when the comparator is a sacral neuraxial block such as CEB.

Fourth, although the selected outcomes (time to first analgesic request and FLACC scores) are clinically relevant, they provide a limited view of postoperative recovery. Pediatric- and caregiver-centred metrics such as opioid consumption, block success rates, ease of performance, motor block, early mobilization, length of stay, and parent/patient satisfaction would enhance real-world clinical applicability. The authors' focus on POUR is commendable, and the lower incidence observed with ESPB is clinically meaningful.¹ However, POUR may be influenced by factors that are not consistently reported across trials, including neuraxial opioid use, catheterization practices, and the proportion of urological procedures.⁴ Clear reporting of these trial-level details would aid interpretation and support more nuanced patient selection. Finally, we appreciate the authors' call for standardized protocols and detailed reporting in future RCTs. Building on this, future studies may consider prespecifying minimally important differences for pain scales, standardizing or at least stratifying local anesthetic regimens and adjuvants, and adopting core outcome sets that incorporate safety, functional recovery, and patient-reported outcomes. These refinements would improve methodological rigor and facilitate more definitive meta-analytic conclusions. Such efforts are crucial for determining when ESPB can be reliably considered an alternative to CEB, particularly in ambulatory pathways or in children at elevated risk for POUR.

In conclusion, this review provides a timely comparison of ESPB and CEB, indicating similar analgesic performance but highlighting potential benefits of ESPB for avoiding POUR. As pediatric regional anesthesia evolves, more anatomically specific study designs and standardized reporting will be essential for defining when ESPB can reliably replace caudal techniques.

Data availability statement

No new data were created or analysed in this letter. Data sharing is not applicable to this article.

Declaration of AI use

The authors confirm that no AI tools or Language Models (LLMs) were used in the writing or editing of the manuscript.

Authors' contributions

Tuhin Mistry: Conceptualization; methodology; writing-original draft; review & editing.

Abhijit Sukumaran Nair: Validation; writing-review & editing; supervision.

Funding

No funding was received.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

The authors used artificial intelligence-assisted tools ChatGPT-5.2 Plus (OpenAI) for language editing only. All sci-

entific content, analysis, interpretation, and conclusions are the authors' own.

Editor

Liana Azi

References

1. Masiero BB, Cavalcante D, Akbarpoor F, et al. Erector spinae plane block versus caudal epidural block in pediatric surgery: a systematic review and meta-analysis of randomized clinical trials. *Braz J Anesthesiol.* 2025;75:844640.
2. Saadawi M, Layera S, Aliste J, Bravo D, Leurcharusmee P, Tran Q. Erector spinae plane block: A narrative review with systematic analysis of the evidence pertaining to clinical indications and alternative truncal blocks. *J Clin Anesth.* 2021;68:110063.
3. Nair AS, Mistry T, Garg R. Ultrasound-guided sacral multifidus plane block for perioperative analgesia: A comprehensive systematic review, meta-analysis, and trial sequential analysis. *Indian J Anaesth.* 2025;69:1341–58.
4. Cambise C, De Cicco R, Luca E, Punzo G, Di Franco V, Dottarelli A, Sacco T, Sollazzi L, Aceto P. Postoperative urinary retention (POUR): A narrative review. *Saudi J Anaesth.* 2024;18:265–71.

Tuhin Mistry ^{a,*}, Abhijit Sukumaran Nair ^b

^a *Ganga Medical Centre & Hospitals Pvt Ltd, Department of Anaesthesiology and Perioperative Care, Coimbatore, India*

^b *Ibra Hospital, Department of Anaesthesiology, Ibra, Sultanate of Oman*

* Corresponding author.

E-mail: tm.tuhin87@gmail.com (T. Mistry).

Received 6 January 2026; accepted 5 April 2026

Available online 22 April 2026

LETTER TO THE EDITOR

Response to Letter to the Editor: “Erector spinae plane block versus caudal block in children: similar analgesia, different stories beneath the surface”



Dear Editor,

We thank Dr. Mistry and Dr. Nair for their thoughtful Letter to the Editor¹ regarding our systematic review and meta-analysis comparing erector spinae plane block and caudal epidural block in pediatric surgery.² Their comments provide an important opportunity to clarify the interpretation of our findings, particularly regarding the certainty of evidence, heterogeneity, and anatomical classification of ESPB techniques.

Regarding the reported inconsistency in the certainty of evidence,¹ we acknowledge the discrepancy between the abstract and the GRADE assessment in the main text. As noted, the primary outcome (time to first rescue analgesia) was supported by low-certainty evidence due to heterogeneity and imprecision, while higher certainty applied only to selected secondary outcomes. We agree that this distinction is important, and conclusions regarding equivalence should be interpreted in this context. This reflects an inconsistency in reporting rather than differences in the underlying analyses.

We also appreciate the comments regarding the significant heterogeneity observed in the primary outcome ($I^2 = 99\%$).¹ This limitation was acknowledged and discussed in the manuscript and reflects the clinical diversity of the included studies, including differences in patient populations, surgical procedures, block levels, and anesthetic techniques.² Subgroup analyses were not feasible due to inconsistent reporting of key variables and the limited number of studies per comparison.² Sensitivity analyses showed that no single study disproportionately influenced the overall results.²

Regarding the classification of ESPB, we acknowledge that sacral, lumbar, and thoracic approaches may differ in anatomical spread and clinical application.² These differences were considered part of the clinical heterogeneity. ESPB was treated as a single intervention based on its shared underlying principle as a fascial plane block and a similar

technical approach.² We agree that further stratification by block level may provide additional insights when sufficient data are available.

Finally, we agree that broader clinical and patient-centered outcomes would enhance applicability. However, these variables, as well as factors influencing POUR such as neuraxial opioid use and surgical type, were inconsistently reported across the included trials and not available in sufficient detail for meaningful analysis.³⁻¹¹ These limitations reflect gaps in the available literature and highlight the need for more standardized reporting in future studies.

We appreciate the opportunity to clarify these points.

Data availability statement

No new data were created or analyzed in this article. Data sharing is not applicable to this article.

Use of AI and AI-assisted technologies

The authors declare that no artificial intelligence or AI-assisted technologies were used in the preparation of this manuscript.

Author's contributions

Barbara Bombassaro Masiero: Conceptualization; data curation; methodology; writing-original draft; project administration. Rafael Arsky Lombardi: Conceptualization; data curation; methodology; writing; project supervision; validation.

Funding

This research received no grant from any funding agency.

Conflicts of interest

All authors report no relationships that could be considered conflicts of interest. All authors take responsibility for all

aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

Editor

Liana Azi

References

1. Mistry T, Nair AS. Erector spinae plane block versus caudal block in children: similar analgesia, different stories beneath the surface. *Braz J Anesthesiol.* 2026;76:844759.
2. Masiero BB, Cavalcante D, Akbarpoor F, et al. Erector spinae plane block versus caudal epidural block in pediatric surgery: a systematic review and meta-analysis of randomized clinical trials. *Braz J Anesthesiol.* 2025;75:844640.
3. Bansal T, Kumar P, Kadian Y, et al. Comparison of ultrasound-guided sacral erector spinae plane block and caudal epidural block for analgesia in paediatric patients undergoing hypospadias repair: A double-blind, randomised controlled trial. *Indian J Anaesth.* 2024;68:725–30.
4. Abotaleb AM, Negm EE, Abdelwahed WM. A comparative study of preoperative ultrasound-guided lumbar erector spine plane block and preoperative ultrasound-guided caudal block for postoperative pain control in pediatric lower limb surgeries: A randomized controlled trial. *Egyp J Anaesth.* 2023;39:802–9.
5. Guan J, Liu L, Yang Y, Zheng Z, Li J, Zheng Z. Erector spinae plane block versus caudal block for postoperative analgesia in pediatric patients undergoing inguinal hernia repair: a randomized controlled trial. *Ann Med.* 2023;55:2298868.
6. Abdelrazik AN, Ibrahim TI, Farghaly AE, Shadwa RM. Ultrasound-guided Erector Spinae Muscle Block Versus Ultrasound-guided Caudal Block in Pediatric Patients Undergoing Lower Abdominal Surgeries. *Pain Physician.* 2022;25:E571–80.
7. Mandour O, Abdel-Aal I, Salem C, Refaat A. Analgesic effect of ultrasound-guided erector spinae plane block versus ultrasound-guided caudal block in pediatric open renal surgeries: A randomized comparative study. *Bali J Anaesthesiol.* 2023;7:60.
8. Özen V, Şahin AS, Ayyıldız EA, Açık ME, Eyileten T, Özen N. Comparison of Caudal Block and Sacral Erector Spina Block for Postoperative Analgesia following Pediatric Circumcision: A Double-Blind, Randomized Controlled Trial. *Urol Int.* 2024;108:292–7.
9. Pandey A, Ahmad Z, Jain S, et al. The analgesic efficacy of ultrasound-guided erector spinae plane block versus ultrasound-guided caudal epidural block for abdominal surgery in pediatric patients – A patient and assessor-blind, randomized controlled study. *Saudi J Anaesth.* 2024;18:55–61.
10. Elshazly M, Shaban A, Gouda N, Rashad M, Soaida SM. Ultrasound-guided lumbar erector spinae plane block versus caudal block for postoperative analgesia in pediatric hip and proximal femur surgery: a randomized controlled study. *Korean J Anesthesiol.* 2023;76:194–202.
11. Elbadry AA, Elrahwan SM, Mostafa TAH, Elahwal L. Analgesic effect of sacral erector spinae, penile and caudal block after hypospadias surgery: A randomized single blind controlled trial. *Egypt J Anaesth.* 2023;39:563–70.

Barbara Bombassaro Masiero ^{a,*},
Rafael Arsky Lombardi ^b

^a *Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil*

^b *University of Nebraska, Lincoln, Kansas, United States of America*

* Corresponding author.

E-mail: barbarabmasiero@gmail.com (B.B. Masiero).

Received 17 April 2026; accepted 17 June 2026

Available online 11 July 2026

