



**PMAC**  
PRINCE MAHIDOL  
AWARD CONFERENCE 2026

E-ISSN 2228-8082

Volume 78, Number 3, March 2026



# S M I J

## Siriraj Medical Journal

The world-leading biomedical science of Thailand

**MONTHLY**

**ORIGINAL ARTICLE**

**REVIEW ARTICLE**

### AI-Assisted Digital Mammography: A Pilot Reader Study



*AI assistance influenced diagnostic performance and reading time, with heterogeneous effects across readers.*

#### Population & Setting



**6**

**radiologists**



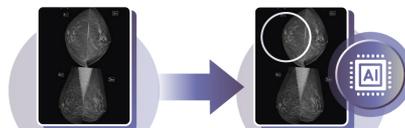
**86**

**mammography cases**



**Siriraj Breast Imaging Center (2018-2019)**

#### Study Intervention/Comparison



**Without AI**

**With AI-CAD**

(Perceptra MMG v0)

**Performance Averages Comparison (Unaided vs. AI-Assisted)**

Metric	Unaided	AI-Assisted
Mean Sensitivity	56.1%	77.5%
Mean Reading Time	32.9 Seconds	21.0 Seconds

#### Study Outcomes



• **Sensitivity increased in most readers**  
Increased in 5 of 6 readers



• **Specificity varied across readers**  
Significant improvement in one readers



• **Reading time decreased**  
32.9 s → 21.0 s per case  
Reduced for both cancer & non-cancer cases



**Pilot study • Exploratory • Larger studies needed**

SCAN FOR FULL TEXT



SIRIRAJ  
MEDICAL  
JOURNAL

Patanawanitkul, et al. *Siriraj Med J* 2026;78(3):185-195.

©Siriraj Medical Journal. All rights reserved. Graphical abstract by Y Krimouy

Indexed by

**Scopus**<sup>®</sup>

**DOAJ**

**EBSCO**



<https://he02.tci-thaijo.org/index.php/sirirajmedj/index>

E-mail: [sijournal92@gmail.com](mailto:sijournal92@gmail.com)

### ORIGINAL ARTICLE

Volume **78** Number **3**

**March** 2026

- 175** Prevalence, Predictive Factors, and Surgical Outcomes of Strabismus in High Myopia  
*Thammanoon Surachatkumtonekul, Piyaphat Jaruniphakul*
- 
- 185** Diagnostic Performance of AI-CAD Digital Mammography for Breast Cancer: Experience from Siriraj Breast Imaging Center  
*Rujira Patanawanitkul, Voraparee Suvannareg, Shanigarn Thiravit, Kobkun Muangsomboon, Pornpim Korpraphong*
- 
- 196** Predictive Models for Screening of Postoperative Cognitive Dysfunction in Older Surgical Patients  
*Arunotai Siriussawakul, Patumporn Suraarunsumrit, Varalak Srinonprasert, Pawit Somnuke, Panop Limratana, Unchana Sura-amonrattana, Ekkaphop Morkphrom, Busadee Pratumvinit, Surapa Tornsatitkul, Chalita Jiraphorncharas*
- 
- 207** Effects of Forward Leaning Characteristics on Protective Steps when Performing Voluntary-induced Stepping Response in Young and Older Adults: A Cross-Sectional Study  
*Pornprom Chayasit, Rumpa Boonsinsukh*
- 
- 218** Development of a Nomogram That Predicts Outcomes After Radical Cystectomy for Bladder Cancer Using Data from Siriraj Hospital, Thailand  
*Kanawut Sooksatian, Kantima Jongjitaree, Thitipat Hansomwong, Varat Woranisarakul, Patkawat Ramart, Siros Jitpraphai, Ekkarin Chotikawanich, Tawatchai Taweemonkongsap*
- 
- 229** Clinical Characteristics and Surgical Outcomes of Renal Epithelioid Angiomyolipoma: A Comparison with the Classic Type  
*Nattaporn Wanvimolkul, Ekkarin Chotikawanich, Siros Jitpraphai, Varat Woranisarakul, Thitipat Hansomwong, Kantima Jongjitaree, Pongsatorn Laksanabunsong, Ngoentra Tantranont, Tawatchai Taweemonkongsap*

### REVIEW ARTICLE

- 240** Genotoxicity and Cytotoxicity among Pesticide- Exposed Workers: A Systematic Review and Meta- Analysis  
*Achmad Ilham Tohari, Muhammad Rijal Fahrudin Hidayat, Nabil Athoillah, Muhammad Yuda Nugraha, Elly Nurus Sakinah, Supangat, Saekhol Bakri, Athira Nandakumar*



## Executive Editor:

Professor Apichat Asavamongkolkul, Mahidol University, Thailand

## Editorial Director:

Professor Aasis Unnanuntana, Mahidol University, Thailand

## Associate Editors

Assistant Professor Adisorn Ratanayotha, Mahidol University, Thailand

Pieter Dijkstra, University of Groningen, Netherlands

Professor Phunchai Charatcharoenwitthaya, Mahidol University, Thailand

Professor Varut Lohsiriwat, Mahidol University, Thailand



## Editor-in-Chief:

Professor Thawatchai Akaraviputh,  
Mahidol University, Thailand

## International Editorial Board Members

Allen Finley, Delhousie University, Canada

Christopher Khor, Singapore General Hospital, Singapore

Ciro Isidoro, University of Novara, Italy

David S. Sheps, University of Florida, USA

David Wayne Ussery, University of Arkansas for Medical Sciences, USA

Dennis J. Janisse, Medical College of Wisconsin, USA

Dong-Wan Seo, University of Ulsan College of Medicine, Republic of Korea

Folker Meyer, Argonne National Laboratory, USA

Frans Laurens Moll, University Medical Center Utrecht, Netherlands

George S. Baillie, University of Glasgow, United Kingdom

Gustavo Saposnik, Unity Health Toronto, St. Michael Hospital, Canada

Harland Winter, Harvard Medical School, USA

Hidemi Goto, Nagoya University Graduate School of Medicine, Japan

Ichizo Nishino, National Institute of Neuroscience NCNP, Japan

Intawat Nookaew, University of Arkansas for Medical Sciences, USA

James P. Doland, Oregon Health & Science University, USA

John Hunter, Oregon Health & Science University, USA

Karl Thomas Moritz, Swedish University of Agricultural Sciences, Sweden

Kazuo Hara, Aichi Cancer Center Hospital, Japan

Keiichi Akita, Institute of Science Tokyo, Japan

Kyoichi Takaori, Kyoto University Hospital, Japan

Marcela Hermoso Ramello, University of Chile, Chile

Marianne Hokland, University of Aarhus, Denmark

Matthew S. Dunne, Institute of Food, Nutrition, and Health, Switzerland

Mazakayu Yamamoto, Tokyo Women's Medical University, Japan

Mitsuhiro Kida, Kitasato University & Hospital, Japan

Moses Rodriguez, Mayo Clinic, USA

Nam H. CHO, Ajou University School of Medicine and Hospital, Republic of Korea

Nima Rezaei, Tehran University of Medical Sciences, Iran

Noritaka Isogai, Kinki University, Japan

Philip A. Brunell, State University of New York At Buffalo, USA

Philip Board, Australian National University, Australia

Ramanuj Dasgupta, Genome Institution of Singapore

Richard J. Deckelbaum, Columbia University, USA

Robert W. Mann, University of Hawaii, USA

Robin CN Williamson, Royal Postgraduate Medical School, United Kingdom

Sara Schwanke Khilji, Oregon Health & Science University, USA

Seigo Kitano, Oita University, Japan

Seiji Okada, Kumamoto University

Shomei Ryozaawa, Saitama Medical University, Japan

Shuji Shimizu, Kyushu University Hospital, Japan

Stanley James Rogers, University of California, San Francisco, USA

Stephen Dalton, Chinese University of HK & Kyoto University

Tai-Soon Yong, Yonsei University, Republic of Korea

Tomohisa Uchida, Oita University, Japan

Victor Manuel Charoenrook de la Fuente, Centro de Oftalmologia Barraquer, Spain

Wikrom Karnsakul, Johns Hopkins Children's Center, USA

Yasushi Sano, Director of Gastrointestinal Center, Japan

Yik Ying Teo, National University of Singapore, Singapore

Yoshiki Hirooka, Nagoya University Hospital, Japan

Yozo Miyake, Aichi Medical University, Japan

Yuji Murata, Aizenbashi Hospital, Japan

## Editorial Board Members

Vitoon Chinswangwatanakul, Mahidol University, Thailand

Jarupim Soongswang, Mahidol University, Thailand

Jaturat Kanpittaya, Khon Kaen University, Thailand

Nopphol Pausawasdi, Mahidol University, Thailand

Nopporn Sittisombut, Chiang Mai University, Thailand

Pa-thai Yenchitsomanus, Mahidol University, Thailand

Pornprom Muangman, Mahidol University, Thailand

Prasit Wattanapa, Mahidol University, Thailand

Prasert Auewarakul, Mahidol University, Thailand

Somboon Kunathikom, Mahidol University, Thailand

Supakorn Rojananin, Mahidol University, Thailand

Suttipong Wacharasindhu, Chulalongkorn University, Thailand

Vasant Sumethkul, Mahidol University, Thailand

Watchara Kasinrerker, Chiang Mai University, Thailand

Wiroon Laupattrakasem, Khon Kaen University, Thailand

Yuen Tanniradorn, Chulalongkorn University, Thailand

**Editorial Assistant:** Nuchpraweeapawn Saleeon, Mahidol University, Thailand

**Proofreader:** Amornrat Sangkaew, Mahidol University, Thailand, Nuchpraweeapawn Saleeon, Mahidol University, Thailand

# Prevalence, Predictive Factors, and Surgical Outcomes of Strabismus in High Myopia

Thammanoon Surachatkumtonekul, M.D.\*<sup>ID</sup>, Piyaphat Jaruniphakul, M.D.<sup>ID</sup>

Department of Ophthalmology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.

## Prevalence, Predictive Factors and Surgical Outcomes of Strabismus in High Myopia



**991**  
High myopia  
Patients

Retrospective cohort study  
At Siriraj Hospital  
Jan 2021 - June 2023

Strabismus 6.7% (66 patients)



Surgery was performed in  
25 patients (37.9%)

Success rate 76%



### Conclusion



Strabismus occurs more frequently in patients with high myopia. Axial length is a strong predictor, with defined cut-off values aiding early detection. Surgical correction yields favorable outcomes comparable to non-myopic patients, supporting targeted management in this population.

**Objective:** To determine the prevalence and predictive factors of strabismus in patients with high myopia and to evaluate the surgical outcomes of strabismus correction in this population.

### Factors Associated

#### Patients with strabismus



Younger age



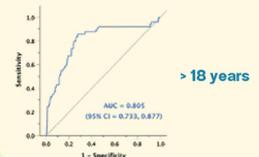
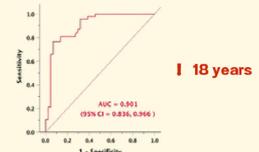
Longer Axial length

#### Age-specific AL cut-offs predict strabismus

! < 18 y : AL 27.5 mm (AUC 0.90)

! > 18 y : AL 29.0 mm (AUC 0.80)

#### AUROC for predict strabismus



SCAN FOR  
FULL TEXT



**SMJ**

SIRIRAJ  
MEDICAL  
JOURNAL

Surachatkumtonekul, et al. *Siriraj Med J* 2026;78(3):175-184.

©Siriraj Medical Journal. All rights reserved. Graphical abstract by T Surachatkumtonekul

\*Corresponding author: Thammanoon Surachatkumtonekul

E-mail: si95thim@gmail.com

Received 21 November 2025 Revised 20 January 2026 Accepted 20 January 2026

ORCID ID: <http://orcid.org/0000-0002-0037-6863>

<https://doi.org/10.33192/smj.v78i3.278985>



All material is licensed under terms of the Creative Commons Attribution 4.0 International (CC-BY-NC-ND 4.0) license unless otherwise stated.

## ABSTRACT

**Objective:** To determine the prevalence and predictive factors of strabismus in patients with high myopia and to evaluate the surgical outcomes of strabismus correction in this population.

**Materials and Methods:** This retrospective cohort study included 991 patients with high myopia (1,880 eyes) who attended Siriraj Hospital between January 2021 and June 2023. High myopia was defined as a spherical equivalent (SE)  $\leq -5.0$  diopters or an axial length (AL)  $\geq 26.5$  mm. Collected data included demographics, visual acuity, refractive error, axial length, strabismus type, and surgical outcomes. Surgical success was defined as postoperative alignment within 8 prism diopters (PD) at the 8-week follow-up. ROC curve analysis was performed to determine optimal AL cut-off values for predicting strabismus.

**Results:** Strabismus was identified in 6.7% (66/991) of patients. The mean SE and AL were  $-11.41 \pm 4.87$  D and  $28.35 \pm 2.06$  mm, respectively. Patients with strabismus were younger and had longer AL compared to those without strabismus ( $p < 0.01$ ). ROC analysis showed optimal AL thresholds of 27.5 mm for patients  $\leq 18$  years and 29.0 mm for patients  $> 18$  years, demonstrating high sensitivity, specificity, and predictive accuracy (AUC = 0.90 and 0.80, respectively). Surgical intervention was performed in 25 patients (37.9%), achieving a success rate of 76%.

**Conclusions:** Strabismus is more prevalent in individuals with high myopia than in the general population. Axial length is a strong predictor, and defined cut-offs can guide early detection. Surgical outcomes are favorable and comparable to those in non-myopic patients, supporting targeted management strategies in this population.

**Keywords:** Myopia; strabismus; prevalence; predictive value of tests; predictive factors; ophthalmologic surgical procedures (Siriraj Med J 2026;78(3):175-184)

## INTRODUCTION

Myopia, also known as nearsightedness, is a condition where distant objects appear blurry, while near objects remain clear. The World Health Organization defines myopia as a spherical equivalent (SE) of less than or equal to  $-0.5$  diopters (D), and high myopia as an SE of  $-5.0$  D or lower.<sup>1-3</sup> As the global prevalence of myopia continues to rise, it is projected that by 2050, nearly 9.8% of the world's population, or approximately 938 million individuals, will have high myopia.<sup>2,3,6</sup> High myopia is associated with several ocular pathologies, including retinal breaks, retinal detachment, myopic macular degeneration, glaucoma, and strabismus.<sup>4,5</sup>

Strabismus, or misalignment of the eyes, has an estimated prevalence of 2.5-2.9% in the general population, though rates vary by age, ethnicity, and geographic region.<sup>7</sup> In individuals with high myopia, strabismus occurs more frequently due to anatomical changes in the eye, particularly the elongation of the globe. Studies have shown increased prevalence of strabismus among high myopia patients compared to the general population, with studies showing varying results depending on the severity of myopia.<sup>8</sup>

The relevance of understanding this relationship is profound. Identifying predictive factors for strabismus in high myopia could lead to better preventive strategies, earlier interventions, and more targeted treatments, potentially

reducing the burden of visual impairment and improving surgical outcomes. Additionally, a better understanding of these factors could inform both clinicians and patients on the prognosis and management of strabismus in the context of high myopia.

This study aims to investigate the prevalence of strabismus in patients with high myopia, identifying predictive risk factors and evaluating surgical outcomes for strabismus correction in this patient group. In particular, we aim to examine how axial length (AL) serves as a key predictive factor for the development of strabismus in high myopia patients, with particular attention to age and refractive error as potential influencing factors.

## MATERIALS AND METHODS

### Study design

This investigation was a retrospective cohort study conducted at Siriraj Hospital between January 2021 and June 2023. The study was approved by the Siriraj Institutional Review Board (COA No. Si 751/2023) and adhered to the tenets of the Declaration of Helsinki. The sample size was calculated using the formula for estimating an infinite population proportion. The expected prevalence (P) of strabismus in high myopia was obtained from a previous study by *Tanaka et al.* A Z value of 1.96 was applied for a 95% confidence level, with an expected proportion (P) of 0.182 and a precision (d) of 0.025, resulting in a

required sample size of **916 participants**. To account for potential missing data or dropouts, an additional 10% was included, resulting in a total sample size of **1,008 participants**. The inclusion criteria consisted of patients diagnosed with high myopia, defined as spherical equivalent (SE) of -5.0 diopters (D) or less, or an axial length (AL) of 26.5 mm or greater. All participants underwent a detailed ophthalmic evaluation by an ophthalmologist, which included testing for best corrected visual acuity (BCVA), refractive error measurement, assessment of axial length, and evaluation of ocular alignment.

### Patient selection

Patient data were collected through a retrospective chart review. Individuals diagnosed with high myopia were identified using the ICD-10 code H44.2. During the study period, 1,676 patients were initially identified. Subsequently, patients who did not meet the inclusion criteria or had incomplete clinical data were excluded from the analysis. A total of 991 patients (1,880 eyes) with high myopia were included in this study. The inclusion process involved a review of medical records from Siriraj Hospital. The patients were assessed for demographic data such as age, sex, and clinical characteristics including BCVA, refractive error, and axial length. Particular attention was given to strabismus diagnoses and type of strabismus (e.g., esotropia, exotropia, etc.).

### Strabismus assessment

Strabismus was defined as any clinically diagnosed ocular misalignment. Strabismus type was classified based on the direction of deviation: horizontal (esotropia or exotropia), vertical (hypotropia or hypertropia), or combined horizontal and vertical strabismus. The degree of deviation was measured in prism diopters (PD) using standard clinical methods.

### Surgical treatment

Strabismus surgery was performed in 25 cases (37.9%). Most underwent horizontal muscle recession and resection (24 cases), and one patient underwent muscle transposition. Surgical success was defined as postoperative ocular alignment within 8 prism diopters at the 8-week follow-up.

### Outcome measures

The primary outcome was the prevalence of strabismus among patients with high myopia. Secondary outcomes included the relationship between axial length and strabismus, the effectiveness of strabismus surgery, and the correlation between axial length measurements and age subgroups ( $\leq 18$  years vs.  $>18$  years). Surgical

success was determined by postoperative alignment within 8 prism diopters (PD) at the 8-week follow-up.

### Statistical analysis

Descriptive statistics were used to summarize demographic and clinical data of the participants. Categorical variables were presented as frequencies and percentages, while continuous variables were summarized using means and standard deviations (SD). Comparisons between the strabismus and non-strabismus groups were performed using either the two-sample t-test or the Mann-Whitney U test, depending on data distribution. Receiver Operating Characteristic (ROC) curve analysis was utilized to assess the diagnostic accuracy of axial length as a predictor of strabismus. The optimal cut-off values for axial length, in terms of sensitivity and specificity, were determined for predicting strabismus in high myopia patients. A p-value of less than 0.05 was considered statistically significant. All statistical analyses were conducted using SPSS Statistics version 29.0.0 (IBM Corp., USA).

## RESULTS

### Demographic and clinical data

A total of 991 patients (1,880 eyes) were included in the analysis, with a mean age of  $45.80 \pm 19.80$  years. Among these, 664 (67.0%) were female. Most participants (89.7%) had bilateral myopia. The mean spherical equivalent (SE) was  $-11.41 \pm 4.87$  D, and the mean axial length (AL) was  $28.35 \pm 2.06$  mm. [Table 1](#) presents a summary of the demographic data for the study cohort.

### Prevalence of strabismus

Among the 991 patients, 6.7% (95% CI = 5.2% to 8.4%) were diagnosed with strabismus, and nine patients with strabismus (13.64%) were identified as having Heavy Eye Syndrome (HES). The most frequently observed type of strabismus was exotropia (47.0%), followed by esotropia (42.4%) and a combination of horizontal and vertical strabismus (6.0%). Only a small percentage of patients had vertical strabismus (4.5%), with hypertropia and hypotropia being the least common ([Table 2](#)).

### Comparison of strabismus and non-strabismus groups

[Table 1](#) compares the clinical characteristics of patients with and without strabismus. Patients in the strabismus group were significantly younger than those in the non-strabismus group ( $29.40 \pm 22.20$  years vs  $47.00 \pm 19.10$  years) ( $p < 0.01$ ). Furthermore, the axial length in the strabismus group was also significantly longer than in the non-strabismus group ( $29.78 \pm 0.21$  mm vs.  $28.11 \pm 0.64$  mm,  $p < 0.01$ ).

**TABLE 1.** Demographic data (Total number = 991 subjects, 1,880 eyes).

Characteristics of 991 patients	
Sex	
Female, n (%)	664 (67.0)
Laterality	
Bilateral, n (%)	889 (89.7)
Age (years), (Mean $\pm$ SD)	45.8 $\pm$ 19.8
Prevalence of strabismus in high myopia, n (%)	66 / 991 = 6.7% (95% CI = 5.2%–8.4%)

Characteristics of 1,880 eyes	n (eyes)	
BCVA (logMAR), mean $\pm$ SD	1,854	0.80 $\pm$ 0.24
Spherical equivalent (D), median (IQR)	1,716	-10.25 (-8.00, -14.00)
Axial length (mm), mean $\pm$ SD	848	28.30 $\pm$ 1.87

Comparison between strabismus and non-strabismus groups

Characteristics	n	Number (%)		p -value
		Strabismus (n = 66)	No Strabismus (n = 925)	
<b>Sex</b>				
Male	327	24 (7.3)	303 (92.7)	0.55
Female	664	42 (6.3)	622 (93.7)	
<b>Laterality</b>				
Unilateral	102	7 (6.9)	95 (93.1)	0.93
Bilateral	889	59 (6.6)	830 (93.4)	
Age (years), (Mean $\pm$ SD)	991	29.4 $\pm$ 22.2	47.0 $\pm$ 19.1	<0.01

Table 3 presents the comparison of visual acuity, spherical equivalent, and axial length. The strabismus group had a mean axial length of  $29.78 \pm 0.21$  mm, which was significantly longer than the non-strabismus group's mean of  $28.11 \pm 0.64$  mm ( $p < 0.01$ ). Additionally, the spherical equivalent was more myopic in the strabismus group (SE =  $-11.75 \pm 4.74$  D) compared to the non-strabismus group (SE =  $-10.13 \pm 4.84$  D); However, this difference was not statistically significant ( $p = 0.87$ ).

### ROC curve analysis

Receiver Operating Characteristic (ROC) curve analysis evaluated the diagnostic ability of axial length for predicting the presence of strabismus. In patients aged  $\leq 18$  years, the optimal axial length threshold was

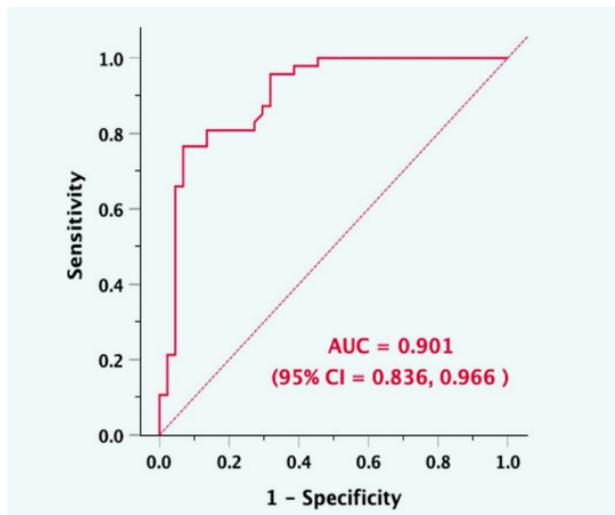
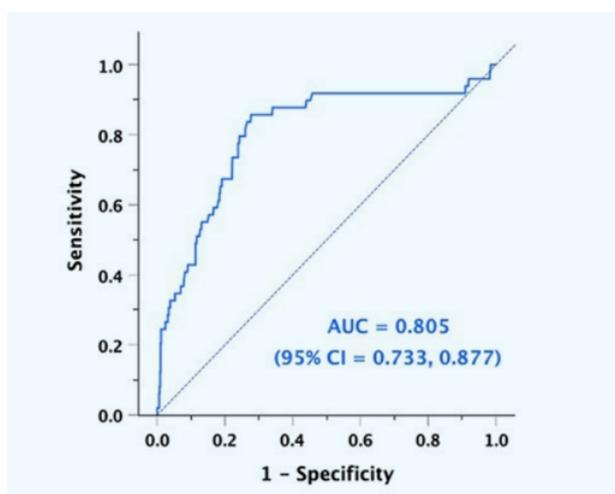
27.5 mm, with an AUC of 0.90, a sensitivity of 80.9%, and a specificity of 75% (Fig 1). For patients older than 18 years, the optimal cut-off value for axial length was 29.0 mm, with an AUC of 0.80, sensitivity of 85.7%, and a specificity of 70.2% (Fig 2). These results indicate that axial length is a reliable predictor of strabismus, with the highest diagnostic accuracy achieved at these cut-off points for both age groups [see Additional file 1].

### Surgical outcomes

Surgical intervention was performed in 25 patients (37.9%), including 24 horizontal recession and resection procedures and one muscle transposition. The overall surgical success rate, defined as postoperative alignment within 8 prism diopters (PD), was 76% (Table 4). The

**TABLE 2.** Clinical characteristics of the strabismus group.

Characteristics	Number (%)
<b>Strabismus type</b>	
Horizontal	
Esotropia	28 (42.4)
Exotropia	31 (47.0)
Vertical	
Hypotropia	0
Hypertropia	3 (4.5)
Horizontal with Vertical	
Exotropia with Hypertropia	2 (3.0)
Esotropia with Hypertropia	1 (1.5)
Esotropia with Hypotropia	1 (1.5)
<b>Deviation (PD), Median (IQR)</b>	
Horizontal	25 (16, 40)
Vertical	10 (5, 20)
<b>Treatment</b>	
Observe	41 (62.1)
Surgery	25 (37.9)

**Fig 1.** Receiver operating characteristic (ROC) curve for predicting strabismus in high myopia, Age  $\leq$  18 years, using axial length.**Fig 2.** Receiver operating characteristic (ROC) curve for predicting strabismus in high myopia, Age  $>$  18 years, using axial length.

**TABLE 3.** Comparison of clinical parameters between strabismus and non-strabismus groups.

Parameters	Strabismus		No Strabismus		p-value
	n (eyes)	value	n (eyes)	value	
BCVA (logMAR), mean ± SD					
≤18 years	72	0.74 ± 0.41	227	0.70 ± 0.43	0.72*
>18 years	43	0.82 ± 0.35	1,512	0.82 ± 0.41	0.68*
Total	115	0.62 ± 0.46	1,739	0.84 ± 0.42	0.43*
Spherical equivalent (D), median (IQR)					
≤18 years	67	-12.00(-15.50, -8.50)	179	-11.00(-15.00, -8.75)	0.82†
>18 years	48	-11.00(-15.38, -7.25)	1,422	-10.00(-14.00, -7.75)	0.69†
Total	115	-11.25(-15.50, -8.00)	1,601	-10.25(-14.00, -8.00)	0.10†
Axial length (mm), mean ± SD					
≤18 years	47	29.13 ± 1.77	44	26.36 ± 1.60	<0.01*
>18 years	49	30.40 ± 2.11	708	28.22 ± 1.72	<0.01*
Total	96	29.78 ± 0.21	752	28.11 ± 0.64	<0.01*

\* 2-sample t-test

† Mann-Whitney U test

**TABLE 4.** Surgical details and outcomes in patients with strabismus (n=25).

Characteristics	Number (%)
History of Previous strabismus surgery	
Yes	6 (24.0)
No	19 (76.0)
Type of surgery	
Recession	11 (44.0)
Resection	5 (20.0)
Combined (recession + resection)	6 (24.0)
Muscle transposition	1 (4.0)
Combined muscle surgery with intraoperative chemodenervation	2 (8.0)
Surgical outcome success	
Yes	19 (76.0)
No	6 (24.0)

Parameters, (Mean ± SD)	Pre-operative	Post-operative	p-value
BCVA (logMAR)	0.50 ± 0.48	0.51 ± 0.55	0.94
Spherical equivalent (D)	-11.75 ± 4.74	-10.13 ± 4.84	0.87
Horizontal deviation (PD)	33.61 ± 7.02	4.67 ± 16.33	<0.01

mean preoperative deviation improved significantly from  $33.61 \pm 7.02$  to  $4.67 \pm 16.33$  PD postoperatively ( $p < 0.01$ ). There were no significant variations in best-corrected visual acuity (BCVA) or spherical equivalent post-surgery.

### Subgroup analysis

In the age-based subgroup analysis, patients  $\leq 18$  years had a significantly shorter mean axial length compared to those  $> 18$  years old ( $29.13 \pm 1.77$  mm vs.  $30.40 \pm 2.11$  mm,  $p < 0.01$ ). However, the prevalence of strabismus was found to be higher in the younger cohort, suggesting that age plays a significant role in the development of strabismus in high myopia.

## DISCUSSION

In this study, the prevalence of strabismus among patients with high myopia was 6.7%, which is notably higher than the prevalence in the general population (1.42% to 2.15%).<sup>9-11</sup> This finding is consistent with previous research, including *Tanaka et al.*, who also reported a high prevalence of strabismus in high myopia, especially horizontal strabismus.<sup>8</sup> The higher prevalence noted in our study could be due to anatomical alterations linked to high myopia, such as increased axial length and globe elongation, which can lead to altered eye muscle function and misalignment.<sup>12</sup>

The higher prevalence of strabismus in younger patients may be partly due to the immaturity of binocular function during childhood. When combined with high myopia, this may lead to reduced visual acuity, which can increase the risk of developing amblyopia. In younger individuals, factors such as immature binocular function, amblyopia, axial length elongation, and changes in the soft tissues surrounding the globe may contribute to ocular misalignment.<sup>13</sup> On the other hand, in older individuals, high myopia may be a result of pathological myopia, which involves progressive axial elongation and ongoing globe enlargement, resulting in significantly greater axial length compared to younger individuals. This continued growth may contribute to more severe forms of ocular misalignment in older adults. Pathological myopia is associated with multiple ocular complications, such as retinal detachment, macular degeneration, and strabismus, due to the continuous elongation of the eye, affecting both the visual system and ocular muscle function.<sup>14</sup>

Our findings reinforce the idea that axial length is a key factor in the development of strabismus, with longer axial lengths being more prevalent. This finding aligns with the work of other researchers, including *Jonas et al.*,

who also found that longer axial length was a key factor in the pathophysiology of strabismus in high myopic eyes.<sup>4</sup>

### Predictive factors for strabismus

Axial length emerged as a strong predictor of strabismus in high myopia, with cut-off values of 27.5 mm for individuals aged  $\leq 18$  years and 29.0 mm for those older than 18 years. This is in line with previous research by *Nakao et al.*, who reported that an axial length greater than 28.0 mm in high myopia patients was linked to a higher risk of developing strabismus.<sup>15</sup> The ROC curve analysis conducted in our study further validates axial length as a reliable diagnostic tool for predicting strabismus in high myopia patients, demonstrating high sensitivity and specificity, particularly among younger patients.

In addition to axial length, our study revealed that younger age was also linked to an increased likelihood of strabismus. This result aligns with the findings of *Tanaka et al.*, who observed that children with high myopia were more likely to develop strabismus than adults.<sup>8</sup> Younger patients may have a higher likelihood of developing strabismus due to ongoing development of ocular structures and the pronounced impact of axial elongation during this period.

### Strabismus surgery outcomes

The surgical success rate in our study, defined as a postoperative alignment within 8 prism diopters (PD), was 76%. This result is comparable to other studies on strabismus surgery in high myopia patients, including those by *Kampanartsanyakorn et al.*, who reported a similar success rate for horizontal strabismus surgery.<sup>17</sup> The fact that our study's surgical success rate mirrors those of normal refractive strabismus surgeries indicates that high myopia-related strabismus can achieve similar surgical outcomes, provided that appropriate surgical techniques are used.

Interestingly, our study found that patients who underwent surgery had a significant reduction in horizontal deviation (from  $33.61 \pm 7.02$  PD preoperatively to  $4.67 \pm 16.33$  PD postoperatively). This also aligns with the findings of *Yetkin et al.*, who noted that horizontal strabismus surgeries in high myopic patients resulted in significant improvements in alignment.<sup>16</sup> Despite the promising surgical outcomes, it is important to recognize that surgical success may be influenced by several factors, including preoperative strabismus angle, muscle strength, and the presence of other ocular conditions such as amblyopia or anisometropia.<sup>18</sup>

All patients with unsuccessful surgical outcomes had a spherical equivalent more myopic than  $-9.50$  D, indicating that they likely had substantially greater axial elongation than average. Because currently available surgical dosage tables are largely derived from populations with normal axial length, they may underestimate the dose required in this subgroup. In patients with markedly increased axial length, defined as greater than 27.5 mm, or with very high myopia, consideration should be given to increasing the surgical dose or other surgical procedures in future practice.

### Anatomical mechanisms of strabismus in high myopia

Strabismus in high myopia is believed to arise primarily from several anatomical changes linked to axial elongation. One key mechanism is the displacement of the extraocular muscles as the globe elongates, leading to changes in muscle paths and mechanical function.<sup>15,19-24</sup> Prior studies by Yamaguchi et al. and Yokoyama et al. suggest that lengthening of the eye and the resulting shifts in muscle positioning can impair the balance of ocular muscle forces, contributing to strabismus development.<sup>12,20</sup>

In addition to muscle path changes, globe displacement and changes in the pulley system of the eye have also been suggested to be contributing factors in the development of strabismus in high myopia. These anatomical alterations are further exacerbated by visual impairment and the need for accommodation, which may put additional strain on the ocular muscles, leading to misalignment.<sup>21</sup>

Our findings further support this anatomical and pathophysiologic framework and are consistent with the mechanisms underlying Heavy Eye Syndrome (HES), a condition traditionally associated with extreme axial myopia. HES is characterized by superotemporal globe prolapse relative to the superior and lateral rectus muscles, resulting in severe esotropia and hypotropia. Previous studies have shown that in HES, the lateral rectus (LR) is displaced at a sharp angle of  $179.9^\circ \pm 30.8^\circ$  when compared to the superior rectus (SR), reflecting the mechanical displacement of the extraocular muscles as a result of axial elongation.<sup>25</sup> Although orbital imaging was not performed in our study, the strong association between longer axial length and increased strabismus prevalence observed in our cohort is consistent with the mechanical displacement of extraocular muscles described in HES. Previous MRI-based studies by Yamaguchi, Demer, and colleagues have demonstrated that progressive axial elongation leads to distortion of the muscle cone and degeneration of the LR–SR band, culminating in ocular misalignment typical of HES. Our results, therefore,

represent the epidemiologic and clinical counterpart of these anatomic mechanisms, suggesting that high axial length may serve as an early indicator of the same pathologic process that, in its advanced form, manifests as Heavy Eye Syndrome.<sup>12,21,26</sup>

Sagging eye syndrome (SES) may also represent a relevant pathophysiologic consideration in this population, particularly among older patients. SES is characterized by age-related degeneration and attenuation of the lateral rectus–superior rectus (LR–SR) band, resulting in inferior displacement of the lateral rectus pulley and a divergence-insufficiency pattern of esotropia. Studies comparing the displacement angles in SES revealed that the LR was displaced at a shallower angle of  $104^\circ \pm 11^\circ$  compared to the SR, suggesting a less severe degree of displacement than that seen in HES.<sup>25</sup> Unlike heavy eye syndrome, which is typically associated with marked axial elongation and superotemporal globe prolapse, SES can occur even in eyes without extreme axial myopia and may therefore account for strabismus in selected patients whose axial length measurements are less remarkable. Although our study did not incorporate orbital imaging to differentiate mechanical etiologies, the coexistence of axial elongation and age-related connective tissue changes likely represents a continuum of anatomical alterations that can influence ocular alignment.<sup>21,27</sup>

Understanding these anatomical mechanisms provides a crucial framework for interpreting our findings, emphasizing that early identification of high axial length may help prevent progression to severe mechanical strabismus requiring complex surgical correction.

### Limitations

This study has several limitations that warrant consideration. The retrospective design introduces inherent selection bias and restricts the ability to infer causal relationships between axial length and the development of strabismus. Because the study was conducted at a single tertiary referral center, the cohort may include a higher proportion of complex cases, which could limit the generalizability of the findings to other clinical settings. In addition, although axial length emerged as a strong predictor of strabismus, orbital imaging was not routinely performed at Siriraj Hospital because of cost constraints and long waiting times. As a result, orbital imaging was not obtained for all patients, precluding direct assessment of the anatomical mechanisms underlying ocular misalignment, including potential extraocular muscle displacement and features related to Heavy Eye Syndrome and Sagging Eye Syndrome. Collectively, these factors underscore the need for prospective, multicenter

investigations with standardized imaging protocols to more clearly delineate the structural pathways and clinical implications of strabismus in high myopia.

## CONCLUSION

The prevalence of strabismus among patients with high myopia was 6.7%, notably higher than that seen in the general population. This emphasizes the elevated risk of strabismus in individuals with high myopia and highlights the need for early identification and intervention. Axial length was identified as a strong predictive factor for strabismus, with cut-off values of 27.5 mm for individuals aged  $\leq 18$  years and 29.0 mm for those over 18 years. These findings provide valuable insights into how axial length can be utilized as a diagnostic tool to identify high-risk individuals.

The observed association between increasing axial length and higher strabismus prevalence supports the proposed pathophysiologic continuum of Heavy Eye Syndrome (HES). Progressive globe elongation may initiate subtle extraocular muscle displacement that precedes the severe mechanical misalignment characteristic of HES.

The surgical success rate of 76% observed in this study indicates that strabismus surgery in high myopia patients can achieve results comparable to those seen in non-myopic populations. This reinforces the notion that surgical intervention can be an effective treatment modality for strabismus in high myopia, even though the outcomes may be influenced by factors such as age and preoperative strabismus angle.

Although the findings of this study provide important insights into strabismus in high myopia, the retrospective design and lack of a control group limit the ability to establish a clear cause-and-effect link between axial length and strabismus. Future prospective studies incorporating advanced imaging techniques, such as MRI, to evaluate the anatomical changes in the eye and extraocular muscles, as well as including control groups, will be crucial for further validating these findings and exploring the underlying mechanisms in greater detail.

## Data Availability Statement

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

## ACKNOWLEDGEMENTS

We would like to acknowledge Piyaphat Jaruniphakul for creating all figures and tables. Our sincere thanks to Suchawadee Leelasrisoonton for preparing the manuscript

for publication. We also acknowledge Asst. Prof. Dr. Chulaluk Komoltri for conducting the statistical analyses.

## DECLARATIONS

### Grants and Funding Information

This research project does not have any funding allocated.

### Conflict of Interest

The authors declare that they have no competing interests.

### Registration Number of Clinical Trial

Not applicable. This was a retrospective study and was not registered as a clinical trial.

### Authors' Contributions

Conceptualization and methodology, T.S.; Investigation, Data collection and Data analysis, P.J.; Writing-original draft preparation, review and editing, T.S., P.J.; Supervision, T.S. All authors read and approved the final manuscript.

### Use of Artificial Intelligence

During preparation of the manuscript, the authors used Grammarly and ChatGPT 5.2 to refine grammar and enhance readability.

### Ethics Approval and Consent to Participate

The study was approved by the Siriraj Institutional Review Board (COA No. Si 751/2023).

## REFERENCES

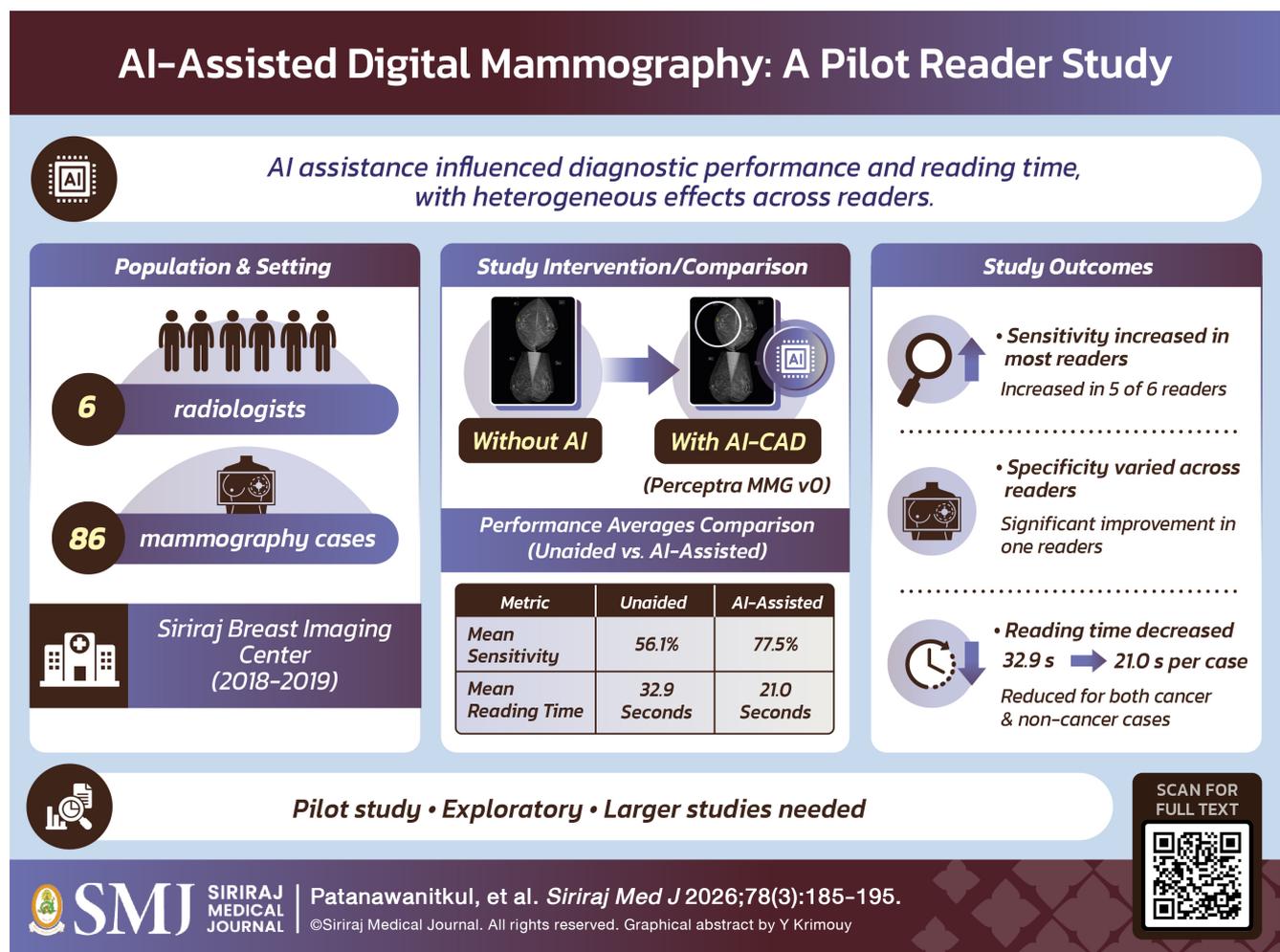
1. World Health Organization. The High Impact of Myopia and High Myopia. WHO; 2017. Accessed Nov. 14, 2019.
2. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology*. 2016;123(5):1036-42.
3. Surachatkumtonekul T. Strabismus in high myopia. In: Wongsakittirak S, Somkijrungrroj T, editors. *Thai Academy of Ophthalmology (Jaksu Vichakan) by the Royal College of Ophthalmologists of Thailand, Vol. 7*. Bangkok: Thammasat University Press; 2023. p. 153-66.
4. Jonas JB, Bikbov MM, Kazakbaeva GM, Wang YX, Xu J, Nangia V, et al. Positive and Negative Associations of Myopia with Ocular Diseases in Population-Based Studies. *Ophthalmology*. 2024;131(12):1427-35.
5. Haarman AEG, Enthoven CA, Tideman JW, Tedja MS, Verhoeven VJM, Klaver CCW. The Complications of Myopia: A Review and Meta-Analysis. *Invest Ophthalmol Vis Sci*. 2020; 61(4):49.
6. Foster PJ, Jiang Y. Epidemiology of myopia. *Eye (Lond)*. 2014;28(2):202-8.

7. Fieß A, Elflein HM, Urschitz MS, Pesudovs K, Münzel T, Wild PS, et al. Prevalence of Strabismus and Its Impact on Vision-Related Quality of Life. *Ophthalmology*. 2020;127(8):1113-22.
8. Tanaka A, Ohno Matsui K, Shimada N, Hayashi K, Shibata Y, Yoshida T, et al. Prevalence of strabismus in patients with pathologic myopia. *J Med Dent Sci*. 2010;57(1):75-82.
9. Jenchitr W, Jaradaroonchay M, Sakulsiritiwakorn N. Strabismus And Amblyopia In Thailand. *Interprof J Health Sci* [internet]. 2023 Oct. 5 [cited 2025 May 14];11(1):38-43. Available from: <https://li05.tci-thaijo.org/index.php/IJHS/article/view/9>
10. Hashemi H, Pakzad R, Heydarian S, Yekta A, Aghamirsalim M, Shokrollahzadeh F, et al. Global and regional prevalence of strabismus: a comprehensive systematic review and meta-analysis. *Strabismus*. 2019;27(2):54-65.
11. Miyata M, Kido A, Miyake M, Tamura H, Kamei T, Wada S, et al. Prevalence and Incidence of Strabismus by Age Group in Japan: A Nationwide Population-Based Cohort Study. *Am J Ophthalmol*. 2024;262:222-8.
12. Yamaguchi M, Yokoyama T, Shiraki K. Surgical Procedure for Correcting Globe Dislocation in Highly Myopic Strabismus. *Am J Ophthalmol*. 2010;149(2):341-6.e2.
13. Adams GGW, Sloper JJ. Update on squint and amblyopia. *J R Soc Med*. 2003;96(1):3-6.
14. Du R, Xie S, Igarashi Yokoi T, Watanabe T, Uramoto K, Takahashi H, et al. Continued increase of axial length and its risk factors in adults with high myopia. *JAMA Ophthalmol*. 2021;139(10):1096-103.
15. Nakao Y, Kimura T. Prevalence and anatomic mechanism of highly myopic strabismus among Japanese with severe myopia. *Jpn J Ophthalmol*. 2014;58(2):218-24.
16. Yetkin AA. Factors Affecting Surgical Success Rates in Pediatric Horizontal Strabismus Surgery. *Cureus*. 2024;16(11):e74758.
17. Kampanartsanyakorn S, Surachatkumtonekul T, Dulayajinda D, Jumroendarasmee M, Tongsaee S. The outcomes of horizontal strabismus surgery and influencing factors of the surgical success. *J Med Assoc Thai*. 2005;88 Suppl 9:S94-9.
18. Surachatkumtonekul T, Tongsaee S, Sathianvichitr K, Sangsre P, Saiman M, Sermsripong W, et al. Corneal curvature change after strabismus surgery: an experience from a single academic center. *Siriraj Med J*. 2024;76(10):713-21.
19. Demer JL. Muscle paths matter in strabismus associated with axial high myopia. *Am J Ophthalmol*. 2010;149(2):184-186.e1.
20. Yokoyama T, Tabuchi H, Ataka S, Shiraki K, Miki T, Mochizuki K. The mechanism of development in progressive esotropia with high myopia. In: *Proceedings of the transactions of the 26th meeting of European strabismological Association* (ed Faber JT De), Barcelona, 14–16 September 2000.p.218-21.
21. Tan RJD, Demer JL. Heavy eye syndrome versus sagging eye syndrome in high myopia. *J AAPOS*. 2015;19(6):500-6.
22. Aoki Y, Nishida Y, Hayashi O, Nakamura J, Oda S, Yamade S, et al. Magnetic resonance imaging measurements of extraocular muscle path shift and posterior eyeball prolapse from the muscle cone in acquired esotropia with high myopia. *Am J Ophthalmol*. 2003;136(3):482-9.
23. Krzizok TH, Schroeder BU. Measurement of recti eye muscle paths by magnetic resonance imaging in highly myopic and normal subjects. *Invest Ophthalmol Vis Sci*. 1999;40:2554-60.
24. Monga S, Kekunnaya R, Sachdeva V. Exotropia-hypotropia complex in high myopia. *J Pediatr Ophthalmol Strabismus*. 2013;50(6):340-6.
25. Surachatkumtonekul T, Wangpaitoon C, Sermsripong W. Binocular Diplopia After Cataract Surgery: Incidence and Associated Factors in a Tertiary Teaching Eye Center. *Siriraj Med J*. 2026;78(2):133–141. doi:10.33192/smj.v78i2.277766.
26. Phamonvaechavan P, Saksiriwutto P. Surgical treatment of myopic strabismus fixus by loop myopexy augmented with scleral fixation: a case report. *Siriraj Med J*. 2019;71(4):318-21.
27. Chaudhuri Z, Demer JL. Sagging eye syndrome: connective tissue involution as a cause of horizontal and vertical strabismus in older patients. *JAMA Ophthalmol*. 2013;131(5):619-25.

# Diagnostic Performance of AI-CAD Digital Mammography for Breast Cancer: Experience from Siriraj Breast Imaging Center

Rujira Patanawanitkul, M.D., Voraparee Suvannarerg, M.D., Shanigarn Thiravit, M.D., Kobkun Muangsomboon, M.D., Pornpim Korpraphong, M.D.\*

Department of Radiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.



\*Corresponding author: Pornpim Korpraphong

E-mail: Pornpim.keaw@gmail.com

Received 4 September 2025 Revised 27 January 2026 Accepted 28 January 2026

ORCID ID: <http://orcid.org/0000-0002-8435-1814>

<https://doi.org/10.33192/smj.v78i3.277476>



All material is licensed under terms of the Creative Commons Attribution 4.0 International (CC-BY-NC-ND 4.0) license unless otherwise stated.

## ABSTRACT

**Objective:** To evaluate the diagnostic performance of radiologists with varying breast imaging experience when interpreting digital mammograms with and without artificial intelligence (AI). This work represents the initial phase of an AI development program at Siriraj Hospital, aiming toward broader integration of AI into breast cancer detection and clinical practice in Thailand.

**Materials and Methods:** In this retrospective study, six radiologists independently reviewed 86 digital mammograms — 40 confirmed cancer cases and 46 normal cases (including 28 false positives and 18 true negatives) — collected between 2018 and 2019 at the Siriraj Breast Imaging Center. Each radiologist interpreted all cases twice: unaided and AI-assisted, with a two-week washout period to minimize recall bias. Diagnostic performance metrics included sensitivity, specificity, false positive/negative rates, and reading time.

**Results:** With AI assistance, sensitivity increased in five of six readers, with mean sensitivity rising from 56.1% to 77.5%, although this difference did not reach statistical significance. Changes in specificity were variable across readers, with a statistically significant improvement observed in one reader (52.2% to 78.3%,  $P < 0.05$ ). Mean reading time decreased from 32.9 seconds to 21.0 seconds per case with AI assistance ( $P < 0.01$ ), with reductions observed for both cancer cases and normal cases.

**Conclusion:** In this pilot study, AI assistance was associated with trends toward improved diagnostic performance and reduced reading time, with statistically significant improvement observed in only a subset of readers. These preliminary findings require confirmation in larger, adequately powered multi-reader multi-case (MRMC) studies.

**Keywords:** Breast neoplasms; mammography; artificial intelligence; computer-aided detection; observer variation (Siriraj Med J 2026;78(3):185-195)

## INTRODUCTION

Breast cancer is the most prevalent malignancy among women worldwide, and remains the leading cause of cancer-related deaths in women.<sup>1,2</sup> Digital mammography (DM) is the main imaging modality for breast cancer screening in asymptomatic women and for diagnostic evaluation in symptomatic women. It facilitates early detection of breast cancer and has been shown in numerous randomized clinical trials to reduce mortality.<sup>3-5</sup>

In standard practice, a radiologist examines mammograms and classifies the findings using the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) lexicon.<sup>6</sup> Atypical findings on DM usually require further diagnostic workup, which may include additional mammographic views or additional imaging modalities. According to *Lehman et al*,<sup>7</sup> the typical performance metrics for screening mammography by a radiologist are a sensitivity of 86.9% and a specificity of 88.9%.

However, interpreting mammograms is challenging due to subtle differences between lesions and background fibroglandular tissue, variations in lesion types, the non-rigid nature of the breast, differences in radiologists' experience, and the relatively low prevalence of cancer in average-risk screening populations. This results in significant intraobserver and interobserver variations.<sup>8,9</sup>

In women with dense fibroglandular tissue, the false-negative rate can range from 16% to 31%.<sup>10-14</sup> As a result, reducing missed diagnoses and interpretation mistakes are critical to improving diagnostic accuracy in digital mammography.

Some European countries use double reading of mammograms to improve lesion detection and interpretation. This procedure has been shown to enhance sensitivity by 8%-14% and specificity by 4%-10%, respectively.<sup>15</sup> However, the high volume of women screened, combined with the adoption of double reading, creates a high workload that threatens efficiency, particularly given the growing shortage of qualified screening radiologists.<sup>16</sup>

To support radiologists and enhance human detection accuracy, computer-aided detection (CAD) systems have been introduced. However, so far, no study has demonstrated any direct improvement in screening outcomes.<sup>17-20</sup> Most evidence indicates no difference in cost-effectiveness, primarily due to the low specificity of these systems. Traditional CAD systems typically offer 84% sensitivity but only 13% specificity, leading to a high rate of false negatives, increased recall rates, and an inability to alleviate radiologist workload.<sup>18,20,21</sup>

Recent advancements in AI, particularly the use of deep learning algorithms, are decreasing the performance gap between humans and machines in many medical imaging applications.<sup>22</sup> Deep learning-based algorithms

for mammography evaluation have exhibited stand-alone performance comparable to radiologists and have significantly improved radiologists' diagnostic accuracy when used as decision support tools in breast cancer screening.<sup>22-26</sup>

However, significant advances in artificial intelligence (AI), particularly with deep convolutional neural networks (CNNs — also known as deep learning algorithms) are narrowing the performance gap between humans and computers across various medical imaging applications, including breast cancer detection. This new generation of deep learning-based CAD systems has the potential to enhance the effectiveness.<sup>26-30</sup>

While recent studies have primarily focused on populations in Europe and the United States, important differences, such as breast density and other demographic characteristics, exist between those populations and the Thai population.<sup>27,31-34</sup> Therefore, there is a clear need for population-specific research in Thailand. Moreover, the cost of commercial AI systems remains high. To address this, Siriraj Hospital has initiated the development of a locally trained artificial intelligence system for Thais called Perceptra MMG version 0, which uses a database derived from Thai patients.

This study aims to assess whether the program can assist radiologists with early detection, raise the standard of breast cancer diagnosis in the Thai public health system, and increase the quality of life of Thai women. As a result, the goal of this study is to assess the diagnostic performance of radiologists reading mammograms unaided versus assisted by the Perceptra MMG version 0 AI system in a Thai population.

## MATERIALS AND METHODS

### Study design and populations

This retrospective study was conducted at the Siriraj Breast Imaging Center, using data collected over a two-year period from January 2018 to December 2019. Institutional Review Board approval was obtained (certificate of approval number Si 841/2022). The requirement for informed consent was waived due to the retrospective and anonymized nature of the data. Anonymized digital mammograms were obtained from both screening and diagnostic exams. Mammograms with BI-RADS categories 1–5 were included if image quality was deemed adequate and clinical/histopathologic outcomes were available. Exclusion criteria included history of breast cancer, history of breastfeeding, history of breast reconstruction, breast implant, any substance/foreign body injection, and poor image quality.

All eligible examinations were classified as true positive, false positive, or true negative. True positives were confirmed by histopathology. False positives were defined by benign biopsy or no malignancy after 24 months of follow-up. True negatives were verified by negative findings at 24-month follow-up. A radiologist not involved in the reading sessions reviewed all cases to exclude those with poor image quality or overt features to minimize recall bias. The final data set used 86 mammographic cases (40 were true-positive cancer cases and 46 negative cases, of which 28 were false positive results and 18 true negative results). [Tables 1 and 2](#) present the demographic characteristics of the selected patient cohort.

**TABLE 1.** Demographic and clinical characteristics of the study population and selected digital mammographic examinations.

Characteristic	All (86)	Disease (40)	Non-disease (46)	P-value
Age (years)				<0.05
Mean	54.62	57.7	51.93	
Median	54	58	50.5	
Range	33-85	37-85	33-85	
Interquartile range	15.5	15.75	16.5	
Breast density				0.517
Almost entirely fat	3.5% (3/86)	5 (2/40)	2.2 (1/46)	
Scattered areas of fibroglandular density	9.3% (8/86)	12.5 (5/40)	6.5 (3/46)	
Heterogeneously dense	75.6% (65/86)	75 (30/40)	76.1 (35/46)	
Extremely dense	11.6% (10/86)	7.5 (3/40)	15.2 (7/46)	

**TABLE 2.** Characteristics of the 40 malignant cancers in the selected dataset.

Characteristic	Number of cases
<b>Histologic type</b>	
Invasive ductal carcinoma	22 (55.0 %)
Ductal carcinoma in situ	15 (37.5%)
Invasive lobular carcinoma	1 (2.5%)
Other *	2 (5.0%)
<b>Lesion type</b>	
Mass	23 (57.5%)
Calcification	14 (35.0%)
Asymmetry	1 (2.5%)
Architectural distortion	2 (5.0%)

\* Other included 2 cases of papillary carcinoma and 1 case of lymphoma

### AI system

The AI system was called Perceptra MMG version 0, which was specifically developed for breast cancer detection on 2-D digital mammograms. This AI system was trained using data from 94,817 cases collected between 2007 and 2021, sourced from the BIS (Breast Information System), which includes details of mass characteristics, calcifications, breast density, ultrasound findings, and BI-RADS categories. From this database, 10,650 mammographic examinations were selected for AI development, including cases with reported abnormalities and negative examinations, based on BIS report categorization. These cases were randomly divided into a training set (7,388 cases), a validation set (1,681 cases), and internal test set (1,581 cases). Malignancy status was determined using histopathology for cancer cases and benign pathology or follow-up imaging for non-malignant cases. In the test set evaluation, the stand-alone AI system demonstrated: Area under the receiver operating characteristic curve (AUC) was 0.968, sensitivity and specificity of 94.3% and 88.3%, PPV of 94.8% and NPV of 87.4%. This program is an in-house AI software system developed by Siriraj Hospital. This research was conducted with permission from the development team, without personal financial conflict of interest and external funding. The Perceptra MMG v0 model was finalized and locked prior to initiation of the reader study. All 86 mammographic examinations used in the reader study were explicitly excluded from

the AI training, validation, and internal testing datasets. Case identifiers were cross-checked to ensure no overlap between AI development data and reader-study cases. The AI system was evaluated as a decision-support tool to assist radiologists during mammography interpretation and was not intended to replace human readers. This study is reported in accordance with the CLAIM and STARD-AI reporting guidelines.

This AI software generated separate gray-scale images for each view (craniocaudal [CC] and mediolateral oblique [MLO]) of each breast. These images included: an overall per-breast abnormality score (0% to 100%) and a heatmap highlighting areas of abnormality using lines of varying thickness to indicate the probability of abnormality. If multiple areas of abnormality were identified, the region with the highest abnormality score was displayed at the bottom of the screen.

### Reader test

The reading sessions involved six radiologists. Two were certified by the Royal College of Radiologists of Thailand and subspecialized in breast imaging, with 10 and 4 years of experience, respectively, each interpreting more than 3,000 mammograms annually. Two radiologists were breast imaging fellows, and two were diagnostic radiology residents.

Readers were blinded to clinical history, prior imaging, and final diagnosis. During the unaided session, no AI outputs were visible; in the AI-assisted session, radiologists viewed AI heatmaps and abnormality scores in a separate but synchronized dual-display setting. For each case, readers reported the location and type of the most suspicious abnormality when a suspicious finding was identified. Interpretation time was recorded individually for each reader and defined as the interval from initial display of the case to the reader's final case-level decision, indicated by selection of either "malignant" or "benign". Although the time was recorded, the measurement was hidden from the readers. The initial reading session was conducted without the AI system, and, after a two-week washout period, the readers re-evaluated the mammograms with the AI system. The washout interval was intended to minimize potential bias from prior exposure to AI usage.<sup>35</sup> The reading environment remained identical for all readers across both sessions.

### Statistical analysis

The main objective of this pilot study was to explore changes in diagnostic performance at the individual reader level between unaided interpretation and AI-assisted reading, using sensitivity, specificity, and reading time

as performance metrics. For each reader, differences in sensitivity and specificity across conditions were analyzed using McNemar's test, with two-sided 95% confidence intervals and corresponding P-values. Reading time per case was automatically measured by the workstation software used for observer evaluation. Reading time differences were analyzed using paired comparisons between unaided and AI-assisted readings. All analyses were performed using PASW Statistics version 18 (formerly SPSS Statistics, SPSS Inc., Chicago, IL, USA). A P-value of less than 0.05 was considered statistically significant.

## RESULTS

### Sensitivity and specificity changes with use of AI

Sensitivity increased for five of six readers when using AI and remained unchanged for the most experienced reader, although none of these changes reached statistical significance (Table 3). Specificity increased for four of six readers, with Reader 2 (second-most experienced) showing the largest improvement (+0.261 when using AI support; P-value < 0.05). An example of a true-negative case is shown in Fig 1.

### Cancer-detection rate, false-positive and false-negative rate changes with use of AI

AI use was associated with a trend toward higher cancer detection rates in five of six readers, with no improvement in the most experienced reader, who already

had the highest cancer detection rate. Similarly, the use of AI reduced the false-positive rate in four of six readers, with Reader 2 again showing the largest reduction, with a decrease of approximately 26.1%.

Five out of six readers showed a trend toward reduced false-negative rates with AI. The most experienced reader showed no change but maintained the lowest false-negative rate without AI. Fig 2 summarizes cancer detection rates, false-positive rates, and false-negative rates, along with the percentage improvement from AI use. Examples of false-negative and false-positive cases are shown in Figs 3 and 4, respectively.

An illustration of a true-positive case in which AI reduced the false-negative rate is shown in Fig 5. Without AI, only the most experienced reader detected the cancer; however, with AI, all readers were able to detect a suspicious lesion.

### Reading time changes with use of AI

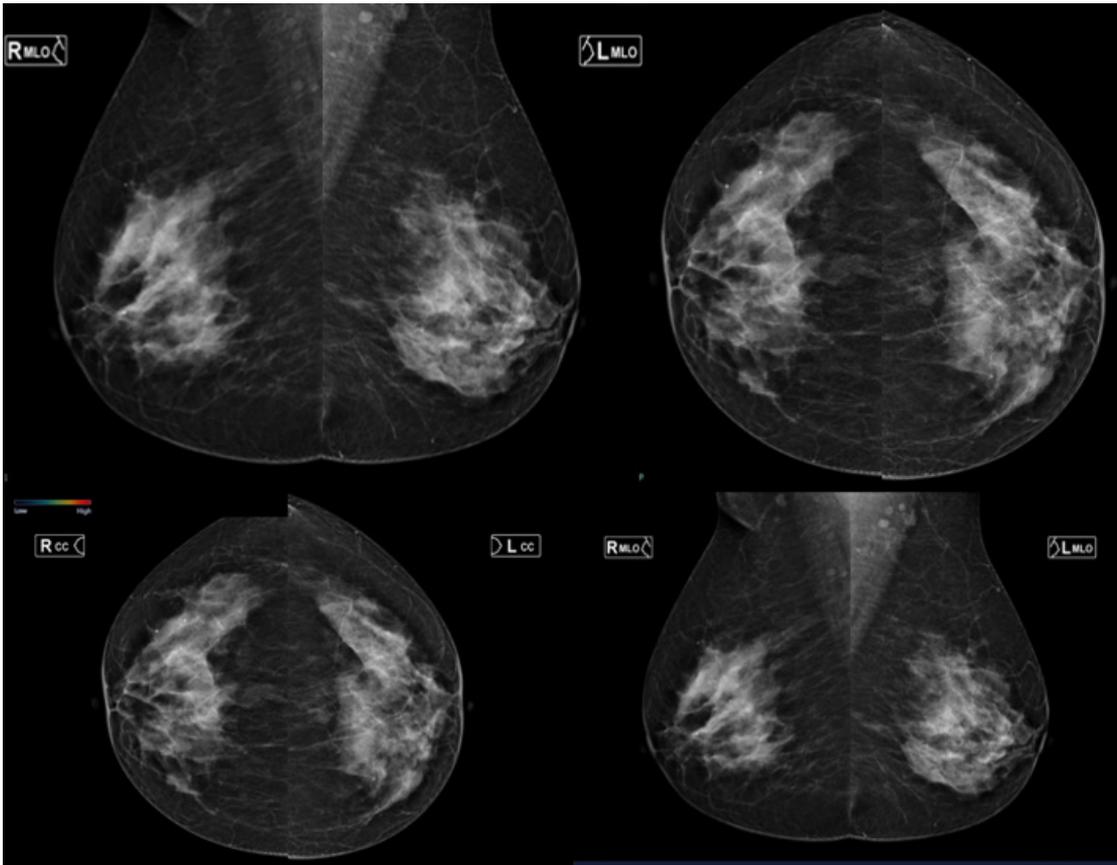
Table 4 shows the overall average reading times, which demonstrated a statistically significant reduction with AI support (P-value < 0.01).

During the first reading session (unaided readings), the average reading time was 27.98 seconds for cancer cases (95% CI: 24.75, 31.21) and 37.22 seconds for negative cases (95% CI: 33.21, 41.24). In the second session (with AI support), the average reading time was 19.07 seconds for cancer cases (95% CI: 17.19, 20.95), and 22.66 seconds

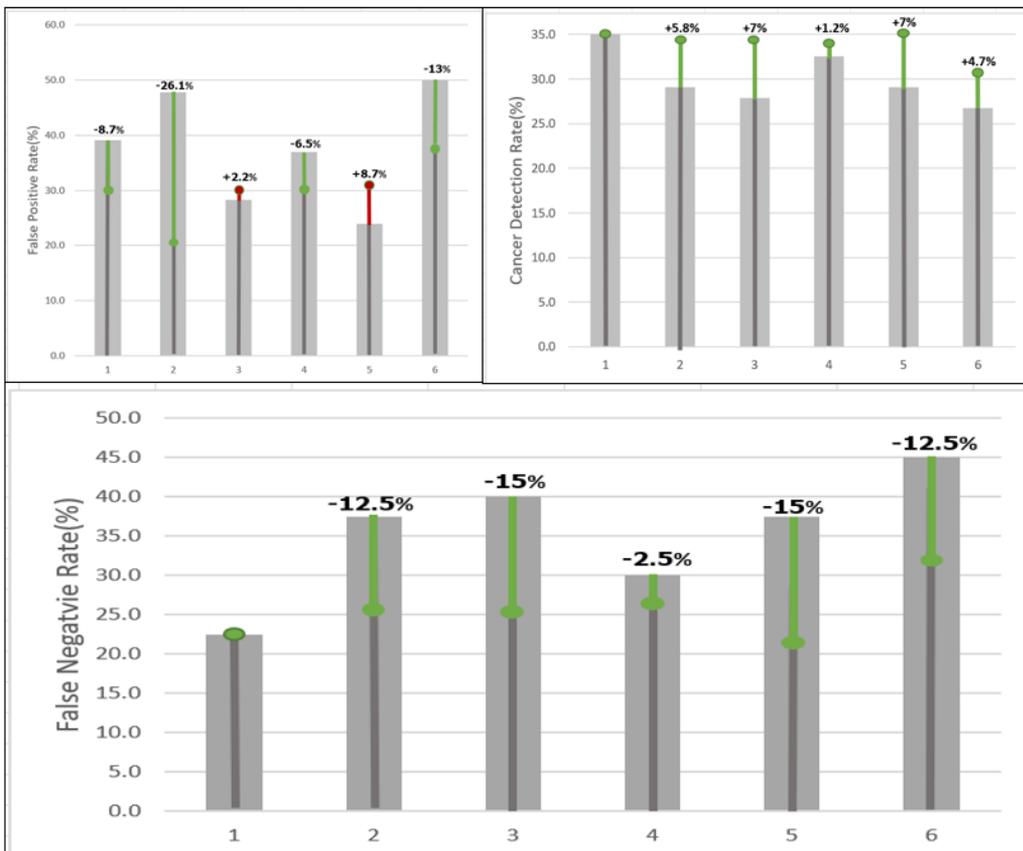
**TABLE 3.** Sensitivity and specificity of each reader.

Reader	Parameter	R	R+A	Δ	P-value
1	Sensitivity	0.775 (0.615, 0.891)	0.775 (0.615, 0.891)	0.000 (-0.157, 0.157)	1
	Specificity	0.608 (0.453, 0.749)	0.695 (0.542, 0.823)	0.087 (-0.263, 0.089)	0.381
2	Sensitivity	0.625 (0.458, 0.773)	0.75 (0.588, 0.873)	0.125 (-0.281, 0.032)	0.228
	Specificity	0.522 (0.369, 0.671)	0.783 (0.636, 0.891)	0.261 (-0.418, -0.104)	< 0.05
3	Sensitivity	0.600 (0.433, 0.751)	0.75 (0.588, 0.873)	0.080 (-0.307, 0.007)	0.152
	Specificity	0.717 (0.565, 0.840)	0.695 (0.542, 0.825)	-0.080 (-0.134, 0.178)	0.819
4	Sensitivity	0.700 (0.535, 0.834)	0.725 (0.561, 0.854)	0.025 (-0.180, 0.130)	0.805
	Specificity	0.630 (0.475, 0.767)	0.696 (0.542, 0.823)	0.066 (-0.228, 0.096)	0.508
5	Sensitivity	0.625 (0.458, 0.772)	0.775 (0.615, 0.892)	0.066 (-0.228, 0.096)	0.143
	Specificity	0.761 (0.612, 0.874)	0.674 (0.520, 0.805)	-0.087 (-0.240, 0.066)	0.256
6	Sensitivity	0.561 (0.397, 0.715)	0.675 (0.509, 0.814)	0.114 (-0.281, 0.053)	0.291
	Specificity	0.489 (0.331, 0.642)	0.630 (0.475, 0.768)	0.141 (-0.311, 0.029)	0.174

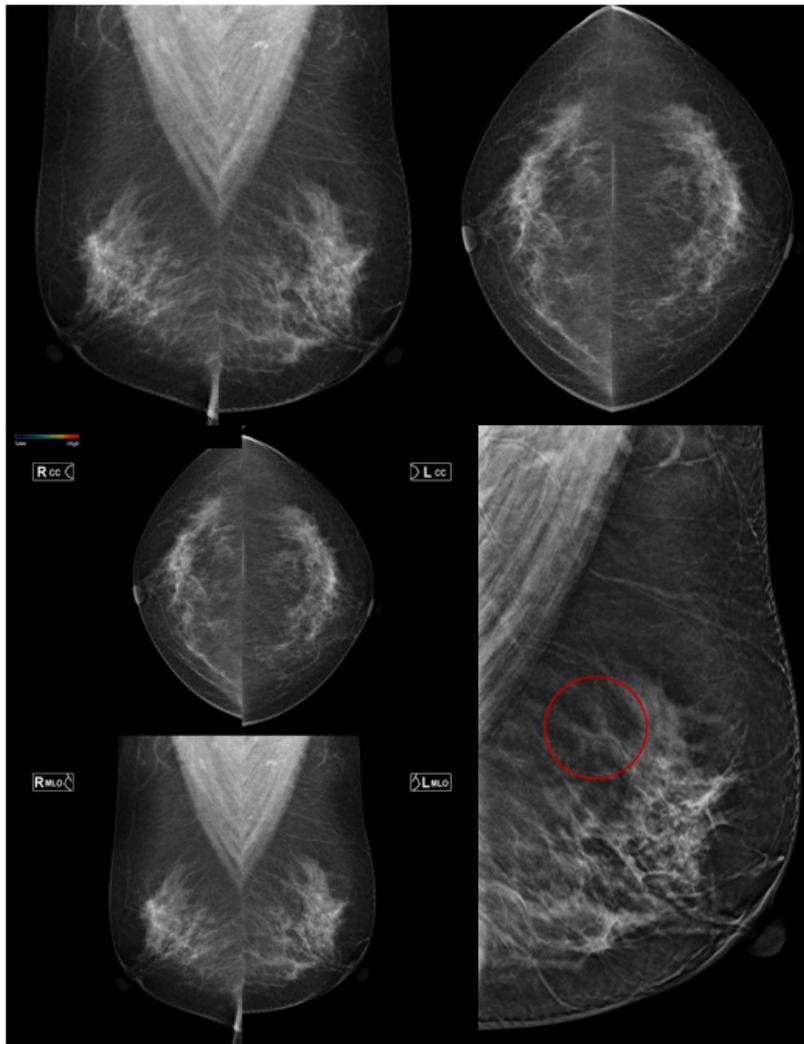
**Note:** Sensitivity and specificity for reading conditions with (R+A) and without (R) the AI system. Numbers in parentheses are 95% CI values.



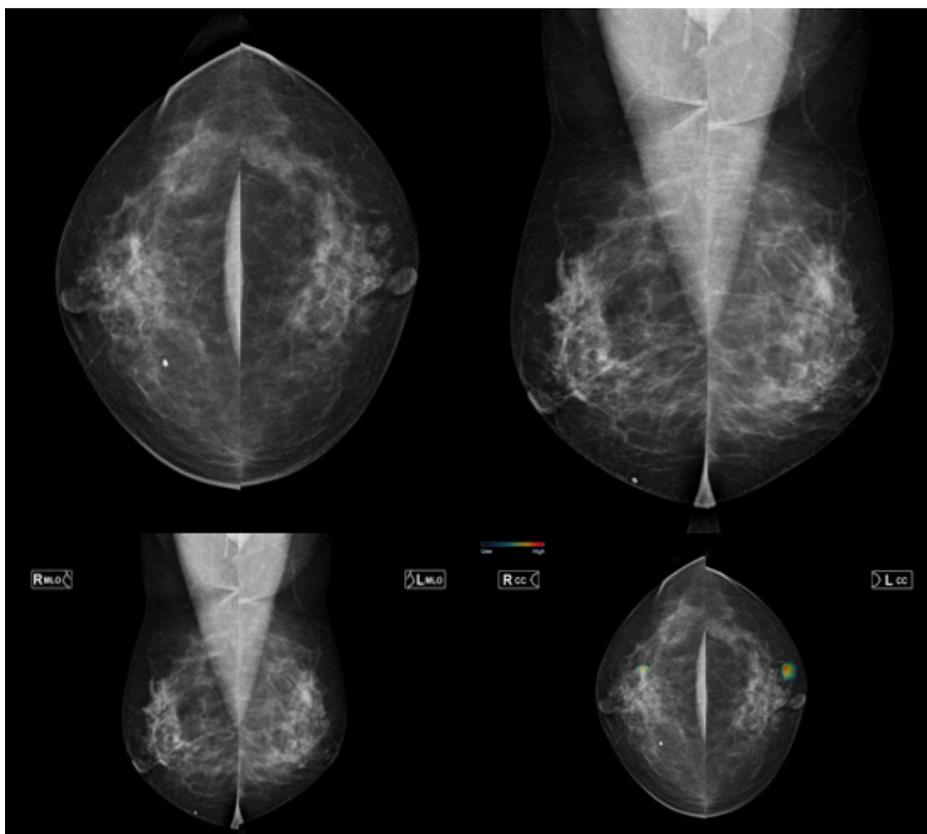
**Fig 1.** A 54-year-old woman without cancer. This case is one of the true-negative cases included in the dataset. No outlined areas and scores are shown in the viewer of AI system. In the unassisted reading session, three readers localized a suspicious lesion. However, after using the AI system for assistance, all readers changed their interpretation to no lesion.



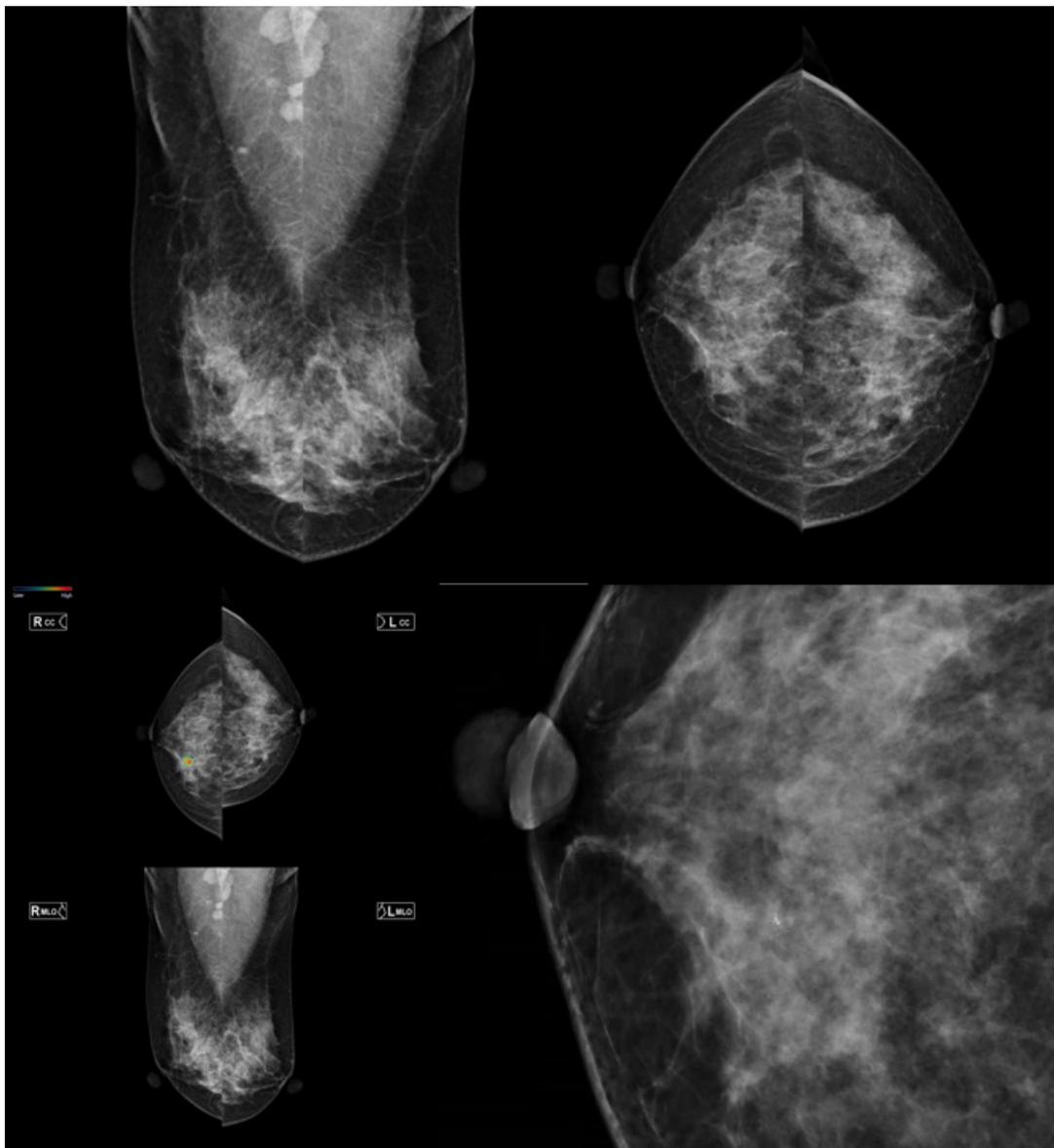
**Fig 2.** Cancer detection rate and percentage improvement brought by the use of artificial intelligence (AI) system: false-positive and false-negative rates, with percentage decreases as a result of AI use. Green bars represent percentage improvement brought about by the AI system, resulting in an increased cancer detection rate, a decreased false-positive rate in four readers, and no increase in false-negative rate in all readers.



**Fig 3.** A 52-year-old woman with invasive ductal carcinoma. This case is one of the false-negative ones included in the dataset. No outlined areas are shown in the viewer of the AI system. All the readers missed the lesion during the unaided and aided reading session. Her mammogram revealed subtle spiculated mass in the left breast's UIQ (Circle), which was more clearly visible after tomosynthesis.



**Fig 4.** A 53-year-old woman without cancer. This case is one of the false-positive cases included in the dataset. The AI system's viewer displays the outline area and abnormality score. In the unassisted reading session, three readers did not localize a suspicious lesion. However, after using the AI system for assistance, they identified a suspicious lesion. Serial follow-up over two years confirmed lesion stability and no malignancy.



**Fig 5.** A 48-year-old woman diagnosed with ductal carcinoma in situ. Her mammogram revealed suspicious microcalcifications in the right breast's LIQ. All readers except the most experienced missed the lesion in the unaided reading session. On the other hand, all lesions were detected when using the AI system for assistance. This case is one of the true-positive cases included in the dataset. The AI system's viewer displays the outline area and abnormality score.

**TABLE 4.** Average reading time across six readers compared to each reading session.

Reading session	Average reading time per case (seconds)			
	Overall	Positive Cases	Negative Cases	P-value
1	32.92	27.98	37.22	<0.01
2	20.99	19.07	22.66	

for negative cases (95% CI: 19.94, 25.39). These reductions were statistically significant ( $P$ -value  $< 0.01$ ).

The two least experienced readers showed significant time reductions for both positive and negative cases. For Reader 5, times decreased from 23.82 seconds to 15.94 seconds,  $p < 0.01$ , for positive cases, and from 27.56 seconds vs 16.51 seconds,  $p < 0.01$  for negative cases, respectively. For Reader 6, times decreased from 42.23 seconds to 13.37 seconds,  $p < 0.01$ , for positive cases, and from 46.83 seconds vs 11.34 seconds,  $p < 0.01$ , respectively. Fig 6 displays pooled reading time results across readers with and without AI assistance.

## DISCUSSION

In this pilot study, the use of AI assistance was associated with changes in radiologists' diagnostic performance during digital mammography interpretation, including trends toward increased sensitivity, reduced false-negative assessments in most readers, and shorter reading times. However, these effects were not consistent across all readers and did not uniformly reach statistical significance.

In the present study, performance trends associated with AI assistance appeared more noticeable among less experienced readers, although this observation was based on descriptive analyses and was not consistently statistically significant. Similar variability in reader response to decision-support tools has been described in prior observer-performance studies.<sup>36,37</sup> This observation suggests that reader experience may influence how AI assistance is incorporated into interpretation, although this study was not designed to formally assess adoption behavior.

Although AI-aided detection of false-positive cases increased for two readers, specificity did not change significantly because false positives decreased in cases where the AI analysis was negative. Furthermore, our

findings are consistent with prior studies suggesting that AI assistance may influence radiologists' cancer detection behavior, although in this study improvements were not consistently statistically significant.

In contrast to the findings of *Rodriguez-Ruiz et al.*<sup>27</sup>, who found that reading time decreased for low-suspicion cases but increased for high-suspicion cases (without differentiating by reader experience), we observed reduced average reading times for both positive and negative cases, possibly because AI helps radiologists detect lesions more quickly. This reduction in reading time may improve radiologists' clinical efficiency. As expected, the two least experienced readers achieved the shortest average reading time in both positive and negative cases. These findings suggest that AI is unlikely to prolong radiologists' workflow and may in fact shorten reading time in screening contexts. However, in real-world practice, additional factors such as stress, tiredness, and other factors, which were not controlled for in our study, may have an impact on reading time.

The main limitations of this study, similar to those reported by *Rodriguez-Ruiz et al* and *Pacilè et al*,<sup>27,32</sup> stem from the use of a dataset that was not representative of routine screening practice. First, it was enriched with cancer cases, which may have introduced a laboratory effect and contributed to a higher rate of false-positive assessments.<sup>38,39</sup> Future research should ideally evaluate the value of AI support systems in real-world screening settings. Second, the prevalence of cancer and biopsy-confirmed benign cases was higher, which is not consistent with breast cancer screening. However, in this study, our focus was on how radiologists' decision-making changed after receiving AI support. Third, the reading task in this study was more demanding than routine clinical mammography interpretation, as readers were not provided with prior mammograms, additional imaging, or relevant clinical information. Lastly, this study was

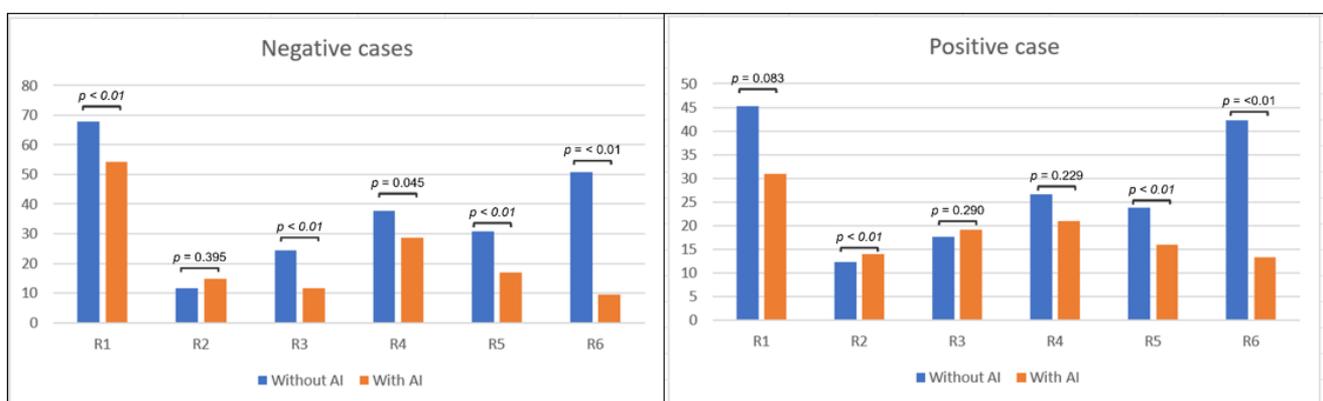


Fig 6. Reading time pooled across readers with and without AI assistance.

Abbreviations: AI = Artificial intelligence, R = Reader

designed as a pilot study at the individual reader level. The study may be underpowered for definitive statistical comparisons.

## CONCLUSION

In this pilot study across individual readers, AI assistance was associated with trends toward improved diagnostic performance and reduced reading time; however, statistically significant improvement was observed in only a subset of readers. Performance gains appeared more pronounced among less-experienced radiologists; however, given methodological limitations and limited statistical power, these findings should be considered preliminary. Larger, adequately powered MRMC studies are required to determine the true clinical impact of AI-assisted mammography interpretation.

## Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available due to patient confidentiality and institutional data protection policies but are available from the corresponding author on reasonable request and with permission from the Siriraj Institutional Review Board.

## ACKNOWLEDGEMENT

The authors would like to thank the staff of the Department of Radiology, Faculty of Medicine Siriraj Hospital, for institutional support during this study.

## DECLARATIONS

### Grants and Funding Information

This research received no external funding.

### Conflict of Interest

The authors declare no conflicts of interest.

### Registration Number of Clinical Trial

Not applicable.

### Author Contributions

Conceptualization and methodology, R.P., V.S., and P.K.; Investigation, R.P., V.S., S.T., and K.M.; Formal analysis, R.P.; Visualization and writing – original draft, R.P.; Writing – review and editing, R.P., V.S., S.T., K.M., and P.K.; Supervision, P.K. All authors have read and agreed to the final version of the manuscript.

### Use of Artificial Intelligence

AI tools were used for grammar and language editing only. All content was reviewed and approved by the authors.

## REFERENCES

1. Khatcheressian JL, Hurley P, Bantug E, Esserman LJ, Grunfeld E, Halberg F, et al. Breast cancer follow-up and management after primary treatment: American Society of Clinical Oncology clinical practice guideline update. *J Clin Oncol.* 2013;31(7): 961-5.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2): 87-108.
3. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L, et al. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2009;151(10): 727-37, W237-742.
4. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer.* 2013;108(11):2205-40.
5. Tabar L, Yen AM, Wu WY, Chen SL, Chiu SY, Fann JC, et al. Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. *Breast J.* 2015;21(1): 13-20.
6. Radiology ACo, D'Orsi CJ. ACR BI-RADS atlas: breast imaging reporting and data system; mammography, ultrasound, magnetic resonance imaging, follow-up and outcome monitoring, data dictionary: ACR, American College of Radiology; 2013.
7. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. *Radiology.* 2017;283(1):49-58.
8. Giger ML, Karssemeijer N, Schnabel JA. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *Annu Rev Biomed Eng.* 2013;15:327-57.
9. Katalinic A, Bartel C, Raspe H, Schreer I. Beyond mammography screening: quality assurance in breast cancer diagnosis (The QuaMaDi Project). *Br J Cancer.* 2007;96(1):157-61.
10. Destounis SV, DiNitto P, Logan-Young W, Bonaccio E, Zuley ML, Willison KM. Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience. *Radiology.* 2004;232(2):578-84.
11. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology.* 1992;184(3): 613-7.
12. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. *Radiographics.* 2003;23(4):881-95.
13. Weber RJ, van Bommel RM, Louwman MW, Nederend J, Voogd AC, Jansen FH, et al. Characteristics and prognosis of interval cancers after biennial screen-film or full-field digital screening mammography. *Breast Cancer Res Treat.* 2016;158(3):471-83.
14. Broeders M, Onland-Moret N, Rijken H, Hendriks J, Verbeek A, Holland R. Use of previous screening mammograms to identify features indicating cases that would have a possible gain in prognosis following earlier detection. *Eur J Cancer.* 2003; 39(12):1770-5.
15. Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. *Acad Radiol.* 1996;3(11):891-897.
16. Rimmer A. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ.* 2017;359.

17. Fenton JJ, Abraham L, Taplin SH, Geller BM, Carney PA, D'Orsi C, et al. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst.* 2011;103(15):1152-61.
18. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med.* 2007;356(14):1399-409.
19. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med.* 2015;175(11):1828-37.
20. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology.* 2001;220(3):781-6.
21. Azavedo E, Zackrisson S, Mejère I, Arnlind MH. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. *BMC Med Imaging.* 2012;12(1):1-12.
22. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciampi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.
23. Kim H-E, Kim HH, Han B-K, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health.* 2020;2(3):e138-e48.
24. Fisches Z, Ball M, Mukama T, Štih V, Payne N, Hickman S, et al. Strategies for integrating artificial intelligence into mammography screening programmes: a retrospective simulation analysis. *Lancet Digit Health.* 2024;6:e803.
25. Hashim HT, Alhatemi AQM, Daraghma M, Ali HT, Khan MA, Sulaiman FA, et al. Artificial intelligence versus radiologists in detecting early-stage breast cancer from mammograms: a meta-analysis of paradigm shifts. *Pol J Radiol.* 2025;90:e1-e8.
26. Eisemann N, Bunk S, Mukama T, Baltus H, Elsner SA, Gomille T, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. *Nat Med.* 2025;31(3):917-24.
27. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology.* 2019;290(2):305-14.
28. Kooi T, Litjens G, Van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal.* 2017;35:303-12.
29. Trister AD, Buist DS, Lee CI. Will machine learning tip the balance in breast cancer screening? *JAMA Oncol.* 2017;3(11):1463-4.
30. Goh S, Goh RSJ, Chong B, Ng QX, Koh GCH, Ngiam KY, et al. Challenges in Implementing Artificial Intelligence in Breast Cancer Screening Programs: Systematic Review and Framework for Safe Adoption. *J Med Internet Res.* 2025;27:e62941.
31. Rodríguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *J Natl Cancer Inst.* 2019;111(9):916-22.
32. Pacilè S, Lopez J, Chone P, Bertinotti T, Grouin JM, Fillard P. Improving breast cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool. *Radiol Artif Intell.* 2020;2(6):e190208.
33. Limchareon S, Kongpromsuk S, Klinwichit P, On-uean A. Radiologist's Role in Artificial Intelligence Era. *Siriraj Med J.* 2022;74(12):891-4.
34. Miyawaki I, Banerjee I, Batalini F, Campello Jorge CA, Celi L, Cobanaj M, et al. Global Disparities in Artificial Intelligence-Based Mammogram Interpretation for Breast Cancer: A Scientometric Analysis of Representation, Trends, and Equity. *Eur J Cancer.* 2025;220:115394.
35. Sung J, Park S, Lee SM, Bae W, Park B, Jung E, et al. Added Value of Deep Learning-based Detection System for Multiple Major Findings on Chest Radiographs: A Randomized Crossover Study. *Radiology.* 2021;299(2):450-9.
36. Hupse R, Samulski M, Lobbes MB, Mann RM, Mus R, den Heeten GJ, et al. Computer-aided detection of masses at mammography: interactive decision support versus prompts. *Radiology.* 2013;266(1):123-9.
37. Tucker L, Gilbert FJ, Astley SM, Dibden A, Seth A, Morel J, et al. Does Reader Performance with Digital Breast Tomosynthesis Vary according to Experience with Two-dimensional Mammography? *Radiology.* 2017;283(2):371-80.
38. Evans KK, Birdwell RL, Wolfe JM. If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One.* 2013;8(5):e64366.
39. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The «laboratory» effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology.* 2008;249(1):47-53.

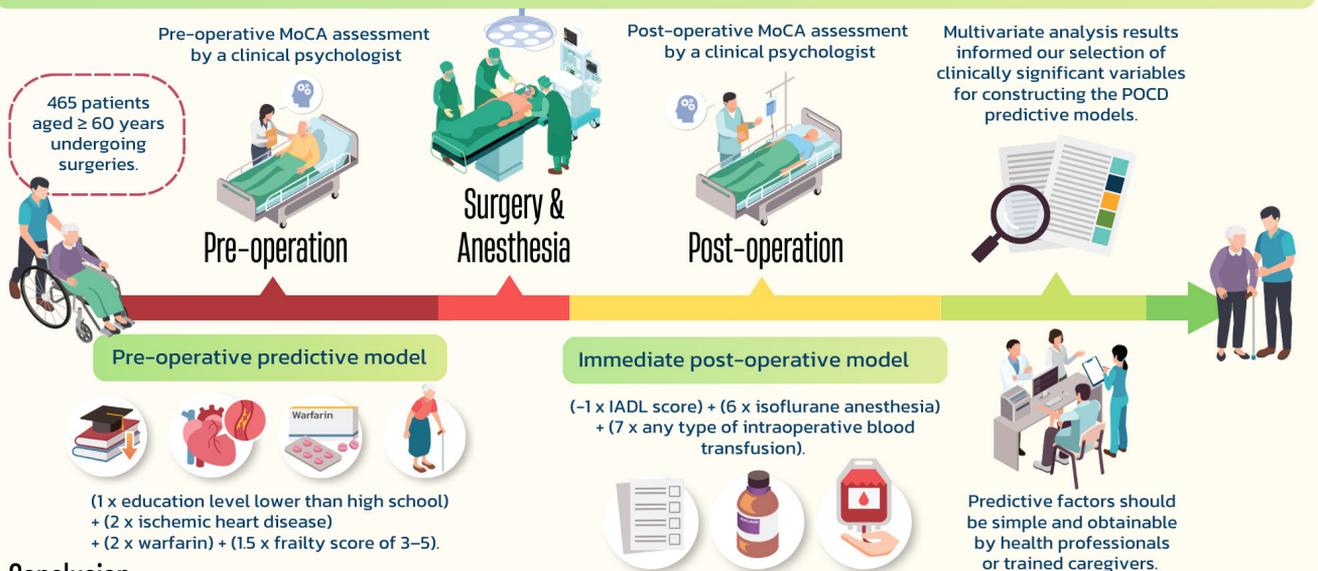
# Predictive Models for Screening of Postoperative Cognitive Dysfunction in Older Surgical Patients

Arunotai Siriussawakul, M.D., Ph.D.<sup>1,2,\*</sup>, Patumporn Suraarunsumrit, M.D., Ph.D.<sup>3</sup>, Varalak Srinonprasert, M.D., MM<sup>2,3,4</sup>, Pawit Somnuek, M.D., Ph.D.<sup>1</sup>, Panop Limratana, M.D.<sup>1</sup>, Unchana Sura-amonrattana, M.D.<sup>3</sup>, Ekkaphop Morkphrom, M.D.<sup>3</sup>, Busadee Pratumvinit, M.D.<sup>5</sup>, Surapa Tornsatitkul, M.D.<sup>1</sup>, Chalita Jiraphorncharas, B.Sc.<sup>2</sup>

<sup>1</sup>Department of Anesthesiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand, <sup>2</sup>Integrated Perioperative Geriatric Excellent Research Center, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand, <sup>3</sup>Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand, <sup>4</sup>Siriraj Health Policy Unit, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand, <sup>5</sup>Department of Clinical Pathology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.

## Predictive Models for Screening of Postoperative Cognitive Dysfunction in Older Surgical Patients

Simple preoperative and early postoperative predictive models reliably identify older surgical patients at risk for early postoperative cognitive dysfunction, facilitating early surveillance and risk stratification.



The preoperative and immediate postoperative POCD predictive models perform reliably. These models may effectively guide early POCD detection and risk modification in older surgical patients.

SCAN FOR FULL TEXT



**ABSTRACT**

**Objective:** Postoperative cognitive dysfunction (POCD) substantially impacts the long-term quality of life of patients and caregivers. Early detection of POCD is essential. We devised quick vigilance screening models for application preoperatively (model one) and during the postoperative period (model two) to predict the development of early POCD (one week after surgery).

**Materials and Methods:** We conducted a cohort study on patients aged  $\geq 60$  years undergoing cardiac or noncardiac surgeries. POCD was defined as a postoperative Montreal Cognitive Assessment decrease of  $\geq 2$  points from the baseline preoperative score. We stipulated that predictive factors should be simple and obtainable by health professionals or trained caregivers. Multivariate analysis results informed our selection of clinically significant variables for constructing the POCD predictive models.

**Results:** Of the 465 patients in the final analysis, the early POCD incidence was 24.9%. The equation used for predictive model one was (1 x education level lower than high school) + (2 x ischemic heart disease) + (2 x warfarin) + (1.5 x frailty score of 3–5). The equation for model two was (-1 x IADL score) + (6 x isoflurane anesthesia) + (7 x any type of intraoperative blood transfusion). Both models displayed well-calibrated curves. The optimal cut-off values of model one and model two to discriminate between a high and low probability of POCD were 2 and 0, respectively.

**Conclusions:** The preoperative and immediate postoperative POCD predictive models perform reliably. These models may effectively guide early POCD detection and risk modification in older surgical patients.

**Keywords:** Cognitive impairment; frailty; functional impairment; postoperative neurocognitive disorders; predictive factors (Siriraj Med J 2026;78(3):196-206)

**INTRODUCTION**

Postoperative cognitive dysfunction (POCD) is characterized by diminished memory orientation, perception, and intellectual performance, that develops after surgery, and can persist for several months.<sup>1</sup> Due to its effect on multiple domains of cognition, POCD diagnosis needs neuropsychological tests, such as neuropsychological (NP) test batteries, Montreal Cognitive Assessment (MoCA), and Mini-Mental State Examination (MMSE), that can be used to assess global cognitive function.<sup>2-7</sup> Although most studies use the NP test, MoCA is easier in daily practice. Moreover, MoCA is better at detecting early cognitive impairment, especially executive dysfunction, than MMSE.<sup>8,9</sup> Despite various diagnostic criteria, POCD is defined as a decline of one or more standard deviations in postoperative cognitive scores, relative to preoperative scores.<sup>2-5</sup> It is commonly found in older surgical patients. One week postoperatively, the incidence of POCD varies with the type of surgery: it has been reported to be 26% – 30% after noncardiac procedures and as high as 71% after cardiac surgeries.<sup>10-13</sup> Even though the incidence drops to 12%–17% at three months and 3% at 12 months, the incidence escalates to over 40% after five years.<sup>12,14</sup> POCD at one week after surgery can be defined as early POCD, and it is not only more prevalent but is also associated with increased 3-month mortality and healthcare utilization.<sup>2,5</sup> However,

currently, there is no specific treatment for POCD. The optimal strategy remains preoperative identification and perioperative risk assessment to facilitate POCD prevention and management.<sup>1</sup>

The pathogenesis of POCD development remains unclear, and several contributing factors have been proposed.<sup>15,16</sup> The functional stability of the central nervous system depends on adequate oxygen and blood supply and sufficient internal environment homeostasis. Mechanisms that lead to hypoxia or alter the homeostatic metabolic state of the brain may lead to POCD. Various etiologies, including preexisting cognitive impairment, anesthetic agents, and metabolic derangements, have been proposed as potential contributors following surgeries. These perioperative variables associated with POCD can be classified into four groups.<sup>9,17-23</sup> Firstly, patient-related factors include being aged over 60, a low education level, cardiovascular comorbidities, alcohol consumption, and restricted physical activity. Secondly, surgical factors are significant, including major surgeries and prolonged operative duration ( $\geq 2$  hours). Thirdly, anesthetic factors include the choice of anesthesia, depth of anesthesia, the drugs and inhalational agents used. Lastly, intraoperative events, such as hypotension, blood transfusion, and desaturation, correlate with POCD occurrence.

Previously proposed models used some variables in patient factor and intraoperative events, which could

predict POCD occurrence; however, those models were only be suitable for the postoperative period.<sup>7,8</sup> The models cannot be applied in the preoperative period. Additionally, MMSE was used to define POCD, which might affect its prevalence. As a result, it could be attributed to the validity of the models. Thus, we need prompt and accurate predictive model assessment to guide a care team's perioperative strategy; thereby, reducing POCD risk and facilitating the monitoring of potential cases. Our study aimed to consolidate multiple POCD contributors into a user-friendly predictive model for non-geriatricians and trained caregivers. Under our comprehensive approach, patients at risk are evaluated twice. The first assessment is conducted preoperatively and draws upon predisposing factors and self-reported instrument data. The second assessment, undertaken after surgery, focuses on baseline function and perioperative precipitating factors, such as anesthetic agents, and intraoperative complications.

## MATERIALS AND METHODS

### *Study design and participants*

This prospective cohort study was conducted at a large university hospital and approved by the Siriraj Institutional Review Board, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand (approval number Si 699/2020). The study adhered to the ethical standards of the 1964 Declaration of Helsinki and subsequent amendments. We recruited Thai-speaking patients aged over 60 who were admitted and scheduled for intermediate to major cardiac or noncardiac surgeries under general anesthesia, with neurosurgical procedures explicitly excluded, from December 2017 to December 2022. All participants provided written informed consent. Excluded were patients unable to communicate in Thai and those with significant visual or auditory impairments, major psychotic disorders, or preoperative delirium.

### *Data collection*

To ensure the accuracy of the study data, research assistants, nurses, psychologists, anesthesiologists, and geriatricians were thoroughly trained in the data required and the collection methods. They obtained information during face-to-face interviews held the day before surgery and again before hospital discharge (approximately one week postoperatively). The data collected from patients included general information (age, sex, education level, alcohol consumption, comorbidities, and medications), preoperative data (American Society of Anesthesiologists physical status classification, laboratory results, and surgery types and sites), and intraoperative variables

(anesthesia type, anesthetic drugs, and adverse events such as hypotension, bleeding, and any type of blood transfusion)

In line with recommendations for geriatric perioperative evaluations, the assessments were conducted alongside traditional practices to meet an acceptable minimum standard. These evaluations encompassed cognition, psychological status, delirium, functional status, frailty, and quality of life (Appendix 1). Cognitive assessments using the Thai version of the MoCA tool were conducted one day before and between 5 and 9 days after surgery. A drop of two or more points from the preoperative score indicated POCD. A previous validation study in the general Thai population reported that the standard deviation (SD) of the MoCA was 2.14.<sup>24</sup>

### *Statistical methods*

To estimate the sample size, we followed statistical guidelines for executing multiple logistic regression analyses. These guidelines suggest that the number of older subjects with POCD should be five to ten times the number of risk factors in a logistic model. Our study had 14 risk factors: age > 60, low education level, alcohol consumption, cerebrovascular disease, diabetes mellitus, renal insufficiency, polypharmacy ( $\geq 5$  medications), cardiac surgery, hypotension, blood transfusion, anesthetic drugs, restricted physical activity, frailty, and psychological derangement.<sup>18-23,25</sup> Accordingly, 70–140 subjects with POCD were needed. Given previous reports showing a 25% POCD incidence, a sample of approximately 450 was deemed sufficient for model development.<sup>10-12</sup>

Statistical analyses were performed using IBM SPSS Statistics version 29 (IBM Corp, Armonk, NY, USA), MedCalc Statistical Software (version 17.6; MedCalc Software Ltd, Ostend, Belgium), and Stata Statistical Software (release 14; StataCorp LLC, College Station, TX, USA). Baseline demographic data were summarized by data type: continuous data with means  $\pm$  SD and categorical data as percentages. Model development utilized data exclusively from the derivation cohort (Appendix 2). The variables known to be related to POCD in older surgical patients were considered. POCD presence or absence was compared using the chi-square test or the independent samples t-test. Our multiple logistic regression model incorporated factors with clinical significance or a p-value < 0.1 from univariate analysis.<sup>26</sup>

The regression coefficients from the multivariable model were used to develop predictive models at two specific time points. First, a predictive model of POCD was developed for the preoperative period. The variables used for the model were predisposing factors and self-

reported instruments. Second, a predictive model of POCD was created for the immediate postoperative period. This model incorporated variables derived from baseline instrumental activities of daily living (IADL) and precipitating factors, such as anesthetic agents, and intraoperative complications.

The calibration of the model, or its fit to the data, was subsequently assessed using the Hosmer–Lemeshow test. The fit was determined by the degree of agreement between the risk score probabilities predicted by the model and the observed probabilities. The model's prognostic ability to discriminate patients with or without risk of POCD was estimated using a receiver operating characteristic (ROC) curve. The estimated shrinkage factor was tested for the performance of the POCD models. The ROC curve that presented the best cut-off point for POCD incorporated Youden's index, sensitivity, specificity, positive predictive value, negative predictive value, AUROC, and 95% confidence interval (CI).<sup>27</sup>

## RESULTS

### *Basic characteristics of patients*

The study enrolled 600 patients undergoing intermediate to major surgery between December 2017 and February 2023. The 465 patients included in the final analysis (Appendix 2) had an age range of 60–89 years (mean 71.31 ± 6.38), and 56.1% (261) were males. Approximately half of the patients (53.8%, 250) underwent cardiovascular surgery, 21.9% (102) had intra-abdominal surgery, and 18.5% (86) received orthopedic surgery. The incidence of early POCD was 24.9% (116/465).

### *Comparison of general and perioperative data between the POCD and non-POCD groups*

Based on cognitive function, patients were divided into a POCD group (n = 116) and a non-POCD group (n = 349). Table 1 presents the baseline characteristics and results from self-reported instruments. Compared to the non-POCD group, patients in the POCD group had a lower education level and were more likely to have a previous history of cerebrovascular and cardiovascular diseases (p < 0.05). Regarding medications, patients with POCD experienced a higher frequency of polypharmacy (p = 0.020). The most common drugs administered to patients with POCD were antihypertensive drugs, warfarin, and diuretics. In terms of self-report instruments related to functional status, patients with POCD had higher frailty scores (p = 0.026) and lower IADL scores (p = 0.019).

### *Development and validation of the POCD prediction*

### *model for the preoperative period*

A multivariate logistic regression analysis was performed using a backward stepwise procedure to obtain an optimized POCD prediction model for the preoperative period (Appendix 3). The analysis incorporated factors reported in the literature<sup>18-23,28</sup> that may influence the development of POCD, along with variables in our univariate analysis with p < 0.1. The best predictive model for POCD in the preoperative period was formulated as follows (Appendix 4):

(1 x education level lower than high school) + (2 x ischemic heart disease) + (2 x warfarin) + (1.5 x frailty score of 3–5)

The AUROC curve value (95% CI) was 0.66 (0.60–0.72). Table 2 shows the optimal cut-off value for distinguishing between a high and low probability of POCD was set at two. This cut-off yielded the highest value for Youden's index (0.25) (Fig 1A), the best AUROC curve value, and optimal values for sensitivity (67.83%, 95% CI 58.47%–76.23%) and specificity (57.27%, 95% CI 51.85%–62.56%). Subsequently, we utilized the bootstrap method, involving repeated sampling 400 times, to determine the internal validation of the preoperative predictive model of POCD (Appendix 5-7). The AUROC curve value for this validation was 0.67 (95% CI 0.61–0.73). Additionally, the calibration curve demonstrated that the results of the predictive model were closely aligned with the observed results. The predictive model also exhibited strong goodness of fit (Hosmer–Lemeshow test,  $\chi^2 = 2.93$ , p = 0.71). Given the moderate discriminatory performance of the model, as reflected by an AUROC, this predictive model is intended to be used primarily as a screening tool for early risk stratification rather than as a diagnostic instrument for POCD.

### *Development and validation of the POCD prediction model for the immediate postoperative period*

A multivariate logistic regression analysis using a backward stepwise procedure was performed to obtain an optimized POCD prediction model for the immediate postoperative period (Appendix 8). We considered variables reported in the literature that may precipitate the development of POCD during the perioperative period along with those with p < 0.1 in our univariate analysis (Table 3). The best predictive model for POCD in the postoperative period was formulated as follows (Appendix 9):

(-1 x IADL score) + (6 x isoflurane) + (7 x any type of intraoperative blood transfusion)

The AUROC curve value (95% CI) was 0.68

**TABLE 1.** Demographic data.

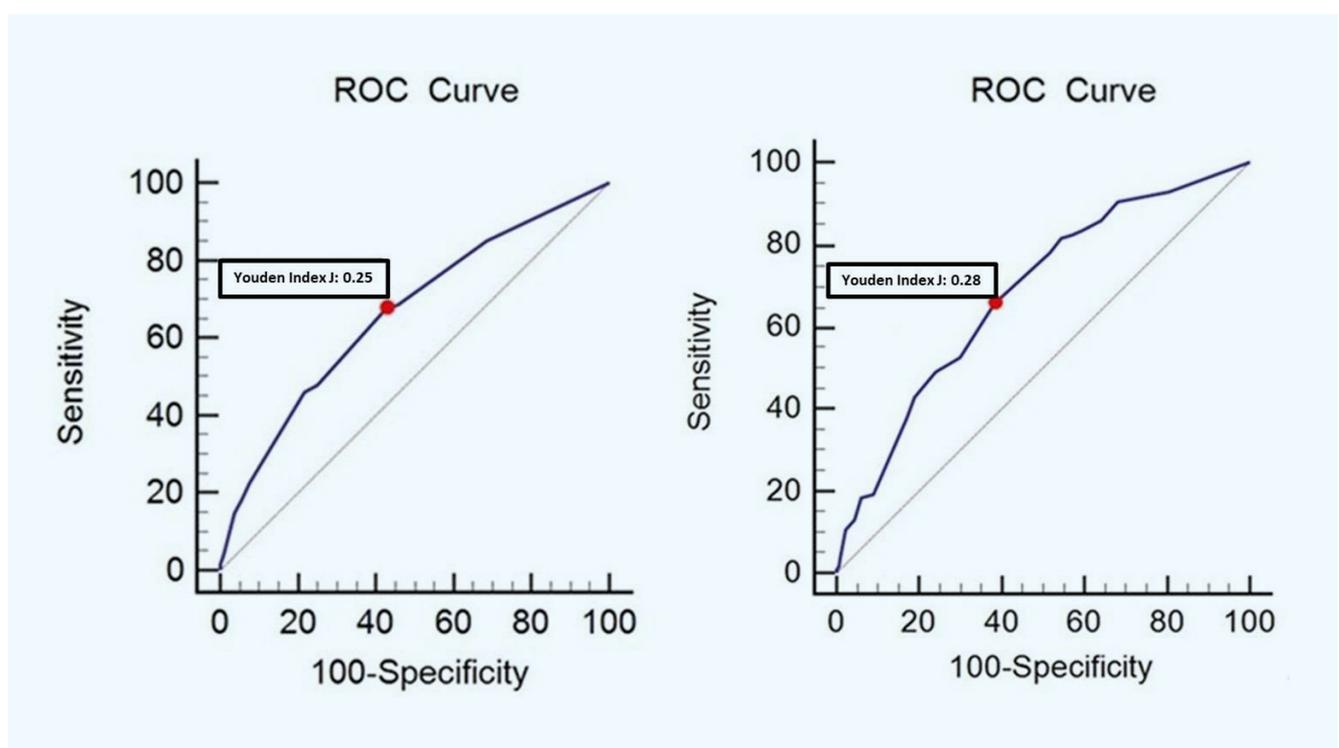
Variables	No POCD (n=349)	POCD (n=116)	p
Sex; male	194 (55.6%)	67 (57.8%)	0.683
Age (years)	71.6±6.3	70.4±6.6	0.071
BMI (kg/m <sup>2</sup> )	24.8±4.2	24.7±4.2	0.828
Education level			
Lower than high school	152 (43.6%)	63 (54.3%)	0.045*
Further/higher education	197 (56.4%)	53 (45.7%)	
Comorbidity			
Hypertension	280 (80.9%)	95 (83.3%)	0.566
Diabetes mellitus	115 (33%)	39 (33.6%)	0.849
Dementia	3 (0.9%)	2 (1.7%)	0.439
Cerebrovascular accident	34 (9.7%)	14 (12.1%)	0.458
CKD stages 3b, 4, and 5	45 (12.9%)	24 (20.7%)	0.036*
Cirrhosis	6 (1.7%)	4 (3.4%)	0.275
Ischemic heart disease/myocardial infarction	123 (35.2%)	68 (58.6%)	<0.001*
Atrial fibrillation	34 (9.7%)	20 (17.2%)	0.032*
Valvular heart disease	69 (19.8%)	36 (31%)	0.013*
Congestive heart failure	35 (10%)	21 (18.1%)	0.023*
Dyslipidemia	238 (68.2%)	85 (73.3%)	0.318
Current smoker	14 (4.0%)	3 (2.6%)	0.479
Alcohol consumption history	332 (95.1%)	114 (98.3%)	0.182
Medications			
Warfarin	24 (6.9%)	17 (14.7%)	0.013*
Antiarrhythmic drug	144 (41.3%)	68 (58.6%)	0.001*
Antihypertensive drug	271 (77.7%)	102 (87.9%)	0.020*
Diuretics	77 (22.1%)	39 (33.6%)	0.014*
Preoperative benzodiazepine use	52 (14.9%)	22 (19%)	0.312
Polypharmacy (current medication ≥ 5)	231 (66.2%)	91 (78.4%)	0.020*
Tools			
Modified IQCODE score	3.13±0.29	3.11±0.31	0.493
9Q ≥ 7 (n %)	19 (5.4%)	11 (9.5%)	0.136
Barthel ADL score	95.12±8.52	94.96±9.96	0.869
IADL score	6.37±1.84	5.88±2.09	0.019*
Frailty; score 3–5 (n %)	38 (10.9%)	22 (19.0%)	0.026*
VAS score	73.88±15.92	70.46±15.97	0.048*
EQ-5D-5L	0.84±0.18	0.85±0.17	0.604

**Abbreviations:** 9Q: Nine-Questions Depression-Rating Scale; Barthel ADL: Barthel activities of daily living; BMI: body mass index; CKD: chronic kidney disease; EQ-5D-5L: European Quality of Life 5 Dimensions 5 Level Version; IADL: instrumental activities of daily living; kg: kilogram; m<sup>2</sup>: square meter; Modified IQCODE: modified informant questionnaire on cognitive decline in the elderly; POCD: postoperative cognitive dysfunction; VAS: visual analogue scale

**TABLE 2.** Clinical scores with associated probabilities of a POCD positive result at the preoperative stage.

Clinical scores	Probabilities	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)	Youden's index
≥ 1.0	18.70	85.22 (77.39 - 91.15)	31.40 (26.52 - 36.59)	29.34 (27.22 - 31.55)	86.40 (79.95 - 91.01)	44.88 (40.27 - 49.56)	0.17
≥ 1.5	22.20	68.07 (59.38 - 77.02)	54.07 (48.64 - 59.43)	33.33 (29.70 - 37.18)	83.78 (79.49 - 87.32)	57.73 (53.07 - 62.30)	0.22
≥ 2.0	26.10	67.83 (58.47 - 76.23)	57.27 (51.85 - 62.56)	34.67 (30.80 - 38.74)	84.19 (80.09 - 87.58)	59.91 (55.27 - 64.43)	0.25
≥ 2.5	30.50	47.83 (38.43 - 57.34)	75.00 (7.08 - 79.49)	39.01 (32.93 - 45.45)	81.13 (78.13 - 83.81)	68.19 (63.71 - 72.43)	0.23
≥ 3.0	35.23	46.09 (36.75 - 55.63)	78.20 (73.46 - 82.45)	41.41 (34.79 - 48.35)	81.27 (78.41 - 83.83)	70.15 (65.74 - 74.31)	0.24

**Abbreviations:** CI: confidence interval; NPV: negative predictive value; POCD: postoperative cognitive dysfunction; PPV: positive predictive value

**Fig 1.** The receiver operating characteristic (ROC) curve for the predictive model of POCD

1A. Preoperative period

1B. Immediate postoperative period

**TABLE 3.** Preoperative and intraoperative data.

Variables	No POCD (n=349)	POCD (n=116)	p
ASA classification			
ASA ≤ 2	120 (34.4%)	16 (13.8%)	<0.001*
ASA > 2	229 (65.6%)	100 (86.2%)	
Charlson comorbidity index > 5	156 (44.7%)	49 (42.2%)	0.693
Site of surgery			
Cardiac surgery	168 (48.1%)	82 (70.7%)	<0.001*
Noncardiac surgery	179 (51.3%)	34 (29.3%)	
Preoperative Hb (g/dL) <sup>a</sup>			
Normal	186 (53.3%)	68 (58.6%)	0.319
Abnormal	163 (46.7%)	48 (41.4%)	
BIS monitoring	63 (18%)	15 (12.9%)	0.211
NIRS monitoring	45 (12.9%)	22 (19%)	0.114
Received blood product	172 (49.3%)	89 (76.7%)	<0.001*
Complications	222 (63.6%)	70 (60.3%)	0.483
Medications			
Midazolam	101 (28.9%)	55 (47.4%)	<0.001*
Dexmedetomidine	45 (12.9%)	22 (19.0%)	0.114
Isoflurane	49 (14%)	36 (31 %)	<0.001*
Morphine	145 (41.5%)	51 (44.0%)	0.681
Fentanyl	328 (94%)	114 (98.3%)	0.110
Paracetamol	23 (6.6%)	7 (6.0%)	0.822
Nefopam	10 (2.9%)	1 (0.9%)	0.244
Vasopressor use	260 (74.4%)	103 (88.8%)	0.003*

Complications: hypertension, hypotension, and severe arrhythmia. Hb(g/dL)<sup>a</sup> normal Range: 13–17 g/dL for male, 12–15 g/dL for female

**Abbreviations:** ASA: American Society of Anesthesiologists physical status; BIS: bispectral index; dL: deciliter; g: gram; Hb: hemoglobin; NIRS: near-infrared spectroscopy; POCD: postoperative cognitive dysfunction

(0.62–0.74). Table 4 shows an optimal cut-off value of 0 was chosen to discriminate between a high and low probability of POCD. This cut-off offered the highest value for Youden's index (0.28) (Fig 1B), the best AUROC curve value, and optimal values for sensitivity (65.79%, 95% CI 56.32%–74.42%) and specificity (61.76%, 95% CI 56.37%–66.95%). We subsequently performed the bootstrap method, repeating the sampling 400 times, to determine the internal validation of the immediate postoperative predictive model for POCD (Appendix 10-12). This process yielded an AUROC curve value of 0.68 (95% CI

0.63–0.74). In addition, the calibration curve showed good agreement between the predicted results of the model and the observed results. Moreover, the predictive model exhibited satisfactory performance regarding goodness of fit (Hosmer–Lemeshow test,  $\chi^2 = 7.49$ ,  $p = 0.38$ ). Given the moderate discriminatory performance demonstrated by the AUROC, this immediate postoperative predictive model is intended to function as a screening tool for early risk identification and postoperative surveillance rather than as a diagnostic model for POCD.

**TABLE 4.** Clinical scores with associated probabilities of a POCD positive result at the immediate postoperative stage.

Clinical scores	Probabilities	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)	Youden's index
≥ -3	19.23	82.46 (74.21-88.94)	42.94 (37.61-48.39)	32.64 (29.95-35.45)	87.95 (82.80-91.72)	52.86 (48.16-57.53)	0.25
≥ -2	21.26	81.58 (73.23-88.22)	45.59 (40.21-51.05)	33.45 (30.61-36.42)	88.07 (83.14-91.70)	54.63 (49.92-59.27)	0.27
≥ -1	23.45	78.07 (69.35-85.28)	48.24 (42.81-53.69)	33.58 (30.51-36.81)	86.77 (82.02-90.42)	55.73 (51.02-60.36)	0.26
≥ 0	25.79	65.79 (56.32-74.42)	61.76 (56.37-66.95)	36.59 (32.32-41.0)	84.34 (80.46-87.56)	62.78 (58.15-67.24)	0.28
≥ 1	28.27	52.63 (43.06-62.06)	70.00 (64.82-74.83)	37.04 (31.68-42.74)	81.51 (78.21-84.41)	65.64 (61.07-70.00)	0.23
≥ 2	30.89	49.12 (39.64-58.65)	75.88 (70.97-80.33)	40.58 (34.37-47.10)	81.65 (78.62-84.32)	69.16 (64.69-73.38)	0.25

**Abbreviations:** CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value

## DISCUSSION

Our research focused on developing POCD prediction models for application before and after surgery. The preoperative and immediate postoperative POCD predictive models could be applied before and immediately after surgery, respectively. The two models perform reliably. The significant demographic features (low education level and poor functional status), anesthetic agents (isoflurane), and intraoperative events (such as any type of blood transfusion) incorporated into these models are consistent with reported risk factors for POCD.<sup>9,18-20,23,29</sup>

In the preoperative predictive model, frailty and warfarin were included. However, some controversies exist regarding whether warfarin might reduce POCD incidence, or if frailty correlates with POCD.<sup>30-32</sup> Proposed evidence supported the argument that warfarin being taken by patients with atrial fibrillation (AF) can prevent strokes and silent cerebral infarcts without clinical strokes that can cause cognitive impairment.<sup>30</sup> Contrastingly, our findings indicate that warfarin consumption could elevate the risk of POCD. This finding could be explained by the synergistic effects between warfarin and cardiovascular co-morbid diseases that might be attributed to POCD. In cases involving presurgical anticoagulation bridging, patients may be at a higher risk of bleeding and intraoperative hypotension, both of which could contribute to POCD

development. Conversely, without bridging therapy, they might be more susceptible to postoperative stroke, leading to changes in cognitive function postoperatively. Accordingly, we concluded that warfarin use as the current medication before surgery is a fitting variable for inclusion in the preoperative predictive model for POCD.

The previous studies proposed similar inflammatory mechanisms between frailty and POCD that lead to immunological alterations, particularly increased C-reactive protein (CRP) and interleukin-6 (IL-6).<sup>33,34</sup> This evidence could suggest that frailty could be associated with POCD, although the previous study did not find a relationship.<sup>31</sup> The difference in time to detect POCD could influence this finding as frailty was not associated with POCD at 3 months, but an association between frailty and POCD at 1 week was reported.<sup>32</sup> Likewise, our study found a significant connection between frailty and the development of early POCD. Including frailty also enhances the predictive value of established risk scores, such as the Society of Thoracic Surgeons Predicted Risk of Mortality or Major Morbidity.<sup>35</sup> Hence, it is reasonable to consider preoperative frailty as part of the POCD predictive model since frail surgical patients may be at risk of developing POCD.

Regarding the postoperative model prediction, it

includes poor functional status, intraoperative isoflurane, and receiving of intraoperative blood transfusion. These factors align with previous evidence of POCD triggers.<sup>9,19,20,23</sup> Shiraboina et al. identified the intraoperative administration of more than two units of transfused blood (odds ratio 4.32, 95% CI 1.59–11.70,  $p = 0.004$ ) was a single independent predictor for the development of early POCD after cardiac surgery.<sup>6</sup> Another recent study introduced the use of a nomogram model for POCD prediction. It incorporated preoperative features (age  $\geq 70$ , body mass index  $< 18.5$  kg/m<sup>2</sup>, presence of cerebrovascular disease, white blood cell count  $> 10 \times 10^9/L$ , hemoglobin level  $< 120$  g/L) and intraoperative events (intraoperative blood loss  $> 400$  ml, operative time  $> 8$  hours). The model showed promising discrimination with an acceptable AUROC value of 0.710 (95% CI 0.645–0.775).<sup>7</sup> The calibration value was also satisfactory, with the fit confirmed according to the Hosmer–Lemeshow test ( $\chi^2$  5.133,  $p = 0.274$ ).<sup>7</sup>

However, the previous models were limited to specific surgical populations and included intraoperative variables irrelevant to the preoperative period.<sup>7,8</sup> In contrast, our two models cater to geriatric patients in preoperative and immediate postoperative periods. This attribute broadens their applicability in predicting POCD risk across cardiac and noncardiac surgical patients. Our postoperative model prediction also added intraoperative isoflurane as the potential anesthetic agent with an increasing probability of detecting early POCD.

Our study has several strengths. Firstly, the study identified variables best suited for the POCD predictive model. They were either consistent with or different from factors used in earlier models.<sup>6,7</sup> Despite the variance in components, our predictive model demonstrated comparable AUROC and fit tests to prior studies.<sup>7</sup> Secondly, our approach capitalized on the strength of prospective cohort studies by meticulously collecting all pertinent parameters and confounders during the perioperative period. Lastly, our multidisciplinary team of skilled healthcare personnel also evaluated POCD with the MoCA tool. This instrument has a higher sensitivity than the MMSE screening tool and has proven to be an optimal and effective standalone resource for cognitive screening.<sup>27</sup>

This study has some limitations. While we employed a larger sample size than previous studies, only older patients were included. Given that their mean age was approximately 71 years, the study findings may be less applicable to younger patients. The number of patients in our study was also insufficient to classify POCD risk by surgery type. Furthermore, we created a POCD predictive formula by combining various pre- and intraoperative

parameters from other studies to craft the most suitable model. The external validity of this approach requires further verification. Despite these challenges, the model successfully demonstrated satisfactory discrimination and fitness in our tests. Future research could explore the application of this predictive model to various surgical procedures and potentially across different age groups.

## CONCLUSIONS

The preoperative and immediate postoperative POCD predictive models could be beneficial for implementation. Early identification of patients at risk for POCD may enable targeted counseling and preventive measures, enhancing patient care and outcomes.

## Data Availability Statement

The principal investigator (A.S.) can provide the data that support the findings of this study upon reasonable request.

## ACKNOWLEDGMENTS

The Integrated Perioperative Geriatric Excellent Research Center facilitated this study. Thanks to Prof. Panita Limpawattana, Faculty of Medicine, Khon Kaen University, for assistance with manuscript writing. Thanks to Mr. Monai Sauejui and Mrs. Nipaporn Sangarunakul for curating the data, and thank Mrs. Rinrada Preedachitkun for her statistical assistance. Finally, the authors gratefully acknowledge the professional editing of this paper by Mr. David Park.

## DECLARATIONS

### Grants and Funding Information

The National Research Council of Thailand supported this study. The funder was not involved in the study design, data collection, analysis, publishing decisions, or preparation of the manuscript.

### Conflict of Interest

The authors assert that the research was undertaken without any commercial or financial relationships that might be construed as potential conflicts of interest.

### Registration Number of Clinical Trial

This trial was registered in the Thai Clinical Trials Registry on the 15th of January, 2019 (registration number: TCTR20190115001).

### Authors' Contributions

Conceptualization, A.S., P.S., and V.S. ; Methodology, A.S. and S.T. ; Validation, P.S. ; Formal analysis, A.S.

; Managed resources, V.S. ; Evaluated the data, C.J. ; Writing – original draft, A.S. and P.So. ; Writing – review and editing, A.S., P.So., P.S., P.L., U.S., E.M., B.P., S.T. and V.S. ; Funding acquisition, A.S.; Supervision, A.S. All authors have read and agreed to the final version of the manuscript.

### Use of Artificial Intelligence

The authors declare that no artificial intelligence tools were used in this study or in the preparation of the manuscript.

### Ethics Approval and Consent to Participate

The Institutional Review Board of Siriraj Hospital, Faculty of Medicine Siriraj Hospital, Mahidol University, approved this study (approval number Si 699/2020). It was conducted in accordance with the ethical guidelines set forth in the 1964 Declaration of Helsinki and its subsequent amendments. All participants provided written informed consent.

### REFERENCES

- Kotekar N, Shenkar A, Nagaraj R. Postoperative cognitive dysfunction - current preventive strategies. *Clin Interv Aging*. 2018;13:2267-73.
- Monk TG, Weldon BC, Garvan CW, Dede DE, van der Aa MT, Heilman KM, Gravenstein JS. Predictors of cognitive dysfunction after major noncardiac surgery. *Anesthesiology*. 2008;108(1):18-30.
- Steinmetz J, Christensen KB, Lund T, Lohse N, Rasmussen LS. Long-term consequences of postoperative cognitive dysfunction. *Anesthesiology*. 2009;110(3):548-55.
- Deiner S, Liu X, Lin HM, Jacoby R, Kim J, Baxter MG, et al. Does Postoperative Cognitive Decline Result in New Disability After Surgery? *Ann Surg*. 2021;274(6):e1108-e14.
- Suraarunsumrit P, Pathonsmith C, Srinonprasert V, Sangarunakul N, Jiraphorncharas C, Siriussawakul A. Postoperative cognitive dysfunction in older surgical patients associated with increased healthcare utilization: a prospective study from an upper-middle-income country. *BMC Geriatr*. 2022;22(1):213.
- Shiraboina M, Ayya S, Srikanth Y, Kumar R, Durga P, Gopinath R. Predictors of postoperative cognitive dysfunction in adult patients undergoing elective cardiac surgery. *Indian J Anaesth*. 2014;58(3):334-6.
- Huang H, Chou J, Tang Y, Ouyang W, Wu X, Le Y. Nomogram to predict postoperative cognitive dysfunction in elderly patients undergoing gastrointestinal tumor resection. *Front Aging Neurosci*. 2022;14:1037852.
- Liu J, Huang K, Zhu B, Zhou B, Ahmad Harb AK, Liu L, Wu X. Neuropsychological Tests in Post-operative Cognitive Dysfunction: Methods and Applications. *Front Psychol*. 2021;12:684307.
- Needham MJ, Webb CE, Bryden DC. Postoperative cognitive dysfunction and dementia: what we need to know and do. *Br J Anaesth*. 2017;119(Suppl 1):i115-i25.
- Moller JT, Cluitmans P, Rasmussen LS, Houx P, Rasmussen H, Canet J, et al. Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. International Study of Post-Operative Cognitive Dysfunction. *Lancet*. 1998;351(9106):857-61.
- Huang C, Mårtensson J, Gögenur I, Asghar MS. Exploring Postoperative Cognitive Dysfunction and Delirium in Noncardiac Surgery Using MRI: A Systematic Review. *Neural Plast*. 2018;2018:1281657.
- Paredes S, Cortínez L, Contreras V, Silbert B. Post-operative cognitive dysfunction at 3 months in adults after non-cardiac surgery: a qualitative systematic review. *Acta Anaesthesiol Scand*. 2016;60(8):1043-58.
- Relander K, Hietanen M, Rantanen K, Rämö J, Vento A, Saastamoinen KP, et al. Postoperative cognitive change after cardiac surgery predicts long-term cognitive outcome. *Brain Behav*. 2020;10(9):e01750.
- Newman MF, Kirchner JL, Phillips-Bute B, Gaver V, Grocott H, Jones RH, et al. Longitudinal assessment of neurocognitive function after coronary-artery bypass surgery. *N Engl J Med*. 2001;344(6):395-402.
- Green CM, Schaffer SD. Postoperative cognitive dysfunction in noncardiac surgery: A review. *Trends in Anaesthesia and Critical Care*. 2019;24:40-8.
- Berger M, Terrando N, Smith SK, Browndyke JN, Newman MF, Mathew JP. Neurocognitive Function after Cardiac Surgery: From Phenotypes to Mechanisms. *Anesthesiology*. 2018;129(4):829-51.
- Cheng H, Clymer JW, Po-Han Chen B, Sadeghirad B, Ferko NC, Cameron CG, Hinoul P. Prolonged operative duration is associated with complications: a systematic review and meta-analysis. *J Surg Res*. 2018;229:134-44.
- Hua M, Min J. Postoperative Cognitive Dysfunction and the Protective Effects of Enriched Environment: A Systematic Review. *Neurodegener Dis*. 2020;20(4):113-22.
- Suenghataiphorn T, Songwisit S, Tornsatitkul S, Somnuke P. An Overview on Postoperative Cognitive Dysfunction; Pathophysiology, Risk Factors, Prevention and Treatment. *Siriraj Med J*. 2022;74(10):705-13.
- Zhu S-H, Ji M-H, Gao D-P, Li W-Y, Yang J-J. Association between perioperative blood transfusion and early postoperative cognitive dysfunction in aged patients following total hip replacement surgery. *Uppsala Journal of Medical Sciences*. 2014;119(3):262-7.
- Plaschke K, Hauth S, Jansen C, Bruckner T, Schramm C, Karck M, Kopitz J. The influence of preoperative serum anticholinergic activity and other risk factors for the development of postoperative cognitive dysfunction after cardiac surgery. *J Thorac Cardiovasc Surg*. 2013;145(3):805-11.
- Rossi A, Burkhart C, Dell-Kuster S, Pollock BG, Strebel SP, Monsch AU, et al. Serum anticholinergic activity and postoperative cognitive dysfunction in elderly patients. *Anesth Analg*. 2014;119(4):947-55.
- Yang X, Huang X, Li M, Jiang Y, Zhang H. Identification of individuals at risk for postoperative cognitive dysfunction (POCD). *Ther Adv Neurol Disord*. 2022;15:17562864221114356.
- Tangwongchai S, Charernboon T, Phanasathit M, Akkayagorn L, Hemrungron S, Phanthumchinda K, et al. The validity of Thai version of the montreal cognitive assessment (MoCA-T). 2009. 172 p. Available from: [https://www.researchgate.net/publication/311715441\\_The\\_validity\\_of\\_thai\\_version\\_of\\_the\\_montreal\\_cognitive\\_assessment\\_MoCA-T](https://www.researchgate.net/publication/311715441_The_validity_of_thai_version_of_the_montreal_cognitive_assessment_MoCA-T)

25. Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. *BMC Geriatrics*. 2017;17(1):230.
26. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. *Source Code Biol Med*. 2008;3:17.
27. Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE. Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimers Dement*. 2013;9(5):529-37.
28. Litaker D, Locala J, Franco K, Bronson DL, Tannous Z. Preoperative risk factors for postoperative delirium. *Gen Hosp Psychiatry*. 2001;23(2):84-9.
29. Bowden T, Hurt CS, Sanders J, Aitken LM. Predictors of cognitive dysfunction after cardiac surgery: a systematic review. *Eur J Cardiovasc Nurs*. 2022;21(3):192-204.
30. Prystowsky EN, Padanilam BJ. Preserve the brain: primary goal in the therapy of atrial fibrillation. *J Am Coll Cardiol*. 2013;62(6):540-2.
31. Mahanna-Gabrielli E, Zhang K, Sieber FE, Lin HM, Liu X, Sewell M, et al. Frailty Is Associated With Postoperative Delirium But Not With Postoperative Cognitive Decline in Older Noncardiac Surgery Patients. *Anesth Analg*. 2020;130(6):1516-23.
32. Nomura Y, Nakano M, Bush B, Tian J, Yamaguchi A, Walston J, et al. Observational Study Examining the Association of Baseline Frailty and Postcardiac Surgery Delirium and Cognitive Change. *Anesth Analg*. 2019;129(2):507-14.
33. Androsova G, Krause R, Winterer G, Schneider R. Biomarkers of postoperative delirium and cognitive dysfunction. *Front Aging Neurosci*. 2015;7:112.
34. Yao X, Li H, Leng SX. Inflammation and immune system alterations in frailty. *Clin Geriatr Med*. 2011;27(1):79-87.
35. Afilalo J, Mottillo S, Eisenberg MJ, Alexander KP, Noiseux N, Perrault LP, et al. Addition of frailty and disability to cardiac surgery risk scores identifies elderly patients at high risk of mortality or major morbidity. *Circ Cardiovasc Qual Outcomes*. 2012;5(2):222-8.

# Effects of Forward Leaning Characteristics on Protective Steps when Performing Voluntary-induced Stepping Response in Young and Older Adults: A Cross-Sectional Study

Pornprom Chayasit, Ph.D.\*<sup>ip</sup>, Rumpa Boonsinsukh, Ph.D.<sup>ip</sup>

Department of Physical Therapy, Faculty of Physical Therapy, Srinakharinwirot University, Nakhon Nayok 26120, Thailand.



\*Corresponding author: Pornprom Chayasit

E-mail: pornprom@g.swu.ac.th

Received 4 October 2025 Revised 31 January 2026 Accepted 31 January 2026

ORCID ID: <http://orcid.org/0000-0002-9250-9589>

<https://doi.org/10.33192/smj.v78i3.278066>



## ABSTRACT

**Objective:** To examine the effect of forward leaning distance and aging on protective step length, movement strategies, and postural stability during voluntary-induced stepping response (VSR).

**Materials and Methods:** Thirty healthy young adults ( $19.5 \pm 0.7$  years) and ten healthy older adults ( $68.9 \pm 4.4$  years) participated in a cross-sectional study. Young adults performed VSR under two conditions: short-distance ( $D_s$ ) and long-distance ( $D_l$ ), based on pelvis displacement ratios. Older adults performed VSR at their preferred leaning distance ( $D_p$ ) and were instructed to recover balance with a single step. Protective step length, movement initiation and cessation strategies, and postural stability were assessed using 2D video analysis. Statistical comparisons were conducted using paired t-tests and chi-square tests ( $\alpha = 0.05$ ).

**Results:** Step length was significantly greater in  $D_l$  than  $D_s$  ( $p < 0.001$ ). Older adults in  $D_p$  showed no significant difference in step length compared to young adults in  $D_l$ . However, older adults more frequently used trunk bending (28% versus 5% of trials), rigid body strategies (71% versus 30%), grasping (13% versus 0%), body sway, and multi-step responses (27% versus 3%) ( $p < 0.001$ ).

**Conclusion:** Forward-leaning distance influences step length during VSR. Aging is linked to altered movement strategies and reduced postural control. These findings suggest that VSR may serve as a targeted intervention to enhance dynamic stability in older adults, supporting broader goals of health promotion and overall well-being.

**Keywords:** Postural balance; accidental falls; gait; health; motor activity (Siriraj Med J 2026;78(3):207-217)

## INTRODUCTION

Falls represent a global public health concern, accounting for an estimated 684,000 deaths annually worldwide, and ranking second only to road traffic injuries among causes of unintentional injury-related mortality.<sup>1,2</sup> Beyond their immediate physical consequences, falls often lead to psychological distress, reduced confidence, social withdrawal, and functional decline, particularly in later life.<sup>2,3</sup> The burden rises sharply with age: fall-related mortality among older adults has shown an increasing trend over recent decades, with rates exceeding 50 deaths per 100,000 population aged 60 years and older, whereas younger adults typically exhibit much lower rates, below 10 deaths per 100,000. This striking contrast highlights the disproportionate vulnerability of older populations.<sup>2,4</sup> These data underscore the need for preventive strategies tailored to aging populations.<sup>5,6</sup> Task-specific interventions targeting balance recovery, such as protective stepping response training, have shown promise in mitigating fall risk among older adults.<sup>7</sup>

Voluntary-induced stepping response (VSR) training has emerged as a low-cost, accessible method for enhancing protective stepping ability. A single 50-minute session has been shown to improve stepping performance without the need for specialized equipment. Participants with stroke demonstrated an average increase of approximately 10% in affected step length and greater affected-leg initiation (from 20.2% to 27% of trials). The participants also showed reductions in multiple steps (from 28.6%

to 12.3%) and grasping (from 30.4% to 21.5%).<sup>8</sup> VSR involves two sequential actions: initiating an internal perturbation via forward leaning using the ankles as the axis of rotation, followed by executing a protective step. Although protective step length may appear similar across age groups, older adults demonstrate distinct movement patterns, including a greater reliance on trunk bending strategies (28% versus 10% of trials in young adults), potentially reflecting reduced postural control or lower tolerance for perturbation.<sup>9</sup>

The effectiveness of protective stepping depends not only on training methods such as VSR but also on how aging and perturbation magnitude shape the body's ability to recover balance. While higher perturbation intensities increase body momentum and the likelihood of stepping, older adults tend to initiate steps earlier under lower perturbation thresholds compared to younger individuals.<sup>10</sup> These findings suggest age-related declines in the capacity to counteract destabilizing forces. However, it remains unclear how voluntary forward leaning distance modulates perturbation magnitude during VSR and affects protective step execution and stability. Clarifying this relationship is important because VSR training relies on self-induced perturbations to strengthen balance recovery strategy. Understanding the dose-response effect of leaning distance is therefore crucial, as it positions VSR not merely as a training exercise but as a practical, scalable approach for integrating self-induced perturbations into fall prevention programs.

Therefore, this study aimed to determine whether increasing forward leaning distance during VSR leads to larger perturbations and longer protective steps, and to assess whether older adults demonstrate different movement strategies compared to young adults under identical conditions. We hypothesized that greater leaning distances would result in larger perturbation magnitudes and longer protective steps, and that aging would be associated with increased reliance on trunk bending strategies and greater instability during stepping.

## MATERIALS AND METHODS

### Study design

This study was cross-sectional in design and was approved by The Institutional Review Board (PTPT2021-006).

### Sample size

The sample size for young adults was determined using G\*Power 3.1, with an alpha level of 0.05 and a power of 80%. The total sample size was calculated based on an estimated effect size of  $d = 0.47$ , derived from normalized step length data obtained from a pilot study of five young adults across different leaning conditions. Consequently, 30 young adults were conveniently sampled for the primary objective. These participants were recruited from university students and alumni, representing healthy young adults who were not athletes.

Additionally, data from 10 healthy older adults were obtained from a previously published study for the secondary objectives.<sup>9</sup> Older adults were eligible if they were at least 60 years of age, able to stand and walk independently without assistive devices for at least 6 meters, and had no cognitive deficits as assessed by the Mini-Mental State Examination ( $MMSE \geq 24$ ). Exclusion criteria included prior experience with perturbation testing or training within the past year, visual problems, or any neurological, cardiovascular, or musculoskeletal conditions that could interfere with task performance. Recruitment of older adults in the original study was conducted through local community centers and word-of-mouth invitations among acquaintances of faculty members and hospital staff, ensuring participants were community-dwelling.<sup>9</sup>

### Participants

Healthy young adults aged 18 to 26 years were recruited for this study. Participants were excluded if they had visual, neurological, or musculoskeletal conditions that could impact protective stepping ability. All participants provided written informed consent prior to participation.

### Procedures

Data on age, gender, weight, height, and body mass index (BMI) were collected from all participants prior to the VSR assessment. For older adults, falling history, fear of falling, and experience with perturbation training or testing were recorded using a simple binary question.

Additionally, older adults underwent functional and cognitive assessments to confirm that they were cognitively intact and physically independent. These included the Mini-Mental State Examination (MMSE),<sup>11</sup> the Activities-specific Balance Confidence (ABC) scale,<sup>12</sup> the Five-Times Sit-to-Stand Test (FTSST),<sup>13,14</sup> the Timed Up and Go (TUG) test, the Dynamic Gait Index (DGI),<sup>15</sup> and items 16-18 from The Balance Evaluation Systems Test (BESTest).<sup>16</sup> These clinical measures were employed to characterize health status, balance confidence, and functional mobility in older adults; accordingly, they were not administered to young adults.

### Voluntary-induced stepping response assessment

To assess VSR, participants were instructed to lean their entire body forward using the ankles as the axis of rotation and to take a step upon sensing postural instability. For the primary objective, two controlled leaning conditions were examined: short-distance ( $D_s$ ) and long-distance ( $D_l$ ). The leaning distance was calculated in centimeters to facilitate setting a rope target for each leaning condition. The rope was positioned in front of the pelvis to define the leaning threshold. Young adult participants were instructed to lean forward until their pelvis contacted the rope and then took a step.

Leaning distances were pre-determined using the formula:  $D = \text{Anterior Superior Iliac Spine (ASIS) displacement ratio} \times \text{participant height}$ . Ratios were derived from previous studies,<sup>9</sup> with  $D_l$  based on young adult data (0.13) and  $D_s$  designed to reflect the typical maximum leaning distance observed in older adults. The ASIS displacement ratios were calculated as follows: ASIS displacement ratio of  $D_l = \text{Average ASIS displacement of young adults} \div \text{body height}$ ; and ASIS displacement ratio of  $D_s = \text{Average ASIS displacement of older adults} \div \text{body height}$ . Specifically,  $D_s$  was defined using a ratio of 0.10, corresponding to the average pelvic displacement ( $D_p$ ) recorded in older adults during maximal voluntary leaning. This approach ensured that the short-distance condition represented a leaning distance similar to that naturally adopted by older adults, thereby facilitating meaningful comparisons across conditions.

To control for variability, leaning velocity was fixed at 0.39 m/s in both conditions. Participants synchronized

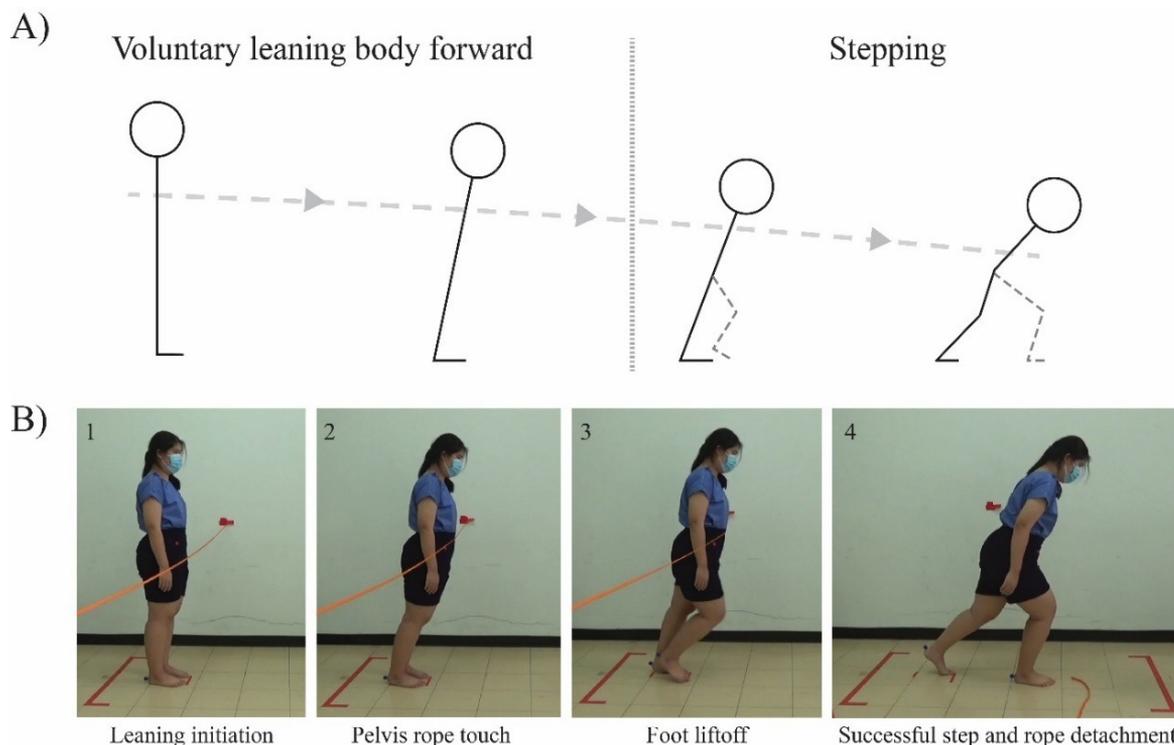
their movement with two consecutive beep sounds: the first signaling initiation and the second marking the endpoint, corresponding to the moment the pelvis contacted the rope. The interval between the first and second beep was calculated by dividing each participant's predetermined leaning distance (m) by the fixed velocity of 0.39 m/s. Rest intervals between test conditions were approximately 2–3 minutes, or longer if requested, to minimize fatigue. Each condition was assessed 10 times. The interval between observations was approximately 2–3 minutes, allowing participants to return to their original position and the researcher to recheck.

Participants completed at least 10 practice trials or practiced until they were familiar with the procedure. Practice and testing were conducted one condition at a time to reinforce the leaning technique. Rest periods were provided between conditions or as needed. To minimize the effects of fatigue and motor learning, the sequence of conditions was randomly assigned, and all trials were completed within a single day.

Participants stood barefoot in their preferred foot position, ensuring both feet were aligned at the same level in the anteroposterior direction. Foot placement was marked and consistently monitored throughout practice and testing. For safety, one end of the rope

was secured to a floor-fixed pole, while the other was lightly taped to the wall beside the participant. This setup allowed the rope to detach easily upon contact with the pelvis, preventing tripping or other adverse incidents, as shown in Fig 1B.

Older adult participants performed the VSR by leaning forward as far as possible and attempting to take only a single step to regain stability, maintaining the final position for 5 seconds ( $D_p$  condition). To ensure consistency, participants performed approximately 10 practice trials before the actual assessment. During practice, the researcher guided the ASIS movement to help participants identify their maximum safe leaning distance that still permitted execution of a single protective step. If leaning was insufficient, participants were encouraged to lean further; if excessive leaning resulted in multiple steps, they were asked to reduce the distance slightly. Thus, the "preferred distance" reflected each participant's perceived maximum safe limit of stability. No rope was used in this group. Participants stood barefoot with a naturally comfortable foot width, maintaining the same stance across all trials. For safety, older participants wore safety harness, and a research assistant remained beside them throughout the session.



**Fig 1.** Presents the setup for evaluating a young adult participant. Panel (A) illustrates the VSR action using stick-figure drawings. Panel (B) showcases the assessment setting and an example of a VSR testing trial. Participants performed VSR in the following sequence: initiating the lean (B-1), continuing until the pelvis contacted the rope (B-2), followed by foot liftoff (B-3), and taking a step (B-4). The rope was designed for easy detachment, as shown in Fig B-4.

## Measurement

Kinematic data were collected using a two-dimensional (1920×1080 pixels) video camera, capturing footage at 50 frames per second. The camera was positioned laterally to the participant, orthogonal to the movement plane, at a distance of 4 meters away from the recording area. The frame was calibrated using a perspective grid.

Four 3D foam markers were attached to participants' heels and ASIS before testing to define spatial positions. Heel positions were identified and exported using the KINOVEA software for protective step length calculation. The KINOVEA program has been validated as a reliable tool for measuring marker distances up to 5 meters.<sup>17</sup> During the pilot phase of this study, the researchers conducted interrater reliability for marker placement and identification in KINOVEA and found it to be excellent (ICC3,1 = 0.99, 95%CI 0.98-0.99).

## Data analysis

A trial was deemed successful if the participant's pelvis contacted the rope precisely at the onset of the auditory cue, and foot liftoff occurred subsequently, following pelvic contact at the level of the ASIS marker.

Protective step length was defined as the anteroposterior distance between the heel markers of the stance and stepping limbs at foot touchdown. To account for inter-individual variability, step length was normalized to participant height.

Initiation strategies for voluntary stepping (trunk leaning vs. trunk bending), movement cessation strategies (flexible vs. rigid body), and compensatory responses (e.g., grasping, body sway, multiple steps, or foot adjustments) were evaluated via visual inspection of video recordings. All assessments were performed by a single researcher, and intra-observer reliability was examined prior to data collection. Repeated assessments were separated by a 1-day interval to minimize recall bias while maintaining participant stability. Prior to the main study, pilot testing was conducted with 10 participants, each performing 10 trials per condition (yielding 100 paired observations). These pilot data were analyzed using Kappa statistics and indicated substantial agreement ( $\kappa = 0.81$ ). The pilot confirmed the feasibility of the protocol and the consistency of the measurement procedure; however, these data were not included in the main statistical analysis.

Forward leaning strategy was classified based on the temporal onset of hip and trunk motion. Simultaneous movement of both segments indicated a trunk leaning strategy, whereas earlier trunk motion relative to the hips indicated a trunk bending strategy.

Movement cessation strategy was determined from the first protective step touchdown until postural stability was regained. Trials exhibiting flexion of the neck, trunk, and hips during motion offset, followed by upright re-stabilization, were classified as employing a flexible body strategy. Conversely, trials in which the neck, trunk, hips, and legs remained rigid throughout the post-step phase were categorized as using a rigid body strategy.

## Statistical analysis

Descriptive statistics were used to summarize participant characteristics. Data normality was assessed using the Shapiro–Wilk test. Given the limited sample size in the older adult group, visual inspection of QQ plots and histograms was additionally performed. Between-group comparisons (young vs. older adults) were performed using independent t-tests for normally distributed variables and Mann–Whitney U tests for non-normally distributed variables. Differences in sex distribution were analyzed using the chi-square test.

To address the primary objective, paired t-tests were conducted to compare protective step length between conditions  $D_s$  and  $D_l$  in young adults. For both conditions ( $D_s$  and  $D_l$ ), data from 30 participants were analyzed. Each participant performed 10 test trials per condition, but the mean value of these trials was calculated for each participant before statistical analysis. Thus, 30 paired observations ( $D_s$  versus  $D_l$ ) were included in the paired t-test. For the secondary objective, independent t-tests were used to compare protective step length between young adults in condition  $D_s$  and older adults in condition  $D_p$ . The mean value of 10 trials for each participant was also calculated prior to statistical analysis. Chi-square tests were applied to examine group differences in movement initiation and cessation strategies.

Statistical significance was set at  $p < 0.05$ . All analyses were performed using IBM SPSS Statistics (Version 26.0; IBM Corp., Armonk, NY, USA).

## RESULTS

Data from 30 healthy young adults and 10 healthy older adults were presented. Participants' characteristics are shown in Table 1. One older adult reported a history of falling, but none expressed a fear of falling. There were no significant differences in sex, weight, height, or BMI between young and older adult participants.

### The influence of leaning distance on step length

The results showed that protective step length was significantly longer in condition  $D_l$  than  $D_s$  as shown in Table 2.

### Aging and protective step length

There was no significant difference in normalized protective step length between the  $D_p$  condition in older adults and the  $D_s$  condition in young adults, nor between  $D_p$  in older adults and  $D_l$  in young adults. In young adults, normalized step length was significantly larger for  $D_l$  than for  $D_s$  as shown in Table 2.

### Aging and postural control during VSR

The results revealed that older adults were significantly more likely to employ a trunk-bending strategy during movement initiation, characterized by a forward sequence in which the trunk initiated motion prior to the hips, than young adults ( $p < 0.001$ ), as illustrated in Fig 2A. Regardless of the initial strategy, older adults demonstrated a higher percentage to cease body motion with rigid

body strategy than young adults ( $p < 0.001$ ), as shown in Fig 2B.

In comparable leaning distances under both  $D_s$  and  $D_p$  conditions, no significant difference was observed in the percentage of balance loss between young and older adults ( $p > 0.05$ ). Nevertheless, older adults exhibited more grasping, body sway after the first protective step touchdown, foot adjustments following initial foot touchdown, and multiple stepping, relative to young adults ( $p < 0.001$ ), as illustrated in Fig 3A, B.

### DISCUSSION

The primary finding of this study is that VSR effectively modulates the magnitude of internally generated perturbations, thereby influencing protective step length. Longer leaning distances corresponded to

**TABLE 1.** Participants' characteristics.

	Young adults (n = 30)	Older adults (n = 10)
Age (y)	19.5±0.7	68.9±4.4
Weight (kg)	59.5±12.2	68.6±16.5
Height (m)	1.7±0.1	1.7±0.1
BMI (kg/m <sup>2</sup> )	19.9 (18.5 to 23.8)	24.1 (20.4 to 27.4)
Sex (M, %)	11 (35.5)	4 (40)
Fall history (Yes, %)	-	1 (10)
Fear of falling (Yes, %)	-	0 (0)
Perturbation experience (Y, %)	-	0 (0)
MMSE (/30)	-	29±1.6
ABC (/100)	-	98.0±2.4
FTSST (s)	-	9.5±2.3
TUG (s)	-	8.7±1.3
DGI (/24)	-	23.9±0.3
BESTest item 16-18 (/12)	-	11.6±1.0

**Note:** The data were presented in the mean ± SD. BMI data were not normal distributed, and it was reported in median (interquartile range). Gender, fear of falling, and perturbation experience are presented in frequency (percentage).

**Abbreviations:** BMI = Body Mass Index; MMSE = Mini-Mental State Examination; ABC = Activities-specific Balance Confidence scale; FTSST = Five-times-sit-to-stand Test; TUG = Timed Up and Go; DGI = Dynamic Gait Index; BESTest = Balance Evaluation System Test.

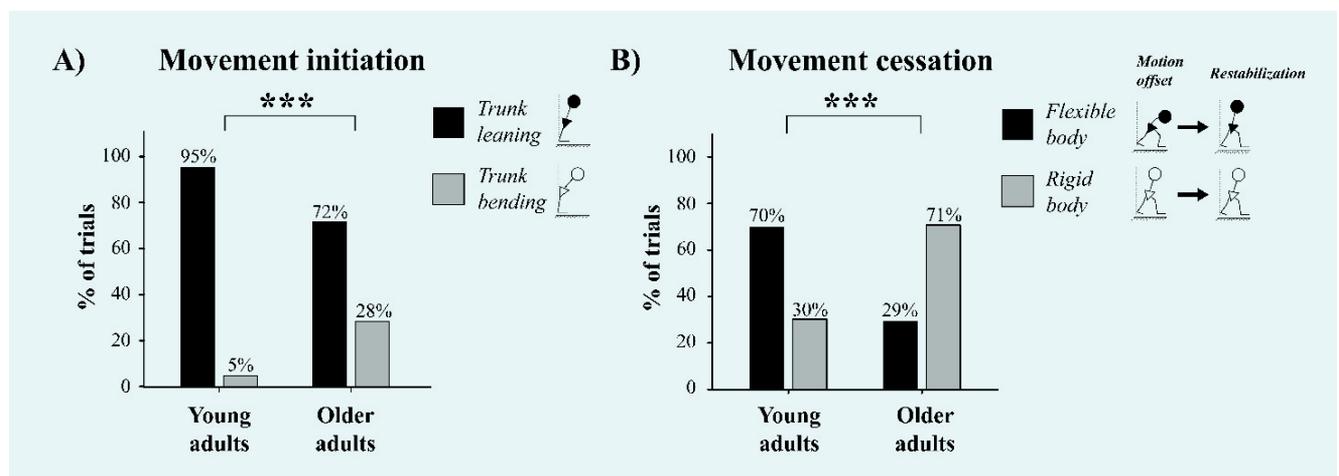
**TABLE 2.** Normalized protective step length of young adults underwent VSR in conditions  $D_s$  and  $D_l$  and older adults who performed VSR in conditions  $D_p$ .

	Young adults		Older adults	p - value
	$D_s$	$D_l$	$D_p$	
Leaning distance (cm)	16.9±0.8	22.4±1.0 <sup>a,b</sup>	17.1±3.7	<sup>a,b</sup> p < 0.001 <sup>a</sup> MD 5.1 ± 0.4 (95% CI: 4.4 to 6.5) <sup>b</sup> MD 5.3 ± 0.6 (95% CI: 3.8 to 6.8)
Normalized step length	0.3 ± 0.1	0.4 ± 0.1 <sup>a</sup>	0.4 ± 0.1	<sup>a</sup> p < 0.001 <sup>a</sup> MD 0.03 ± 0.04 (95% CI: 0.02 to 0.04)

<sup>a</sup> presents significant difference of outcomes between  $D_l$  and  $D_s$ .

<sup>b</sup> presents significant difference of outcomes between  $D_l$  and  $D_p$ .

**Abbreviation and notes:** MD is mean difference. Normalized protective step length = step length / participant height. Leaning distance is presented as mean ± SD in centimeters unit.

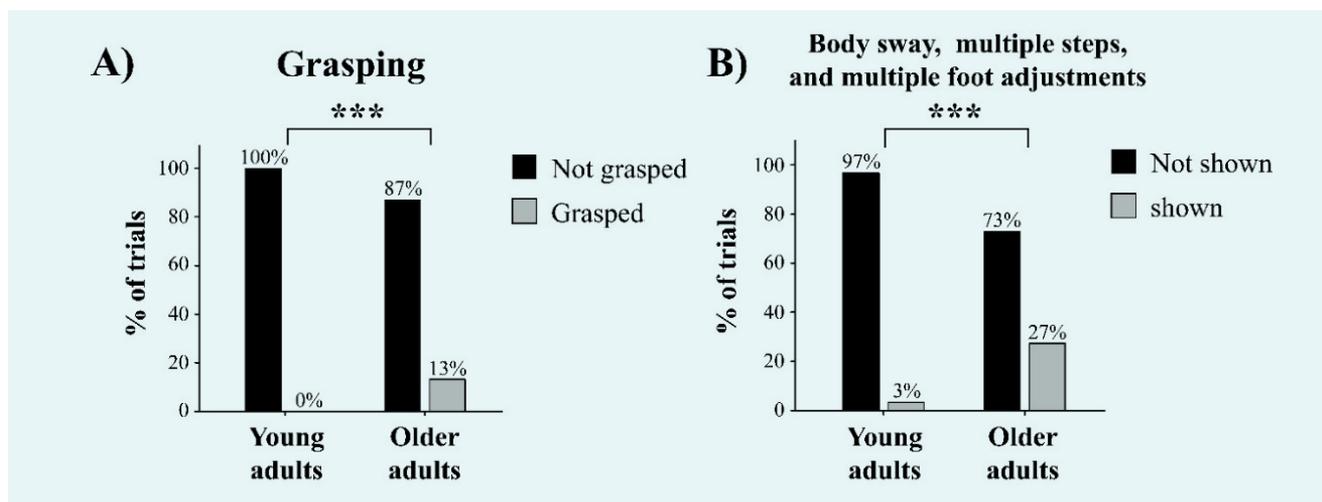


**Fig 2.** The bar charts present movement strategies during VSR in young ( $D_s$ ) and older ( $D_p$ ) adults, segmented into two phases: (A) movement initiation and (B) movement cessation. In phase A, two initiation strategies were identified. The trunk-leaning strategy involved simultaneous movement of the trunk and hips, whereas the trunk-bending strategy was characterized by sequential movement, with the trunk initiating motion ahead of the hips. In phase B, two cessation strategies were observed. The flexible-body strategy featured forward flexion of the neck, trunk, and lower segments at the point of motion offset, followed by gradual regaining of upright posture for restabilization. In contrast, the rigid-body strategy was characterized by persistent stiffness of the neck, trunk, and legs throughout the transition from motion offset to restabilization. \*\*\* p<0.001.

longer compensatory step lengths required to maintain balance. Furthermore, the results revealed age-related differences in perturbation response during VSR. Although older adults attempted to preserve postural stability by employing strategies (such as increased use of a trunk bending strategy during movement onset, using a rigid body strategy during movement cessation, and step

lengthening), there was still a greater incidence of grasping, body sway, multiple foot adjustments, and multiple steps, indicative of postural instability compared to young adults.

The primary finding suggested that VSR replicates key mechanical characteristics of externally induced perturbations, such as those observed in waist-pull



**Fig 3.** The bar charts present percentage of grasping (A) and percentage of body sway, multiple steps, or multiple foot adjustment (B) during VSR in young adults ( $D_s$ ) and older adults ( $D_p$ ).

\*\*\*  $p < 0.001$ .

protocols.<sup>18</sup> By emphasizing anterior displacement of the ASIS in the horizontal plane while maintaining rigid body alignment around the ankle axis, VSR facilitates displacement of the center of mass (CoM) beyond the base of support (BoS). This displacement triggers a protective step to reestablish stability. Such mechanical demands closely resemble real-world destabilizing scenarios (e.g., tripping or slipping), thereby making VSR a functionally relevant task. Among healthy young adults, greater forward lean distances were significantly associated with longer compensatory steps. This association also aligns with biomechanical principles: as the CoM shifts further from the BoS, the resulting increase in destabilizing moments requires a correspondingly faster muscle activation and a longer step to regain balance.<sup>19,20</sup>

While the present study is observational, it offers mechanistic insight into the utility of VSR for perturbation-based training. Prior interventional research has shown that VSR can enhance protective stepping capacity, particularly by promoting change-in-support strategies.<sup>8</sup> Unlike in-place responses, which rely on joint torques and are often insufficient under large perturbations, protective steps offer a dynamic means of restoring stability.<sup>21</sup> VSR generates self-induced instability of sufficient magnitude to trigger protective stepping. The findings also suggest that voluntary leaning distance during VSR influences the intensity of such internal perturbations. This task-specific stimulus supports scalable, progressive protocols to strengthen change-in-support strategies, especially for individuals at high risk of falling.

Secondary findings highlighted age-related differences in protective stepping strategies during VSR. Older adults predominantly employed trunk bending rather

than trunk leaning to initiate movement. This pattern reflected reduced tolerance to perturbation magnitude and an attempt to pre-compensate for instability. By keeping the center of mass (CoM) within the base of support (BoS) and minimizing forward momentum, they sought to maintain stability.<sup>10</sup> In contrast, young adults adopted a flexible body strategy, characterized by continued neck–trunk–hip motion at step touchdown. This strategy facilitated dynamic stabilization by displacing the center of mass (CoM) toward the stepped foot. The cessation strategy promoted a more favorable alignment of the ground reaction force (GRF) vector relative to the center of pressure (CoP), thereby counteracting forward momentum during VSR.<sup>22,23</sup> This favorable GRF–CoP alignment, referring to the spatial relationship between the GRF vector and the CoP, may reduce mechanical demand on the hip extensors.<sup>24</sup> Collectively, these findings underscore that aging is associated with a shift from flexible, momentum-absorbing strategies toward more rigid, cautious responses. This shift increases reliance on hip extensor moments, which are well documented to decline with age, and, together with age-related reduction in quadriceps strength,<sup>10,24</sup> may further limit stabilizing capacity and ultimately compromise postural stability under perturbation.

Protective stepping during VSR may be influenced not only by mechanical demand but also by age-related neural factors. In the present study, older adults in the  $D_p$  condition exhibited protective step lengths comparable to those of young adults in the  $D_l$  condition, despite leaning a shorter distance. This pattern suggests that step lengthening in older adults reflects a compensatory response to reduced postural control. Neurophysiological

evidence supports this interpretation, showing that cortical responses to perturbation, particularly from the supplementary motor area (SMA), can affect postural control during step execution.<sup>25</sup> For example, larger perturbation-evoked N1 potentials have been observed in individuals with poorer balance, indicating heightened cortical engagement during recovery.<sup>25</sup> Moreover, experimental inhibition of SMA and posterior cerebellar activity has been shown to shorten anticipatory postural adjustment (APA) duration and maladapt step execution.<sup>26</sup> These findings align with evidence that aging-related structural changes in the cerebellum and basal ganglia, together with altered motor control in the SMA, may affect the scaling of protective stepping responses.<sup>27,28</sup> Thus, even under equivalent perturbation magnitudes, older adults may exhibit longer step lengths as a compensatory strategy to offset diminished postural control and sensorimotor integration.<sup>29</sup>

### Clinical implications

The findings of this study provide new evidence that VSR can elicit quantifiable age-related differences in protective stepping strategies under controlled conditions. Specifically, older adults were more likely to employ trunk bending and rigid body strategies and showed greater instability indicators (grasping, sway, multiple steps) despite comparable step lengths to young adults. These results highlight that protective step length alone is not a sufficient marker of successful recovery in older adults. Clinically, the use of VSR offers a safe and repeatable method to provoke protective stepping without relying on accidental falls, allowing clinicians to systematically adjust task difficulty by modifying forward leaning distance. Monitoring movement strategies and instability signs during VSR may therefore provide practical markers for tailoring interventions aimed at improving balance recovery in populations at risk of falling. Importantly, the frequent use of multi-step responses observed in older adults carries clinical relevance. Geriatric literature indicates that the inability to recover balance in a single step is a strong predictor of future fall risk.<sup>30</sup> Thus, documenting multi-step responses during VSR may serve as an early indicator of impaired balance recovery capacity and help identify individuals who require targeted fall-prevention interventions.

### Limitations and future directions

It should also be noted that the findings for older adults were based on only 10 participants, which further limits the generalizability of the results. A post-hoc power analysis indicated 38% power at  $\alpha = 0.05$ , underscoring the

limited statistical strength of this subsample. Accordingly, these results are considered pilot findings that provide preliminary insight into age-related differences and should be interpreted with caution. Future research should therefore involve larger and more representative samples of older adults to strengthen statistical power and provide more definitive conclusions.

A key limitation of this study is its exclusive focus on anteroposterior (AP) dynamics during VSR, without direct assessment of mediolateral (ML) stability. Although forward leaning primarily challenges AP balance, ML control also plays an important role in compensatory stepping and whole-body coordination.<sup>18,31</sup> Our use of 2D video analysis limited the ability to capture ML instability or transverse plane rotations as effectively as 3D motion capture, which may be particularly relevant for older adults. Future studies should incorporate ML measures (e.g., frontal-plane kinematics, step width variability, CoM–CoP relationships) to provide a more comprehensive characterization of balance recovery.

Another limitation is that anticipatory postural adjustments (APA) prior to leaning or stepping were not quantified. APA reflect central preparatory mechanisms that optimize dynamic stability and may differ between young and older adults.<sup>32</sup> Without evaluating APA-related parameters (e.g., CoP shifts, EMG activity), it remains unclear whether age-related impairments in VSR arise from deficits in feedforward control or purely reactive mechanisms. Future work should examine APA responses to clarify their contribution to protective stepping.

Finally, anthropometric scaling must be considered when interpreting step length outcomes. We normalized step length to participant height, which is widely adopted, practical, and yields outcomes comparable to leg length normalization.<sup>33–36</sup> Prior studies also highlight the influence of height on postural balance and gait performance.<sup>37,38</sup> Thus, height normalization was deemed appropriate for the present study.

### CONCLUSIONS

This study demonstrates that voluntary leaning modulates perturbation intensity and stepping responses during VSR. However, aging alters postural strategies in ways that may compromise dynamic stability, particularly under challenging conditions.

### Data Availability Statement

The data supporting the findings of this study are not publicly available due to ethical restrictions and participant confidentiality.

## ACKNOWLEDGEMENTS

We would like to thank Chanatip Pureesathit, Natchamon Chawpeng, Natnicha Kerdpaibull, Arissara Sangkaew, Arisara Nooeiad, Natthida Saipet, Premisikan Sathitporn, Ananya Nakaem, Suppaka Chataluek, Supanuch Kerdsiri for their assistance with data collection and participant coordination during the study.

## DECLARATIONS

The authors gratefully acknowledge the Faculty of Physical Therapy, Srinakharinwirot University, for their invaluable support in facilitating data collection and funding for abstract dissemination. Their contribution was instrumental to the successful completion of this study.

## Conflict of Interest

There is no conflict of interest to be declared.

## Registration Number of Clinical Trial

As this study was observational and did not involve any intervention, clinical trial registration was not required.

## Author Contributions

Conceptualization and methodology, P.C. and R.B. ; Investigation, P.C. ; Formal analysis, P.C. and R.B. ; Visualization and writing – original draft, P.C. ; Writing – review and editing, P.C. and R.B. ; Funding acquisition, P.C. ; Supervision, R.B. . All authors have read and agreed to the final version of the manuscript.

## Use of Artificial Intelligence

We would like to declare that generative AI (Copilot) was used solely for assistance in checking and refining the English language in this manuscript. The authors entirely generated the content, ideas, and findings presented in the manuscript without AI assistance. After language editing, the authors reviewed and validated the final version to ensure its accuracy and integrity.

## REFERENCES

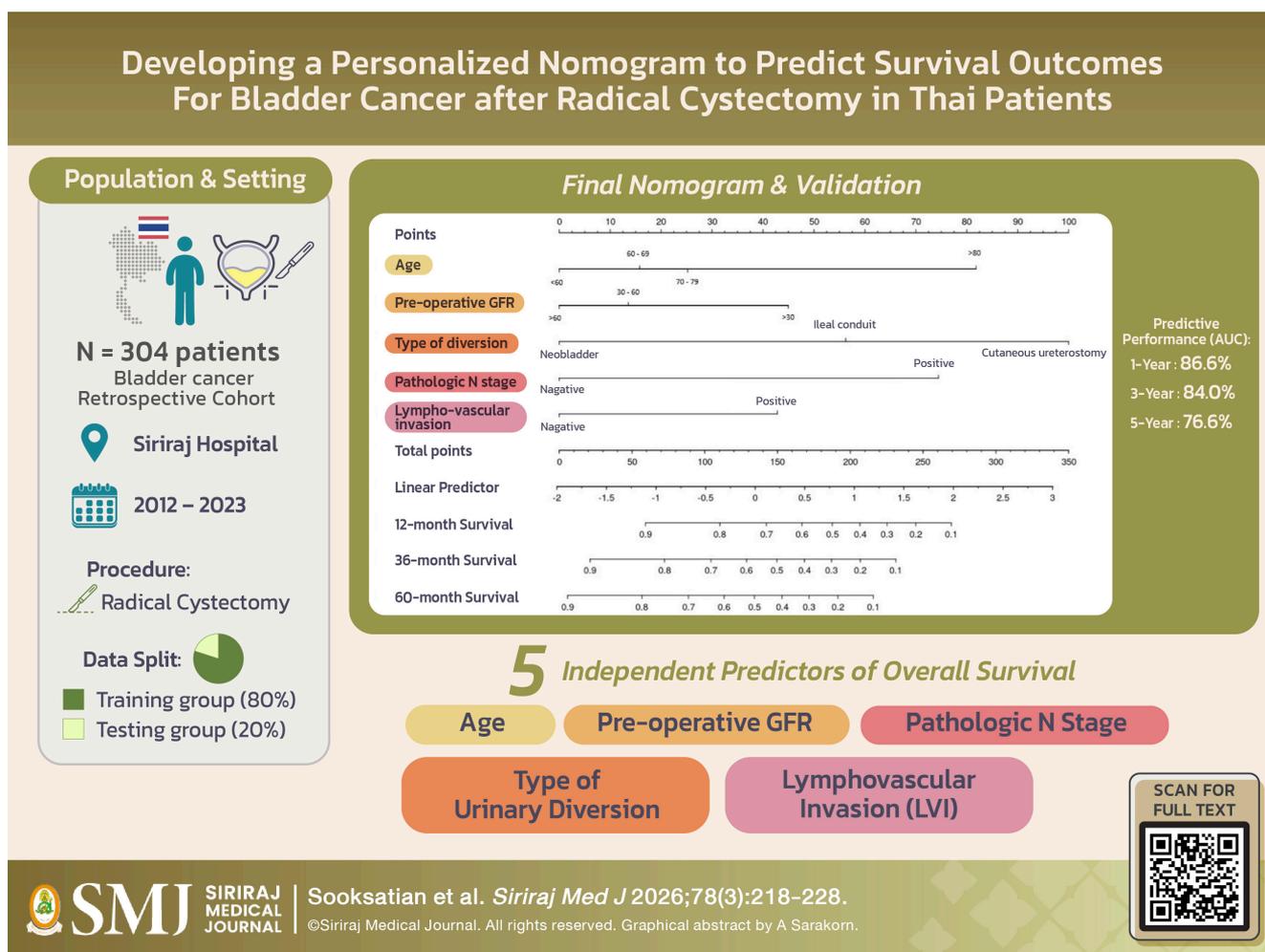
- World Health Organization. Falls [Internet]. 2021. Available from: <https://www.who.int/news-room/fact-sheets/detail/falls/>.
- James SL, Lucchesi LR, Bisignano C, Castle CD, Dingels ZV, Fox JT, et al. The global burden of falls: global, regional and national estimates of morbidity and mortality from the Global Burden of Disease Study 2017. *Inj Prev*. 2020;26(Suppl 1): i3-i11.
- Gambaro E, Gramaglia C, Azzolina D, Campani D, Molin AD, Zeppego P. The complex associations between late life depression, fear of falling and risk of falls. A systematic review and meta-analysis. *Ageing Research Reviews*. 2022;73:101532.
- Monteiro YCM, Vieira MAdS, Vitorino PVdO, Queiroz SJd, Policena GM, Souza ACSE. Trend of fall-related mortality among the elderly. *Rev Esc Enferm USP*. 2021;55:e20200069.
- Montero-Odasso M, van der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, et al. World guidelines for falls prevention and management for older adults: a global initiative. *Age and Ageing*. 2022;51(9):afac205.
- Dajpratham P, Thitisakulchai P, Pongratanakul R, Prapavanond R, Haridravedh S, Muangpaisan W. Effectiveness of Personalized Multifactorial Fall Risk Assessment and Intervention in Reducing Fall Rates Among Older Adults: A Retrospective Study. *Siriraj Medical Journal*. 2025;77(1):64-72.
- Bhatt T, Wang Y, Wang S, Kannan L. Perturbation Training for Fall-Risk Reduction in Healthy Older Adults: Interference and Generalization to Opposing Novel Perturbations Post Intervention. *Front Sports Act Living*. 2021;3:697169.
- Chayasit P, Hollands K, Hollands M, Boonsinsukh R. Immediate effect of voluntary-induced stepping response training on protective stepping in persons with chronic stroke: a randomized controlled trial. *Disabil Rehabil*. 2022;44(3):420-7.
- Chayasit P, Hollands K, Hollands M, Boonsinsukh R. Characteristics of Voluntary-induced Stepping Response in Persons with Stroke compared with those of healthy Young and Older Adults. *Gait Posture*. 2020;82:75-82.
- Jensen JL, Brown LA, and Woollacott MH. Compensatory Stepping: The Biomechanics of a Preferred Response Among Older Adults. *Experimental Aging Research*. 2001;27(4):361-76.
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-98.
- Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. *J Gerontol A Biol Sci Med Sci*. 1995;50A(1):M28-M34.
- Bohannon R, Shove M, Barreca S, Masters L, Sigouin C. Five-repetition sit-to-stand test performance by community-dwelling adults: A preliminary investigation of times, determinants, and relationship with self-reported physical performance. *Isokinet Exerc Sci*. 2007;15:77-81.
- Lord SR, Murray SM, Chapman K, Munro B, Tiedemann A. Sit-to-stand performance depends on sensation, speed, balance, and psychological status in addition to strength in older people. *J Gerontol A Biol Sci Med Sci*. 2002;57(8):M539-43.
- Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. *Phys Ther*. 1997;77(8):812-9.
- Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Phys Ther*. 2009;89(5):484-98.
- Puig-Diví A, Escalona-Marfil C, Padullés-Riu JM, Busquets A, Padullés-Chando X, Marcos-Ruiz D. Validity and reliability of the Kinovea program in obtaining angles and distances using coordinates in 4 perspectives. *PLoS One*. 2019;14(6):e0216448.
- Zhu RT, Lyu PZ, Li S, Tong CY, Ling YT, Ma CZ. How Does Lower Limb Respond to Unexpected Balance Perturbations? New Insights from Synchronized Human Kinetics, Kinematics, Muscle Electromyography (EMG) and Mechanomyography (MMG) Data. *Biosensors (Basel)*. 2022;12(6).
- Bair WN, Prettyman MG, Beamer BA, Rogers MW. Kinematic and behavioral analyses of protective stepping strategies and risk for falls among community living older adults. *Clin Biomech*

- (Bristol). 2016;36:74-82.
20. Lugade V, Lin V, Chou L-S. Center of mass and base of support interaction during gait. *Gait & Posture*. 2011;33(3):406-11.
  21. Mille ML, Rogers MW, Martinez K, Hedman LD, Johnson ME, Lord SR, et al. Thresholds for Inducing Protective Stepping Responses to External Perturbations of Human Standing. *Journal of Neurophysiology*. 2003;90(2):666-74.
  22. Zadavec M, Olenšek A, Rudolf M, Bizovičar N, Goljar N, Matjačić Z. Assessment of dynamic balancing responses following perturbations during slow walking in relation to clinical outcome measures for high-functioning post-stroke subjects. *Journal of NeuroEngineering and Rehabilitation*. 2020;17(1):85.
  23. Horak FB, Nashner LM. Central programming of postural movements: adaptation to altered support-surface configurations. *J Neurophysiol*. 1986;55(6):1369-81.
  24. Ren X, Lutter C, Kebbach M, Bruhn S, Bader R, Tischer T. Lower extremity joint compensatory effects during the first recovery step following slipping and stumbling perturbations in young and older subjects. *BMC Geriatrics*. 2022;22(1):656.
  25. Payne AM, Ting LH. Worse balance is associated with larger perturbation-evoked cortical responses in healthy young adults. *Gait & Posture*. 2020;80:324-30.
  26. Richard A, Van Hamme A, Drevelle X, Golmard J-L, Meunier S, Welter M-L. Contribution of the supplementary motor area and the cerebellum to the anticipatory postural adjustments and execution phases of human gait initiation. *Neuroscience*. 2017;358:181-9.
  27. Radhakrishnan V, Gallea C, Valabregue R, Krishnan S, Kesavadas C, Thomas B, et al. Cerebellar and basal ganglia structural connections in humans: Effect of aging and relation with memory and learning. *Frontiers in Aging Neuroscience*. 2023;Volume 15 - 2023.
  28. Zapparoli L, Mariano M, Paulesu E. How the motor system copes with aging: a quantitative meta-analysis of the effect of aging on motor function control. *Communications Biology*. 2022;5(1):79.
  29. King GW, Akula CK, Luchies CW. Age-related differences in kinetic measures of landing phase lateral stability during a balance-restoring forward step. *Gait & Posture*. 2012;35(3):440-5.
  30. Carty CP, Cronin NJ, Nicholson D, Lichtwark GA, Mills PM, Kerr G, et al. Reactive stepping behaviour in response to forward loss of balance predicts future falls in community-dwelling older adults. *Age Ageing*. 2015;44(1):109-15.
  31. Singer JC, Prentice SD, McIlroy WE. Age-related challenges in reactive control of mediolateral stability during compensatory stepping: A focus on the dynamics of restabilisation. *Journal of Biomechanics*. 2016;49(5):749-55.
  32. Tisserand R, Robert T, Chabaud P, Livet P, Bonnefoy M, Cheze L. Comparison between investigations of induced stepping postural responses and voluntary steps to better detect community-dwelling elderly fallers. *Neurophysiol Clin*. 2015;45(4-5):269-84.
  33. Hof AL. Scaling gait data to body size. *Gait Posture* [Internet]. 1996;4:222-3 [cited 2026 Feb 4].
  34. Pierrynowski MR, Galea V. Enhancing the ability of gait analyses to differentiate between groups: scaling gait data to body size. *Gait Posture*. 2001;13(3):193-201.
  35. Rygelová M, Uchytíl J, Torres IE, Janura M. Comparison of spatiotemporal gait parameters and their variability in typically developing children aged 2, 3, and 6 years. *PLoS One*. 2023;18(5):e0285558.
  36. Mobbs L, Fernando V, Fonseka RD, Natarajan P, Maharaj M, Mobbs RJ. Normative Database of Spatiotemporal Gait Metrics Across Age Groups: An Observational Case-Control Study. *Sensors* [Internet]. 2025;25(2):581.
  37. Eom GM, Kwon YR, Kim DY, Ko J, Kim JW. The influence of height on test-retest reliability of postural balance measures in healthy young adults. *J Mech Med Biol*. 2022;22(9).
  38. Unluer N, Taş S. Effects of anthropometric factors, age, gender, and foot posture on single leg balance performance in asymptomatic subjects. *Fizyoterapi Rehabilitasyon*. 2019;30:154-60.

# Development of a Nomogram That Predicts Outcomes After Radical Cystectomy for Bladder Cancer Using Data from Siriraj Hospital, Thailand

Kanawut Sooksatian, M.D., Kantima Jongjitaree, M.D., Thitipat Hansomwong, M.D., Varat Woranisarakul, M.D., Patkawat Ramart, M.D., Siros Jitraphai, M.D., Ekkarin Chotikawanich, M.D., Tawatchai Taweemonkongsap, M.D.\*

Division of Urology, Department of Surgery, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.



\*Corresponding author: Tawatchai Taweemonkongsap

E-mail: thawatchai.taw@mahidol.ac.th

Received 16 January 2026 Revised 7 February 2026 Accepted 12 February 2026

ORCID ID: <http://orcid.org/0000-0002-8969-0495>

<https://doi.org/10.33192/smj.v78i3.279910>



**ABSTRACT**

**Objective:** This study aimed to develop and validate a prognostic nomogram to estimate individualized overall survival (OS) for bladder cancer patients in Thailand undergoing radical cystectomy (RC), using data from Siriraj Hospital.

**Materials and Methods:** We retrospectively analyzed a cohort of 304 bladder cancer patients who underwent RC at Siriraj Hospital between 2012 and 2023. The patients were randomly allocated to the training (80%) and testing (20%) cohorts. Cox regression analyses were employed to identify predictors of OS from a range of clinical, pathological, and treatment-related variables. A prognostic nomogram was subsequently constructed and its performance was validated using the concordance index and the area under the receiver operating characteristic curve (AUC).

**Results:** The median patient age was 68 years and the majority of patients presented with muscle-invasive disease. The median duration of follow-up was 61 months, with a median overall survival of 51 months. Multivariate analysis identified five independent predictors of OS: age, preoperative glomerular filtration rate, type of urinary diversion, pathological N stage, and presence of lymphovascular invasion. The nomogram demonstrated strong predictive performance, with AUC values of 86.6% at 12 months, 84.0% at 36 months, and 76.6% at 60 months.

**Conclusion:** We have developed and validated a prognostic nomogram tailored for Thai bladder cancer patients undergoing RC. This tool provides individualized survival estimates and may be a valuable aid in patient counseling, risk stratification, and formulation of postoperative management strategies. Future multicenter validation and integration of molecular markers will enhance the clinical utility of the prognostic nomogram.

**Keywords:** Nomogram; cystectomy; bladder cancer; prognostic; survival (Siriraj Med J 2026;78(3):218-228)

**INTRODUCTION**

Bladder cancer is the 12<sup>th</sup> most commonly diagnosed malignancy globally, accounting for 573,278 new cases and 212,536 deaths in 2020.<sup>1</sup> In Thailand, bladder cancer is among the top ten most common malignancies and ranks 9<sup>th</sup> in men. It accounts for approximately 2,700 deaths annually.<sup>2</sup> Approximately 20%–40% of patients are diagnosed with muscle-invasive bladder cancer (MIBC), a condition that, if left untreated, is associated with a mortality rate of up to 85% within two years.<sup>3</sup> Radical cystectomy (RC) remains the standard of care for MIBC; however, survival outcomes are heterogeneous. Reports from Europe and China indicate 5-year survival rates after RC ranging from 54.5% to 68%<sup>4</sup>, whereas data from Thailand suggest survival rates between 47.1% and 50%.<sup>5,6</sup>

In recent years, the field of uro-oncology has witnessed an increasing emphasis on personalized medicine, with nomograms emerging as practical instruments for individualized risk prediction. Nomograms integrate multiple clinical and pathological parameters to generate patient-specific prognostic probabilities, thereby supporting shared decision-making between clinicians and their patients. Several studies have demonstrated that predictive models that incorporate demographic, pathological, and treatment variables can improve the accuracy of survival estimation following RC. Nevertheless, such models have been developed largely using data from Western populations, and their applicability to Southeast Asian cohorts remains uncertain.<sup>7-11</sup> At Siriraj Hospital,

continuous clinical research on bladder cancer and radical cystectomy has been conducted over many years, resulting in the systematic collection of comprehensive clinical, pathological, and treatment-related data.<sup>12-14</sup> This study was conducted to identify relevant prognostic factors and to construct a nomogram to predict overall survival (OS) in Thai bladder cancer patients following RC.

**MATERIALS AND METHODS**

This study was approved by the Institutional Review Board of the Siriraj Hospital Faculty of Medicine (SIRB) (COA no. Si 192/2024). The study was a single-center retrospective analysis of patients with bladder cancer aged  $\geq 18$  years who underwent radical cystectomy at the Division of Urology, Department of Surgery, Siriraj Hospital, between January 1, 2012 and December 31, 2023. Patients were excluded if they had incomplete medical records, metastatic bladder cancer, a concurrent malignancy, or had undergone palliative surgery or non-standard treatments such as alternative medicine. The variables analyzed were selected based on their clinical relevance and availability within the dataset. These variables encompassed demographic and clinical information, pathological tumor characteristics, surgical details and perioperative treatments. The factors were chosen and synthesized from published studies to ensure the inclusion of all potential variables that might influence survival outcomes.<sup>15-23</sup> Collectively, these variables provide a

comprehensive overview of the patients' clinical, surgical and oncological profiles. OS was defined as the time from the date of radical cystectomy to death from any cause. Patients who were alive at the time of last follow-up or had incomplete follow-up were not excluded; instead, they were censored at the date of last known follow-up. Survival outcomes were estimated using the Kaplan-Meier method. Prognostic factors for OS were identified using univariate and multivariate Cox proportional hazard models. Subsequently, a prognostic nomogram was developed based on the independent predictors identified in the multivariate analysis. The model's validation was performed using the concordance index (C-index) and the area under the receiver operating characteristic curve (AUC) at 12, 36 and 60 months. All statistical analyses were conducted using SPSS software (version 21.0; IBM Corp., Armonk, NY, USA) and R program (version 4.3.3).

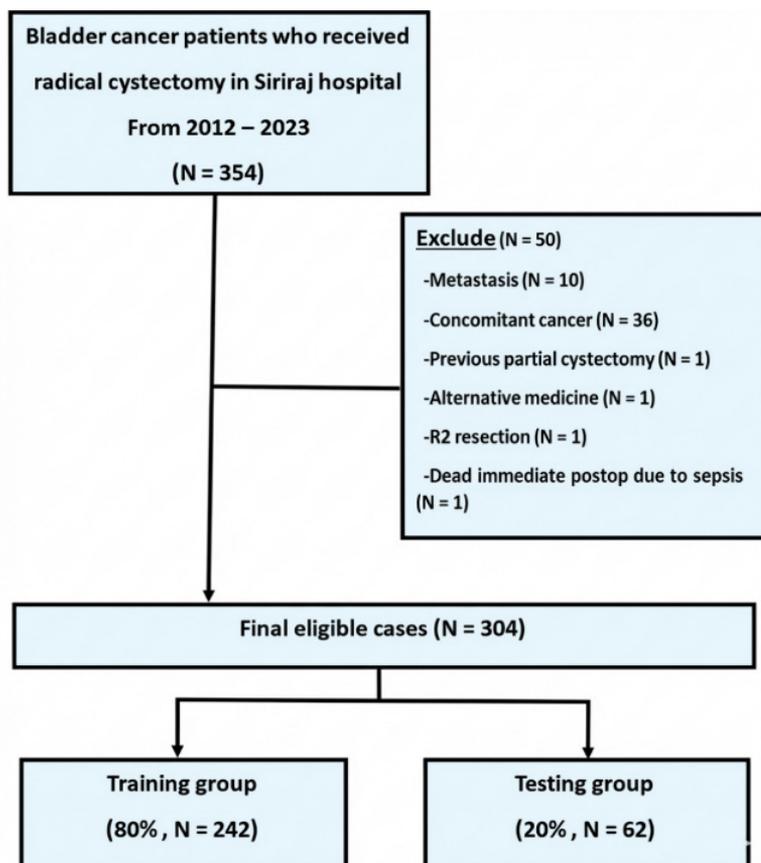
## RESULTS

### Patient characteristics

A total of 304 patients were included in the final analysis. The median age was 68 years and the cohort was predominantly male. Most of the patients presented with muscle-invasive disease and the predominant surgical approach was open radical cystectomy with ileal conduit diversion. A minority of patients received neoadjuvant

chemotherapy. The patient cohort was randomly divided into a training set (n=242, 80%) for model development and a testing set (n=62, 20%) for validation (Fig 1). There were no statistically significant differences in baseline characteristics, including age, disease stage and treatment patterns, between the two cohorts (all P-values > 0.05), confirming the appropriateness of the randomization.

Baseline demographic, preoperative, surgical and oncological characteristics indicated that the majority of patients were older adults, predominantly male, with a median age approaching 70 years. Parameters such as renal function, body mass index and American Society of Anesthesiologists Physical Status (ASA) Classification System were comparable between the groups. A small proportion of patients received neoadjuvant chemotherapy. The majority of the patients underwent open radical cystectomy, with ileal conduit being the most common form of urinary diversion, followed by neobladder and other types. Most of the patients had muscle-invasive disease at the time of surgery and a subset had nodal involvement or lymphovascular invasion (LVI). Overall, no significant differences were observed between the training and testing cohorts, which supports the validity of the random allocation (Table 1 and Table 2). The median follow-up time for this study was 61 months, and the median OS was 51 months.



**Fig 1.** Flow diagram for filtering and selecting patient records from the Siriraj database.

**TABLE 1.** Demographic and preoperative data.

	Total data	Training data (N = 242)	Testing data (N = 62)	P-value
Age (Years)	67.5±14.0	68±14.0	66.5±12.8	0.482
Age group, n (%)				0.103
<60	63(20.7)	53(21.9)	10(16.1)	
60-69	112(36.8)	81(33.5)	31(50.0)	
70-79	97(31.9)	80(33.1)	17(27.4)	
>80	32(10.5)	28(11.6)	4(6.5)	
Sex, n (%)				0.303
Male	240(78.9)	194(80.2)	46(74.2)	
Female	64(21.1)	48(19.8)	16(25.8)	
BMI	23.4±6.0	23.6±6.1	22.4±5.3	0.076
BMI, n (%)				0.515
<18.5	28(9.2)	22(9.1)	6(9.7)	
18.5-25	174(57.2)	135(55.8)	39(62.9)	
>25	102(33.6)	85(35.1)	17(27.4)	
ASA classification, n (%)				0.954
1	57(18.8)	45(18.6)	12(19.4)	
2	167(54.9)	134(55.4)	33(53.2)	
>=3	80(26.3)	63(26.0)	17(27.4)	
Pre-operative GFR	57.1±32.0	59.0±30.0	49.5±33.3	0.050
Pre-operative GFR, n (%)				0.169
>60	135(44.4)	114(47.1)	21(33.9)	
30-60	134(44.1)	102(42.1)	32(51.6)	
<30	80(26.3)	26(10.7)	9(14.5)	
Pre-operative Hydronephrosis, n (%)				0.170
No	156(51.3)	129(53.3)	27(43.5)	
Yes	148(48.7)	113(46.7)	35(56.5)	
Primary tumor size, n (%)				0.010
<4cm	118(38.8)	104(43.0)	14(22.6)	
>=4cm	132(43.4)	97(40.1)	35(56.5)	
Unknown	54(17.8)	41(16.9)	13(21.0)	
Tumor number, n (%)				0.588
Solitary	176(58.3)	143(59.6)	33(53.2)	
Multiple	105(34.5)	80(33.3)	25(40.3)	
Diffuse	21(6.9)	17(7.1)	4(6.5)	
Tumor morphology, n (%)				0.897
Papillary	160(52.6)	129(53.3)	31(50.0)	
Sessile	93(30.6)	73(30.2)	20(32.3)	
Flat lesion	51(16.8)	40(16.5)	11(17.7)	
Primary tumor location, n (%)				0.014
Trigone	55(18.3)	37(15.5)	18(29.0)	
Dome	73(24.3)	56(23.4)	17(27.4)	0.514
Lateral wall	195(64.8)	156(65.3)	39(62.9)	0.870
Anterior wall	81(26.9)	60(25.1)	21(33.9)	0.165
Posterior wall	147(48.8)	116(48.5)	31(50.0)	0.837
Bladder neck	58(19.3)	42(17.6)	16(25.8)	0.143
Time Diagnosis to surgery	91±118.0	91±117.0	90+119.0	0.661
Time Diagnosis to surgery, n (%)				1.000
<=90 day	150(50)	119(50)	31(50.0)	
>90	150(50)	119(50)	31(50.0)	
Clinical T stage, n (%)				0.713
Localized disease (stage <=2)	115(37.8)	92(38)	22(35.5)	
Locally advanced disease (stage>3)	190(62.5)	150(62)	40(64.5)	

**Abbreviations:** BMI, body mass index; ASA, American Society of Anesthesiologists Physical Status (ASA) Classification System; GFR, glomerular filtration rate;

**TABLE 2.** Surgical and Oncological data.

	Total data (N = 304)	Training data (N = 242)	Testing data (N = 62)	P-value
Surgical approach, n (%)				0.901
Open	241 (79.3)	193 (79.8)	48 (77.4)	
Laparoscopic	24 (7.9)	19 (7.9)	5 (8.1)	
Robotic assisted	39 (12.8)	30(12.4)	9 (14.5)	
Type of diversion, n (%)				0.574
Ileal conduit	252 (82.9)	198 (81.8)	54 (87.1)	
Neobladder	29 (9.5)	24 (9.9)	5 (8.1)	
Ureterostomy	23 (7.6)	20 (8.3)	3 (4.8)	
Histology, n (%)				0.202
Urothelial carcinoma	289 (95.1)	232 (95.9)	57 (91.9)	
Non-urothelial carcinoma	15 (4.9)	10 (4.1)	5 (8.1)	
Pathologic T stage, n (%)				0.448
T0	12 (3.9)	11 (4.5)	1 (1.6)	
T1	24 (7.9)	17 (7.0)	7 (11.3)	
T2	93 (30.6)	78 (32.2)	15 (24.2)	
T3	115 (37.8)	90 (37.2)	25 (40.3)	
T4	60 (19.7)	46 (19)	14 (22.6)	
N stage, n (%)				0.190
N0	195 (64.1)	156 (64.5)	39 (62.9)	
N1	45 (14.8)	40 (16.5)	5 (8.1)	
N2	50 (16.4)	36 (14.9)	14 (22.6)	
N3	11 (3.6)	10 (4.1)	4 (6.5)	
Lymphovascular invasion, n (%)				0.739
No	161 (52.7)	127 (52.5)	34 (54.8)	
Yes	143 (47.0)	115 (47.5)	28 (45.2)	
Margin positive, n (%)				0.820
No	242 (79.6)	192 (79.3)	50 (80.6)	
Yes	62 (20.4)	50 (20.7)	12 (19.4)	
Carcinoma in situ, n (%)				0.325
No	252 (82.9)	198 (81.8)	54 (87.1)	
Yes	52 (17.1)	44 (18.2)	8 (12.9)	
Chemotherapy				
Pre-operative	90 (29.6)	76 (31.4)	14 (22.6)	0.174
Post-operative	46 (15.2)	38 (15.8)	8 (12.9)	0.575
Radiation therapy, n (%)				
Pre-operative	1 (0.3)	1 (0.4)	0 (0)	1.000
Post-operative	12 (4.0)	11 (4.6)	1 (1.6)	0.471

**Identification of prognostic factors**

Univariate analysis was performed on 21 potential prognostic factors. Of these, 11 were significantly associated with OS at a significance threshold of  $P < 0.15$ . These factors included age, body mass index (BMI), preoperative glomerular filtration rate (GFR), presence of

hydronephrosis, tumor location, surgical approach, type of urinary diversion, pathological T and N stage, LVI, and margin status (Table 3). Subsequent multivariate Cox regression analysis identified five independent predictors of OS: age, preoperative GFR, type of urinary diversion, pathological N stage, and LVI (Table 4).

**TABLE 3.** Predictive factors of 5-year overall survival.

Factor	Univariate analysis			P-value
	Unadjusted HR (95%CI)			
Age				
<60	Reference			
60-69	1.043	0.632	1.720	0.870
70-79	1.207	0.734	1.985	0.459
>80	2.195	1.222	3.942	0.009
Sex				
Male	Reference			
Female	0.892	0.567	1.404	0.622
BMI				
<18.5	Reference			
18.5-25	0.759	0.427	1.351	0.349
>25	0.607	0.330	1.118	0.109
ASA				
1	Reference			
2	1.361	0.835	2.217	0.217
>=3	1.478	0.849	2.573	0.167
Pre-operative GFR				
>60	Reference			
30-60	1.651	1.119	2.436	0.012
<30	2.495	1.476	4.216	0.001
Pre-operative hydronephrosis				
No	Reference			
Yes	1.348	0.948	1.918	0.097
Primary tumor size				
<4cm	Reference			
>=4cm	0.992	0.671	1.466	0.966
Unknown	0.936	0.571	1.533	0.793
Tumor number				
Solitary	Reference			
Multiple	1.200	0.824	1.749	0.341
Diffuse	1.101	0.529	2.293	0.797
Tumor morphology				
Papillary	Reference			
Sessile	0.934	0.625	1.397	0.741
Flat lesion	0.782	0.473	1.294	0.339
Primary tumor location				
Trigone	1.345	0.855	2.117	0.200
Dome	1.035	0.686	1.564	0.869
Lateral wall	1.193	0.810	1.756	0.372
Bilateral lateral wall	1.234	0.661	2.304	0.509
Anterior wall	1.126	0.757	1.673	0.558
Posterior wall	1.183	0.830	1.685	0.353
Bladder neck	1.394	0.897	2.165	0.140

**TABLE 3.** Predictive factors of 5-year overall survival. (Continue)

Factor	Univariable analysis Unadjusted HR (95%CI)			P-value
Time Diagnosis to surgery				
<=90 day	Reference			
>90	0.988	0.687	1.422	0.949
Surgical approach				
Open	Reference			
Laparoscopic	1.007	0.540	1.880	0.982
Robotic assisted	0.371	0.163	0.845	0.018
Type of diversion				
Ileal conduit	Reference			
Neobladder	0.339	0.138	0.833	0.018
Ureterostomy	2.426	1.408	4.181	0.001
Histology				
Urothelial carcinoma	Reference			
Non-urothelial carcinoma	0.620	0.229	1.680	0.347
Pathologic T stage				
T0	Reference			
T1	4.330	0.541	34.630	0.167
T2	2.974	0.404	21.910	0.285
T3	7.644	1.058	55.250	0.044
T4	8.765	1.197	64.200	0.033
N stage				
N0	Reference			
N positive	3.553	2.483	5.084	<0.001
Lymphovascular invasion				
No	Reference			
Yes	2.656	1.846	3.821	<0.001
Margin positive				
No	Reference			
Yes	1.687	1.131	2.514	0.010
Carcinoma in situ				
No	Reference			
Yes	1.093	0.699	1.710	0.695
Chemotherapy				
Pre-operative	0.769	0.507	1.167	0.216
Post-operative	0.799	0.490	1.303	0.368
Radiation therapy				
Pre-operative	0.000	0.000	0.000	0.995
Post-operative	1.201	0.586	2.461	0.617

**Abbreviations:** BMI, body mass index; ASA, American Society of Anesthesiologists Physical Status (ASA) Classification System; GFR, glomerular filtration rate;

**TABLE 4.** Predictive factors of 5-year overall survival.

Factor	Univariable analysis				Multivariable analysis			
	Unadjusted HR (95%CI)			P-value	Adjusted HR (95%CI)			P-value
Age								
<60	Reference							
60-69	1.043	0.632	1.720	0.870	1.26	0.76	2.10	0.375
70-79	1.207	0.734	1.985	0.459	1.45	0.86	2.44	0.164
>80	2.195	1.222	3.942	0.009	3.33	1.75	6.34	0.000
BMI								
<18.5	Reference							
18.5-25	0.759	0.427	1.351	0.349				
>25	0.607	0.330	1.118	0.109				
Pre-operative GFR								
>60	Reference							
30-60	1.651	1.119	2.436	0.012	1.220	0.807	1.845	0.346
<30	2.495	1.476	4.216	0.001	1.935	1.093	3.428	0.024
Pre-operative hydronephrosis								
No	Reference							
Yes	1.348	0.948	1.918	0.097				
Primary tumor location								
Bladder neck	1.394	0.897	2.165	0.140				
Surgical approach								
Open	Reference							
Laparoscopic	1.007	0.540	1.880	0.982				
Robotic assisted	0.371	0.163	0.845	0.018				
Type of diversion								
Ileal conduit	Reference							
Neobladder	0.339	0.138	0.833	0.018	0.44	0.18	1.09	0.076
Ureterostomy	2.426	1.408	4.181	0.001	1.90	1.07	3.39	0.029
Pathologic T stage								
T0	Reference							
T1	4.330	0.541	34.630	0.167				
T2	2.974	0.404	21.910	0.285				
T3	7.644	1.058	55.250	0.044				
T4	8.765	1.197	64.200	0.033				
N stage								
N0	Reference							
N positive	3.553	2.483	5.084	<0.000	2.98	1.96	4.54	<0.001
Lymphovascular invasion								
No	Reference							
Yes	2.656	1.846	3.821	<0.000	1.88	1.22	2.89	0.004
Margin positive								
No	Reference							
Yes	1.687	1.131	2.514	0.010				

**Abbreviations:** BMI, body mass index; GFR, glomerular filtration rate;

### Development and validation of a prognostic nomogram

A nomogram was constructed incorporating the five independent prognostic factors. This nomogram uses a point-based scoring system to estimate the 1-, 3-, and 5-year OS probabilities for individual patients (Fig 2). The predictive accuracy of the nomogram was confirmed through C-index and ROC analysis. The model demonstrated a high discriminatory power, with AUC values of 86.61% (95% CI 77.2–96.0) at 12 months, 83.97% (95% CI 73.0–94.9) at 36 months, and 76.56% (95% CI 62.0–91.1) at 60 months (Fig 3).

### DISCUSSION

This study represents the first initiative to develop a prognostic nomogram specifically for bladder cancer patients undergoing RC in Thailand. The model identified age, preoperative GFR, type of urinary diversion, lymph

node involvement (N stage) and LVI as key independent prognostic factors. These findings are consistent with the results of earlier studies conducted in Western and Asian populations. It should be noted that pathological T stage including subgroup analysis comparing pT0–2 versus pT3–4 did not retain its significance in the final multivariate model, an observation likely attributable to the multicollinearity with N stage and LVI, both of which more directly reflect tumor aggressiveness and the potential for systemic dissemination.

Each variable included in the final model has previously been established as a powerful prognostic marker in bladder cancer. Age is a known independent factor associated with both cancer-specific and all-cause mortality, as older patients typically present with more comorbidities and reduced physiological reserve.<sup>7</sup> N stage remains one of the most consistent predictors of a poor

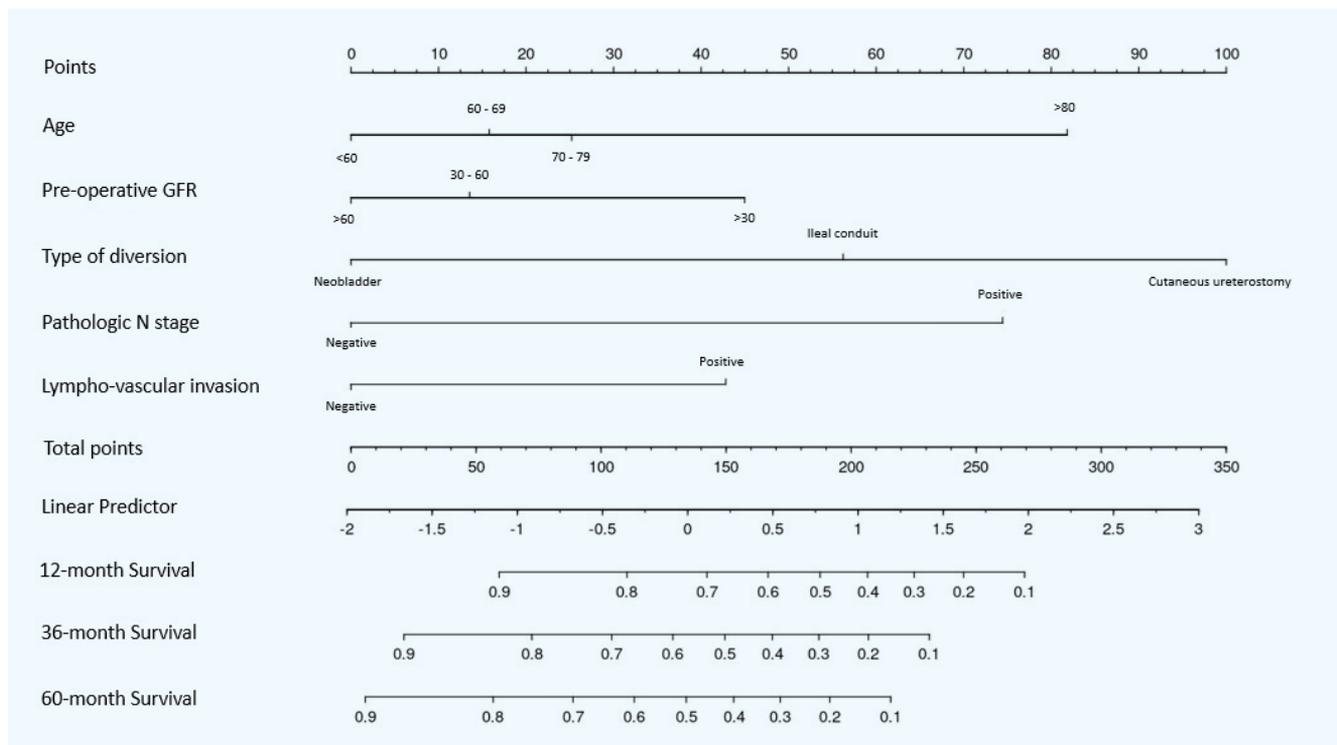


Fig 2. Nomogram predicting overall survival probability of 1- 3- and 5-years after radical cystectomy. Variables include Age, Pre-operative glomerular filtration rate (GFR), Type of diversion, Pathologic N stage, and Lympho-vascular invasion.

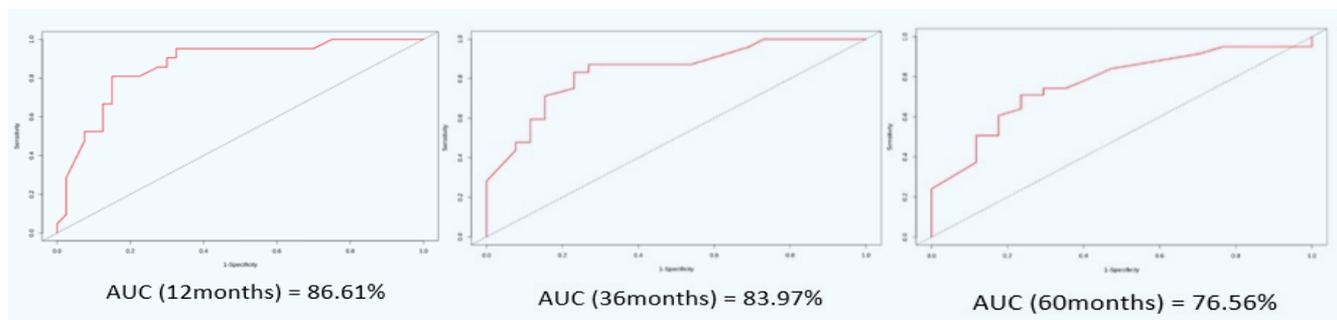


Fig 3. Validation of the nomogram.

prognosis, given that nodal metastasis is indicative of systemic spread and necessitates aggressive multimodal therapy.<sup>11</sup> A reduced preoperative GFR not only restricts eligibility for cisplatin-based chemotherapy but also serves as an indicator of overall health status, which can impact long-term outcomes.<sup>17</sup> Furthermore, LVI is a well-established marker of aggressive tumor biology and metastatic potential.<sup>21</sup> The type of diversion is a significant factor in this study. Patients undergoing neobladder reconstruction typically have fewer comorbidities and a better baseline functional status compared to those receiving other types of diversion, which may contribute to better survival outcomes. However, we still included this factor in the development of the nomogram because the type of diversion is an important postoperative factor applicable to every patient after radical cystectomy.<sup>8</sup>

Neoadjuvant chemotherapy (NAC), despite being recommended by international guidelines for muscle-invasive bladder cancer, was not a significant predictor in our model. This finding is in alignment with a large SEER cohort study of T2-4N0-3M0 MIBC, in which, after the inverse probability of treatment weighting, NAC did not confer a statistically significant OS advantage over adjuvant chemotherapy. However, subgroup analysis revealed that in patients without lymph node involvement (N0), NAC was associated with superior OS and cancer-specific survival compared to adjuvant chemotherapy, suggesting that nodal status may mediate the benefit of NAC in MIBC. The lack of significance in our cohort may be due to selection bias, as patients receiving NAC often present with more advanced disease or worse baseline characteristics, potentially offsetting the therapeutic benefit. Additionally, variability in chemotherapy regimens, incomplete treatment courses, and the absence of a centralized response evaluation may dilute the observed impact. Immune checkpoint inhibitors were not included in the present analysis, as their use was limited to the most recent 2–3 years and involved a very small number of patients, precluding meaningful statistical analysis.

In our cohort, the low uptake of NAC and inconsistencies in treatment documentation could further diminish statistical power. Moreover, our study did not include data on pathologic downstaging post-NAC, which is a critical intermediate endpoint correlated with long-term benefit.<sup>24</sup> Consequently, the non-significance of NAC in this study does not imply a lack of benefit, but rather highlights the challenges of capturing its value in retrospective, real-world datasets.

The nomogram developed in this study provides a clinically practical tool for individualized risk prediction. It enables clinicians to estimate OS at multiple time

points, thereby enhancing prognostic counseling, tailoring surveillance strategies, and informing treatment planning. Patients identified as high-risk by the model may be considered for more intensive follow-up or additional therapeutic interventions.

Several limitations of this study should be acknowledged. The retrospective single-center design limits the generalizability of the findings, and the presence of missing data could have influenced variable selection and statistical significance. Certain potentially relevant prognostic factors, such as smoking status, nutritional parameters and molecular or genetic markers, were not available for analysis. Furthermore, the study lacks external validation in an independent cohort. The relatively small number of patients who received NAC may also have limited the statistical power to detect a survival benefit associated with this treatment.

In practical application, this nomogram can be implemented as a web-based or electronic medical record-integrated tool, allowing clinicians to input patient-specific variables to generate immediate survival estimates. This can help guide patient counseling and facilitate risk-adapted follow-up protocols, ultimately contributing to improved patient outcomes and more efficient resource allocation.

## CONCLUSION

We have successfully developed and validated a nomogram that predicts OS in bladder cancer patients following RC at Siriraj Hospital. This model provides clinicians with an individualized prediction tool tailored to the Thai population, which can assist in shared decision-making and postoperative planning. Future research should focus on prospective multicenter validation and the integration of molecular biomarkers to strengthen the reliability and clinical utility of the model.

## ACKNOWLEDGEMENT

The authors would like to express their gratitude to Dr. Saowalak Hunnangkul and Terasut Numwong for valuable assistance and support in this study.

## DECLARATIONS

### Grants and Funding Information

None.

### Conflict of Interest

All authors declare no personal or professional conflicts of interest, and no financial support from the companies that produce and/or distribute the drug, devices, or materials described in this report.

**Registration Number of Clinical Trial**

None.

**Author Contributions**

Conceptualization and methodology: K.S., T.T., T.H.; Investigation: K.S., V.W.; Formal analysis: K.S., K.J., P.R.; Visualization and writing—original draft: K.S., T.T.; Writing-review and editing: T.T., E.C., S.J.; Supervision: T.T.

**Ethical Approval Statement**

This study was approved by the Institutional Review Board of the Siriraj Hospital Faculty of Medicine (SIRB) (COA no. Si 192/2024).

**Use of Artificial Intelligence**

No artificial intelligence tools or technologies were used in the writing and analysis.

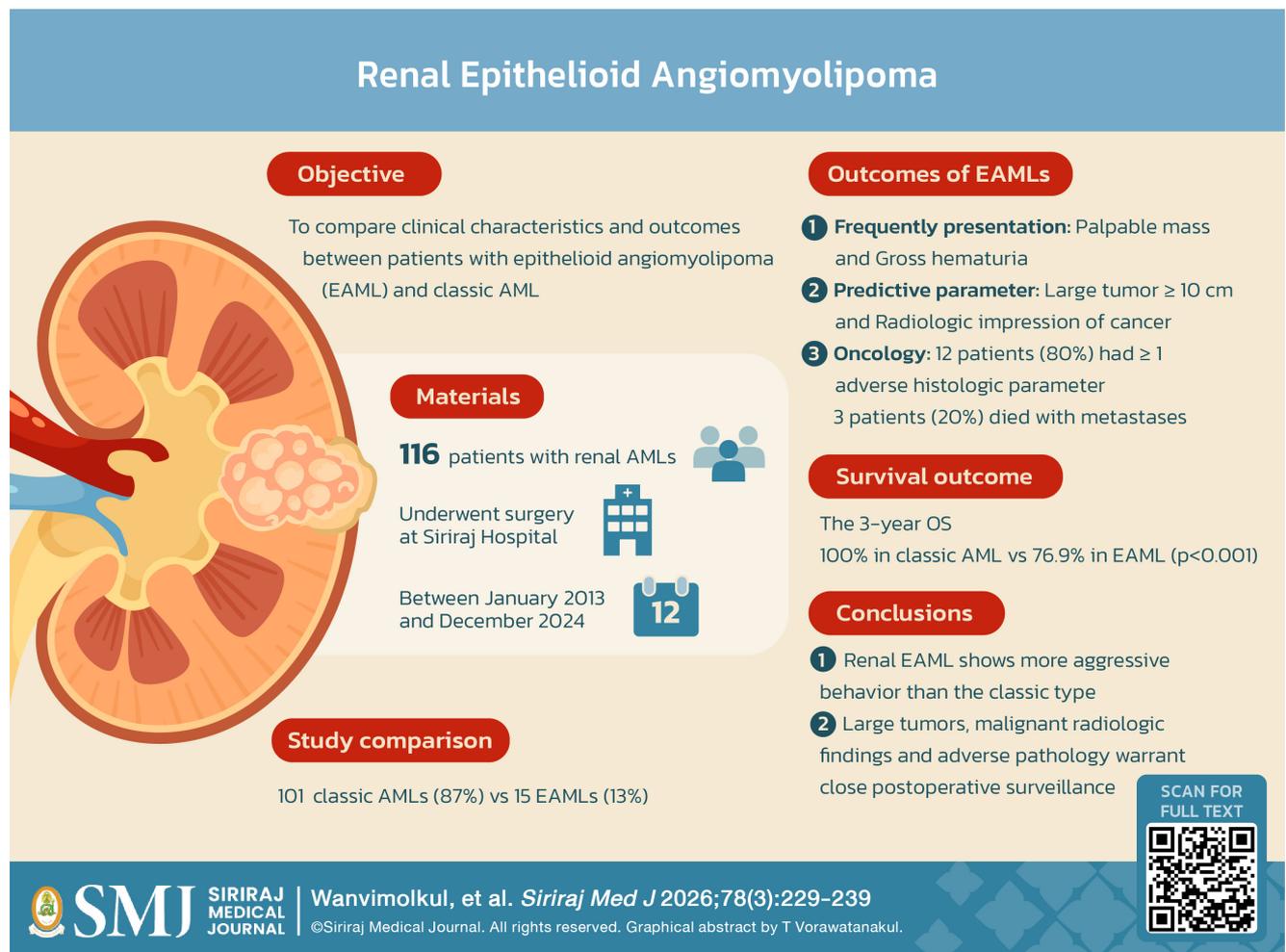
**REFERENCES**

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-49.
- Cancer. IAFRo. Thailand fact sheet: GLOBOCAN 2020. . Lyon, France: International Agency for Research on Cancer; 2020.
- Aboumarzouk OM, Drewa T, Olejniczak P, Chlosta PL. Laparoscopic versus open radical cystectomy for muscle-invasive bladder cancer: a single institute comparative analysis. *Urol Int.* 2013;91(1):109-12.
- Zhang ZL, Dong P, Li YH, Liu ZW, Yao K, Han H, et al. Radical cystectomy for bladder cancer: oncologic outcome in 271 Chinese patients. *Chin J Cancer.* 2014;33(3):165-71.
- Ramart P, Chaiyaprasithi B, Pradniwat K, Ratanarapee S, Amornvesukit T, Taweemongkongsap T, et al. Outcome of open radical cystectomy with pelvic lymph node dissection for bladder urothelial cancer in Siriraj hospital between 1998-2003. *Insight Urology.* 2010;31(1):27-39.
- Siriboonpipattana N, Nualyong C, Taweemonkongsap T, Leewansangtong S, Ramart P, Amornvesukit T. The Oncologic Outcome of Laparoscopic Radical Cystectomy for Invasive Bladder Cancer in Siriraj Hospital Between 2005-2013. *Siriraj Medical Journal.* 2017;69(6):377-83.
- Shariat SF, Karakiewicz PI, Palapattu GS, Amiel GE, Lotan Y, Rogers CG, et al. Nomograms provide improved accuracy for predicting survival after radical cystectomy. *Clin Cancer Res.* 2006;12(22):6663-76.
- Yang Z, Bai Y, Liu M, Hu X, Han P. Development and validation of a prognostic nomogram for predicting cancer-specific survival after radical cystectomy in patients with bladder cancer: A population-based study. *Cancer Med.* 2020;9(24):9303-14.
- Osawa T, Abe T, Takada N, Ito YM, Murai S, Shinohara N. Validation of the nomogram for predicting 90-day mortality after radical cystectomy in a Japanese cohort. *Int J Urol.* 2018;25(7):699-700.
- Liu P, Xu L, Chen G, Shi B, Zhang Q, Chen S. Nomograms for predicting survival in patients with micropapillary bladder cancer: a real-world analysis based on the surveillance, epidemiology, and end results database and external validation in a tertiary center. *BMC Urol.* 2023;23(1):16.
- Wang J, Wu Y, He W, Yang B, Gou X. Nomogram for predicting overall survival of patients with bladder cancer: A population-based study. *Int J Biol Markers.* 2020;35(2):29-39.
- Veerakulwatana S, Suk-Ouichai C, Taweemonkongsap T, Chotikawanich E, Jitraphai S, Woranisarakul V, et al. Perioperative factors and 30-day major complications following radical cystectomy: A single-center study in Thailand. *Heliyon.* 2024;10(13):e33476.
- Mahalelakul A, Assavavirojekul P, Leewansangtong S, Woranisarakul V, Hansomwong T, Srinualnad S. Outcomes of Robot-assisted Radical Prostatectomy in Men Aged 75 Years Old or Older: A Single-center Study in Thailand. *Siriraj Medical Journal.* 2025;77(1):22-8.
- Sornthai W, Teyateeti A, Taweemonkongsap T, Jitraphai S, Woranisarakul V, Jongjitree K, et al. Preoperative myosteotosis and perioperative serum chloride levels predict 180 day major complications after radical cystectomy. *Sci Rep.* 2025;15(1):3184.
- Lee CT, Dunn RL, Chen BT, Joshi DP, Sheffield J, Montie JE. Impact of body mass index on radical cystectomy. *J Urol.* 2004;172(4 Pt 1):1281-5.
- Mayr R, May M, Martini T, Lodde M, Pycha A, Comploj E, et al. Predictive capacity of four comorbidity indices estimating perioperative mortality after radical cystectomy for urothelial carcinoma of the bladder. *BJU Int.* 2012;110(6 Pt B):E222-7.
- Kim D, Nam W, Kyung YS, You D, Jeong IG, Hong B, et al. Effect of decreased renal function on poor oncological outcome after radical cystectomy. *Investig Clin Urol.* 2023;64(4):346-52.
- Cheng L, Neumann RM, Scherer BG, Weaver AL, Leibovich BC, Nehra A, et al. Tumor size predicts the survival of patients with pathologic stage T2 bladder carcinoma: a critical evaluation of the depth of muscle invasion. *Cancer.* 1999;85(12):2638-47.
- Chen H, Hong Y, Yu B, Ruiqian L, Jun L, Hongyi W, et al. Retrospective analysis of bladder cancer morphology and depth of invasion under cystoscopy. *BMC Urol.* 2022;22(1):12.
- Kimura S, Mari A, Foerster B, Abufaraj M, Vartolomei MD, Stangl-Kremser J, et al. Prognostic Value of Concomitant Carcinoma In Situ in the Radical Cystectomy Specimen: A Systematic Review and Meta-Analysis. *J Urol.* 2019;201(1):46-53.
- Kikuchi E, Margulis V, Karakiewicz PI, Roscigno M, Mikami S, Lotan Y, et al. Lymphovascular invasion predicts clinical outcomes in patients with node-negative upper tract urothelial carcinoma. *J Clin Oncol.* 2009;27(4):612-8.
- Chang SS, Hassan JM, Cookson MS, Wells N, Smith JA, Jr. Delaying radical cystectomy for muscle invasive bladder cancer results in worse pathological stage. *J Urol.* 2003;170(4 Pt 1):1085-7.
- Fallara G, Di Maida F, Bravi CA, De Groote R, Piramide F, Turri F, et al. A systematic review and meta-analysis of robot-assisted vs. open radical cystectomy: where do we stand and future perspective. *Minerva Urol Nephrol.* 2023;75(2):134-43.
- Grossman HB, Natale RB, Tangen CM, Speights VO, Vogelzang NJ, Trump DL, et al. Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N Engl J Med.* 2003;349(9):859-66.

# Clinical Characteristics and Surgical Outcomes of Renal Epithelioid Angiomyolipoma: A Comparison with the Classic Type

Nattaporn Wanvimolkul, M.D.<sup>1</sup>, Ekkarin Chotikawanich, M.D.<sup>1</sup>, Siros Jitpraphai, M.D.<sup>1</sup>, Varat Woranisarakul, M.D.<sup>1</sup>, Thitipat Hansomwong, M.D.<sup>1</sup>, Kantima Jongjitaree, M.D.<sup>1</sup>, Pongsatorn Laksanabunsong, M.D.<sup>1</sup>, Ngoentra Tantranont, M.D.<sup>2</sup>, Tawatchai Taweemonkongsap, M.D.<sup>1,\*</sup>

<sup>1</sup>Division of Urology, Department of Surgery, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, <sup>2</sup>Department of Pathology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.



\*Corresponding author: Tawatchai Taweemonkongsap

E-mail: [thawatchai.taw@mahidol.ac.th](mailto:thawatchai.taw@mahidol.ac.th)

Received 19 January 2026 Revised 12 February 2026 Accepted 14 February 2026

ORCID ID: <http://orcid.org/0000-0002-8969-0495>

<https://doi.org/10.33192/smj.v78i3.279923>



## ABSTRACT

**Objective:** To compare clinical characteristics and surgical outcomes between patients with epithelioid angiomyolipoma (EAML) and classic angiomyolipoma (AML), and to identify factors associated with EAML diagnosis.

**Materials and Methods:** All patients with renal AML who underwent surgery at Siriraj Hospital between January 2013 and December 2024 were reviewed. Clinical features and surgical outcomes were compared between patients with classic AML and those with EAML, and predictors of EAML were evaluated using multivariable analyses.

**Results:** Among 116 eligible patients, 101 had classic AML and 15 had EAML (12.9%). Most patients were female and were diagnosed in their fifth decade. Demographics, tumor laterality, prevalence of tuberous sclerosis complex gene mutation, and comorbidities did not differ between the 2 groups. Palpable mass (26.7%) and hematuria (13.3%) were more frequent in patients with EAML than in those with classic AML. Most patients with EAML underwent radical or partial nephrectomy due to suspected malignancy. In multivariable analysis, tumor size  $\geq 10$  cm (odds ratio 15.44;  $P = 0.003$ ) and a radiologic impression of cancer (odds ratio 46.98;  $P < 0.001$ ) independently predicted EAML. Four patients with EAML had adverse pathologic features and experienced poor survival; 3 patients died with metastases. The 3-year overall survival was 100% in classic AML and 76.9% in EAML ( $P < 0.001$ ).

**Conclusions:** Patients with EAML had less favorable surgical outcomes than those with classic AML. Larger tumor size and a preoperative radiologic impression of malignancy were associated with an EAML diagnosis. Adverse pathologic features in EAML suggest malignant potential.

**Keywords:** Angiomyolipoma; epithelioid; renal; surgery; classic (Siriraj Med J 2026;78(3):229-239)

## INTRODUCTION

Angiomyolipoma (AML) is the most common benign renal tumor. Histologically, AML is categorized as classic AML or epithelioid angiomyolipoma (EAML). Classic renal AML is typically benign and comprises a triphasic mixture of smooth muscle, mature adipose tissue, and thick-walled blood vessels. The 2004 World Health Organization classification of renal neoplasms defines EAML as a potentially malignant mesenchymal neoplasm characterized by predominantly epithelioid-cell proliferation.<sup>1</sup> Most AMLs are discovered incidentally on imaging, although flank pain, gross hematuria, and severe retroperitoneal hemorrhage may occur.<sup>2</sup> AML is generally diagnosed on imaging as a fat-containing renal mass; however, reduced internal fat in EAML makes radiologic differentiation from renal cell carcinoma (RCC) challenging. Surgical management depends on symptoms, tumor size and number, growth pattern, and malignant potential.<sup>3</sup> Definitive diagnosis and subtyping rely on histopathologic examination and immunohistochemical staining.

Although the global incidence of EAML is low, it displays distinct biology from classic AML and carries malignant potential; approximately 20% of patients present with local invasion or metastasis.<sup>4,5</sup> Accurate distinction between classic AML and EAML is therefore critical for treatment selection and prognostication. Several studies have associated EAML diagnosis with younger age, male sex, reduced intratumoral fat, and larger tumor size.<sup>6,7</sup> In

contrast, Tsai et al reported no significant associations of age or tumor size between noninvasive and invasive EAML.<sup>8</sup> Based on their analysis of 41 cases and review of 66 published cases, Nese et al proposed that carcinoma-like growth pattern and extrarenal extension or renal vein involvement predict outcomes.<sup>9</sup> More recently, a literature review summarized 46 EAML cases with distant metastasis and recommended long-term follow-up.<sup>10</sup> However, most research comprises case reports or small series, and comprehensive data on clinical and prognostic features remain limited.

We therefore compared clinical characteristics and surgical outcomes between EAML and classic AML in 116 consecutive patients who underwent surgery over 10 years at our center. We also evaluated factors associated with EAML diagnosis. Additionally, we identified EAML cases with malignant behavior and explored predictors of poor outcomes in this subgroup. This is the first EAML cohort study from the Southeast Asian population and may inform future management strategies.

## MATERIALS AND METHODS

### Study design and ethics

The study protocol was approved by the Siriraj Institutional Review Board (SIRB) (COA no. Si 605/2025). We conducted a retrospective review of 116 patients with renal AML who underwent surgery at the hospital between January 2013 and December 2024.

## Cohort and variables

All patients underwent partial or radical nephrectomy, with pathology confirming classic AML or EAML. Collected patient characteristics included sex, age, body mass index, symptoms, and past medical history. Diagnostic variables included radiologic impression, tumor laterality, size, bilaterality, and presence of metastasis. Indications for treatment, management type, and reasons for radical nephrectomy were recorded.

## Pathologic definitions and risk stratification

EAML was diagnosed when specimens contained  $\geq 80\%$  epithelioid cells, with characteristic positivity for melanocytic markers such as HMB-45 and melan-A and variable positivity for smooth muscle markers (Fig 1). Adverse histologic parameters followed the criteria proposed by Nese et al. since 2011.<sup>9</sup> These included presence of tuberous sclerosis and/or concurrent classic AML, tumor size  $\geq 7$  cm, carcinoma-like morphology, perinephric fat and/or renal vein involvement, and necrosis. Cases were stratified as low risk (0–1 parameter), intermediate risk (2–3 parameters), or high risk (4–5 parameters).

## Outcomes and follow-up

Follow-up data included tumor progression, time to recurrence or metastasis, and patient status at last follow-

up. Adverse outcomes were defined as local recurrence, metastasis, or cancer-specific mortality.

## Statistical analysis

Clinical characteristics and surgical outcomes were compared between classic AML and EAML. Continuous variables were summarized as means with standard deviations or medians with interquartile ranges. Between-group differences in categorical variables were tested using the chi-square test or Fisher's exact test, as appropriate. Overall survival was estimated by the Kaplan–Meier method and compared using the log-rank test. Factors associated with EAML diagnosis were identified using univariable and multivariable logistic regression. All tests were 2-sided, with  $P < 0.05$  considered statistically significant. Analyses were performed using PASW Statistics version 18 (SPSS Inc, Chicago, IL, USA).

## RESULTS

### Cohort characteristics

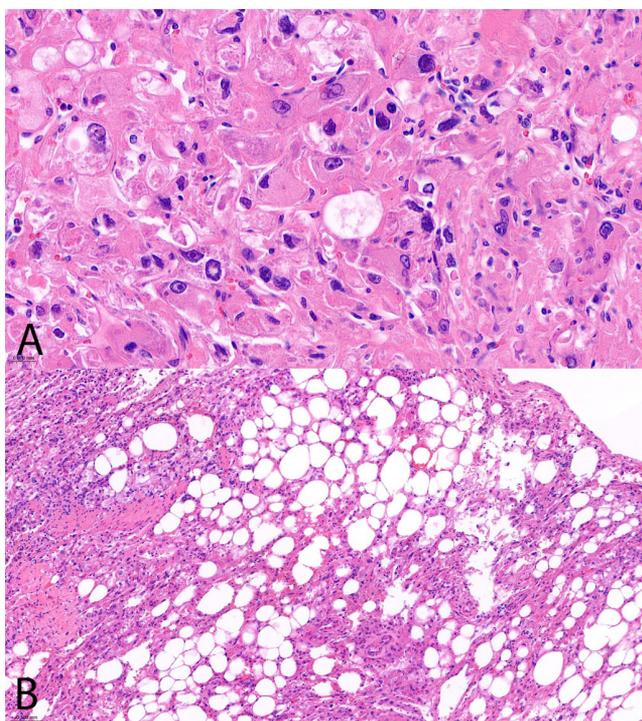
A total of 116 patients were identified: 101 with classic AML (87%) and 15 with EAML (13%). Median age did not differ between the 2 groups (55 vs 53 years;  $P = 0.23$ ), and females predominated in both. Baseline features—including tumor laterality, bilaterality, tuberous sclerosis complex, prior tumor rupture, and comorbidities—were comparable. Clinical characteristics and surgical outcomes are summarized in Table 1.

### Presentation and imaging

Patients with classic AML had higher body mass index (mean body mass index 24.0 vs 20.6;  $P = 0.024$ ) and were more often asymptomatic (59.4% vs 26.7%;  $P = 0.024$ ). By contrast, patients with EAML more frequently presented with gross hematuria (13.3% vs 1.0%;  $P = 0.044$ ) and a palpable mass (26.7% vs 5.0%;  $P = 0.016$ ). On imaging, a fat-containing mass consistent with AML was identified more often in classic AML (66.3% vs 13.3%;  $P < 0.001$ ). An impression of RCC was more common in EAML (73.3% vs 23.8%;  $P < 0.001$ ). Representative Computed Tomography (CT scan) image of epithelioid AML from our cohort was presented in Fig 2.

### Treatment indications and surgical approach

Indications for surgery in classic AML included suspected malignancy (34.6%), increasing tumor size or size  $\geq 4$  cm (34.7%), and symptoms or prior rupture (27.7%). The distribution of indications differed significantly between groups: EAML surgeries were driven by suspected malignancy in 13 cases (86.6%) and by symptoms or prior rupture in 2 cases (13.4%). Surgical approach



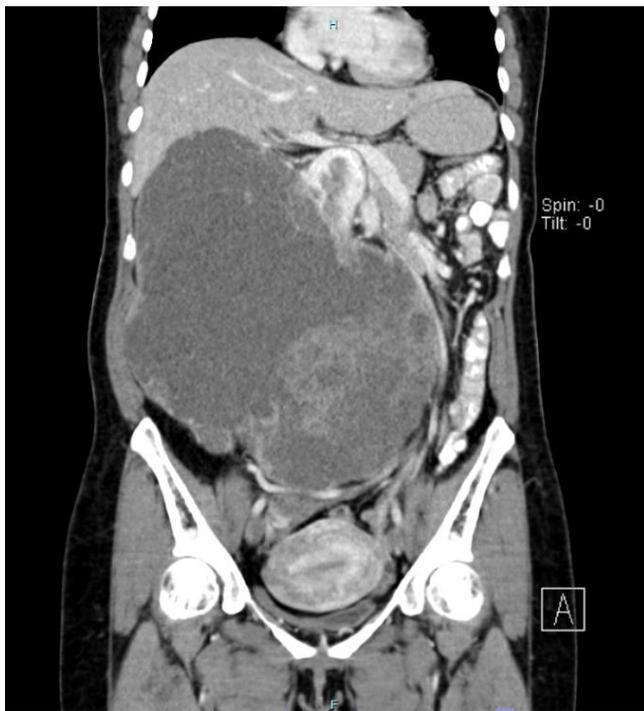
**Fig 1.** Epithelioid angiomyolipoma of the kidney (H&E stain). (A) The tumor comprises predominantly pleomorphic epithelioid cells with eosinophilic cytoplasm. (B) Epithelioid AML with concurrent classic AML.

**TABLE 1.** Baseline clinical characteristics, imaging impressions, and surgical management in classic AML vs EAML (January 2013–December 2024).

Parameter	Classic AML n = 101	EAML n = 15	P value
Age, median (IQR), y	55.0 (20.0)	53.0 (24.0)	0.238
Female sex, n, %	85 (84.16)	11 (73.33)	0.289
Right side, n, %	45 (44.6)	8 (53.5)	0.524
BMI, median (IQR), y	24.0 (5.6)	20.6 (4.9)	0.024
Co-morbidities, n, %			
Hypertension	32 (31.7)	3 (20)	0.548
Diabetes mellitus	14 (13.9)	2 (13.3)	1.000
CVD	2 (2.0)	1 (6.67)	0.342
Presentations, n, %			
Asymptomatic	60 (59.4)	4 (26.7)	0.024
Abdominal pain	35 (34.7)	5 (33.3)	0.920
Gross hematuria	1 (1.0)	2 (13.3)	0.044
Palpable mass	5 (5.0)	4 (26.7)	0.016
History of ruptured tumor	6 (5.9)	2 (13.3)	0.276
Bilateral AML	4 (4.0)	1 (6.7)	0.506
TSC	1 (1.0)	1 (6.7)	0.243
Tumor size, median (IQR), cm	5.6 (6.5)	10.0 (17.8)	0.828
≥ 4 cm	75 (74.3)	11 (73.3)	1.000
≥ 10 cm	26 (25.7)	9 (60)	0.013
Report of CT/MRI, n, %			
AML	67 (66.3)	2 (13.3)	< 0.001
RCC	24 (23.8)	11 (73.3)	< 0.001
Undetermined	10 (9.9)	2 (13.3)	0.653
Indication for treatment, n, %			
Symptomatic	23 (22.8)	1 (6.7)	0.191
Suspicion of malignancy	35 (34.7)	13 (86.7)	< 0.001
History of ruptured tumor	5 (5)	1 (6.7)	0.573
Pseudoaneurysm	3 (3)	0 (0)	1.000
Increasing size	13 (12.9)	0 (0)	0.213
Size ≥ 4 cm	22 (21.8)	0 (0)	0.071
Type of management, n, %			
Partial nephrectomy	94 (93.1)	8 (53.3)	< 0.001
Radical nephrectomy	7 (6.9)	7 (46.7)	< 0.001
Reason for radical nephrectomy			
Suspicion of malignancy	3 (42.9)	6 (85.7)	0.266
Complex mass	4 (57.1)	1 (14.3)	0.266
Follow-up time, median (IQR), mo	33 (62)	28 (44)	0.091

Data are reported as median (IQR) or n (%), as shown in the table. Imaging impressions derive from CT/MRI reports. Tumor size refers to maximal diameter. Follow-up is reported in months. P values are 2-sided.

**Abbreviations:** AML, angiomyolipoma; BMI, body mass index; CT, computed tomography; CVD, cardiovascular disease; EAML, epithelioid angiomyolipoma; IQR, interquartile range; mo, months; MRI, magnetic resonance imaging; RCC, renal cell carcinoma; TSC, tuberous sclerosis complex; y, years

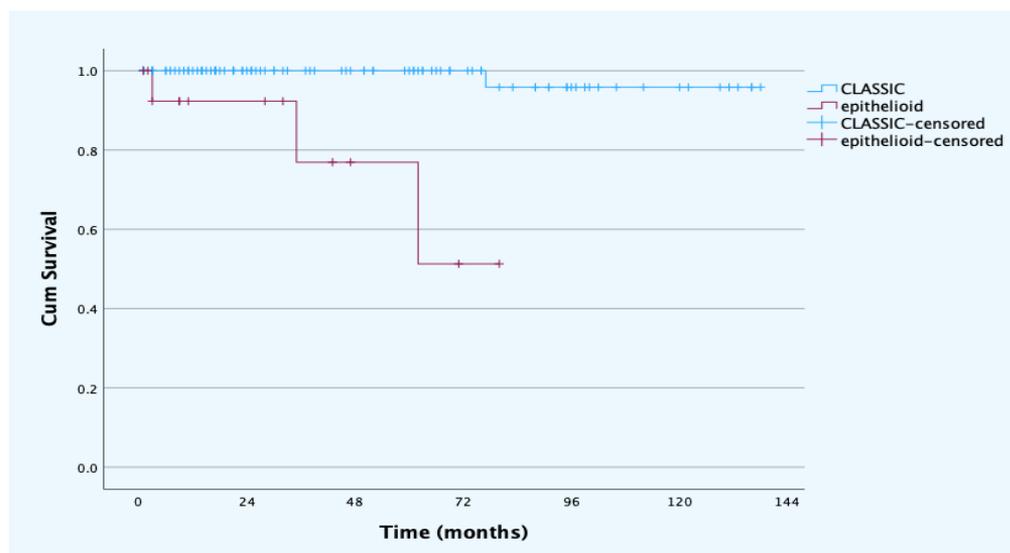


**Fig 2.** Imaging of epithelioid angiomyolipoma from our cohort. CT scan reveals a 24.4 × 15.0 × 24.0 cm irregular, encapsulated cystic mass with internal irregular nodules and septations, originating from the right kidney.

also differed, with partial nephrectomy more common in classic AML ( $P < 0.001$ ) and radical nephrectomy predominating in EAML ( $P < 0.001$ ).

### Follow-up and survival

The median follow-up was 33 months in classic AML and 28 months in EAML. All patients with classic AML remained tumor-free at last follow-up. Four EAML patients (26.7%) experienced tumor progression during follow-up, with mortality of 20% in this group. The 3-year overall survival was 100% in classic AML and 76.9% in EAML ( $P < 0.001$ ; Fig 3).



**Fig 3.** Overall survival after surgery for classic AML vs EAML (Kaplan–Meier). Curves were compared with the log-rank test ( $P < 0.001$ ). Sample sizes were  $n = 101$  for classic AML and  $n = 15$  for EAML, with median follow-up of 33 and 28 months, respectively, as per Table 1. **Abbreviations:** AML, angiomyolipoma; EAML, epithelioid angiomyolipoma

### Predictors of EAML

Univariable and multivariable analyses are summarized in Table 2. Higher body mass index was a protective factor in univariable analysis ( $P = 0.024$ ) but lost significance in multivariable models. Increasing tumor size independently predicted EAML. A tumor size  $\geq 10$  cm was associated with higher odds of EAML (odds ratio 4.32; 95% CI 1.40–13.30). A radiologic impression of RCC on computed tomography/magnetic resonance imaging also independently predicted EAML (odds ratio 15.35 compared with AML; 95% CI 3.17–74.33). Both associations remained significant in multivariable analysis.

### Clinicopathologic features and outcomes of EAML

Clinicopathologic features of patients with EAML are shown in Table 3. Of the 15 patients with EAML, 12 (80%) had  $\geq 1$  adverse histologic parameter. Localized and locally advanced T stage occurred in a 3:2 ratio. Nearly half of the patients with EAML were high risk for malignant behavior (7/15). Four patients (26.7%) experienced malignant progression with local recurrence, metastasis, or cancer-related death. One patient with positive margin and nodal metastasis (R+/N+) developed liver metastasis at 3 months postoperatively. Another patient with a positive margin without nodal metastasis (R+/N0) developed local recurrence at 9 months postoperatively. Two patients with negative margins and nodes (R0/N0) had longer progression-free survival at 29 and 36 months. No patients with distant metastases received adjuvant systemic therapy. Mortality among the 4 patients with malignant EAML was 75% (3/4). Clinicopathologic features and surgical outcomes of malignant EAML are summarized in Table 4.

**TABLE 2.** Univariable and multivariable logistic regression for factors associated with EAML diagnosis.

Variable	Univariable			Multivariable		
	OR	95% CI	P value	OR	95% CI	P value
Age	0.972	0.933–1.013	0.175			
BMI	0.819	0.688–0.974	0.024	0.816	0.656–1.016	0.069
Male sex	0.518	0.149–1.830	0.307			
Bilateral AML	1.732	0.180–16.629	0.634			
TSC	7.143	0.423–120.757	0.173			
Presentations						
Asymptomatic	0.248	0.074–0.834	0.024			
Abdominal pain	0.943	0.299–2.975	0.920			
Gross hematuria	15.385	1.303–181.707	0.030	15.689	0.063–3923.067	0.328
Palpable mass	6.982	1.629–29.922	0.009			
Report of CT/MRI						
AML			0.003			
RCC	15.35	3.172–74.330	< 0.001	46.989	5.890–374.848	< 0.001
Undetermined	6.70	0.846–53.071	0.072	10.116	0.631–162.160	0.102
Tumor size	1.091	1.016–1.171	0.017			
Tumor size ≥ 4 cm	0.868	0.265–2.844	0.815			
Tumor size ≥ 7 cm	2.385	0.787–7.221	0.124			
Tumor size ≥ 10 cm	4.327	1.404–13.330	0.011	15.443	2.593–91.983	0.003
History of ruptured tumor	2.436	0.444–13.361	0.305			

Odds ratios (ORs) are shown with 95% confidence intervals. For continuous tumor size, the OR corresponds to a 1-unit increase (cm). For “Report of CT/MRI,” AML is the reference category. *P* values are 2-sided, with *P* < 0.05 considered significant.

**Abbreviations:** AML, angiomyolipoma; BMI, body mass index; CI, confidence interval; CT, computed tomography; EAML, epithelioid angiomyolipoma; MRI, magnetic resonance imaging; OR, odds ratio; RCC, renal cell carcinoma; TSC, tuberous sclerosis complex

## DISCUSSION

First described by Mai et al, EAML is a rare AML variant with malignant potential.<sup>11</sup> Diagnosis is challenging: specific clinical manifestations are lacking, radiographic distinction is hampered by reduced intratumoral fat, and histologic differentiation from classic AML may be difficult. Surgical management of AML is guided by suspected malignancy, symptoms, and radiologic size; nevertheless, some patients with EAML develop local recurrence or distant metastasis after resection. We therefore retrospectively compared clinical characteristics

and outcomes between EAML and classic AML in 116 consecutive patients who underwent surgery for AML.

Consistent with prior studies, middle-aged patients and women predominated in both the EAML and classic AML groups in our cohort.<sup>7,12</sup> However, Lee et al reported that males were 3.3 times more likely than females to have EAML.<sup>6</sup> Female sex hormones have been proposed to explain women’s higher AML prevalence, but whether epithelioid differentiation affects sex distribution remains uncertain.<sup>13</sup> In our cohort, tumor laterality, bilaterality, tuberous sclerosis complex, prior rupture, and comorbidities

**TABLE 3.** Clinicopathologic features and malignancy-risk stratification of EAML.

Clinicopathologic features	EAML <i>n</i> = 15
Adverse features	12 (80)
LVI	6 (40)
TSC	1 (6.67)
Tumor size > 7 cm	9 (60)
Concurrent classic AML	7 (46.7)
Carcinoma-like morphological pattern	9 (60)
Involvement of perinephric fat/renal vein	9 (60)
Necrosis	9 (60)
Risk	
Low	4 (26.7)
Intermediate	4 (26.7)
High	7 (46.7)
T staging	
T1	5 (33.3)
T2	4 (26.7)
T3a	5 (33.3)
T4	1 (6.7)
Positive margin	4 (26.67)
Recurrence*	2 (13.3)
Metastasis*	3 (20)
Dead	3 (20)

Risk categories follow Nese et al<sup>9</sup>: low risk 0–1 parameter, intermediate risk 2–3, and high risk 4–5, as defined in **Methods**. Pathologic T stage is shown as recorded. One case experienced both local recurrence and metastasis, as indicated by the asterisk in the table.

**Abbreviations:** AML, angiomyolipoma; EAML, epithelioid angiomyolipoma; LVI, lymphovascular invasion; TSC, tuberous sclerosis complex

did not differ between the groups. In a series of 587 patients, Lee et al found no difference in hemorrhage incidence between sporadic AML and EAML.<sup>12</sup> Aydin et al reported that EAML is strongly associated with tuberous sclerosis complex and adverse pathologic features.<sup>14</sup> Correspondingly, 1 EAML patient with the complex in our series exhibited malignant behavior. Our findings add to the clinical characterization of AML; however, the biological distinctions between classic AML and EAML require further investigation.

Patients with classic AML had a higher rate of asymptomatic presentation than those with EAML (59.4% vs 26.7%). This finding aligns with a large surgical series that reported 48% of triphasic AMLs were incidental findings in 209 patients.<sup>2</sup> The proportion of asymptomatic patients may be even higher under active surveillance; a Swedish cohort reported 61% asymptomatic at presentation.<sup>15</sup> These data suggest that asymptomatic presentation reflects the benign behavior of classic AML. Conversely, our patients with EAML more often presented with symptomatic

**TABLE 4.** Clinicopathologic features and surgical outcomes of malignant EAML cases.

Case	Age/ Sex	Presenting symptoms	Surgery type	Tumor size (cm)	Tumor margin (R)/Node status (N)	LVI	Presence of TSC	Concurrent classic AML	Other adverse features*	Risk	Local recurrence/ distant metastasis	Adjuvant treatments	Time to disease progression (mo)	Follow -up time (mo)	Survival status
1	31/M	Hematuria	RN	14.5	R0/N0	N	Y	Y	Y	High	Local/lung	N	29	36	Dead (progression of disease)
2	24/F	Hematuria	RN	14	R0/N0	Y	N	N	Y	High	Bone, lung, liver, brain	Adjuvant RT at bone	36	39	Dead (progression of disease)
3	40/F	Palpable mass	RN	26	R+/N+	Y	N	Y	Y	High	Liver	N	3	3	Dead (progression of disease)
4	66/F	Abdominal pain	PN	10	R+/N0	Y	N	Y	Y	High	Local	Adjuvant RT at tumor base	9	12	Alive

R denotes margin status (R0, negative; R+, positive). N denotes nodal status (N0, negative; N+, positive). "Other adverse features" include carcinoma-like growth, perinephric fat or renal vein involvement, necrosis, and tumor size  $\geq 7$  cm. Time intervals are in months.

**Abbreviations:** AML, angiomyolipoma; EAML, epithelioid angiomyolipoma; LVI, lymphovascular invasion; mo, months; N0/N+, nodal status; PN, partial nephrectomy; R0/R+, resection margin status; RN, radical nephrectomy; RT, radiotherapy; TSC, tuberous sclerosis complex.

disease, notably palpable masses and gross hematuria, which likely relates to larger tumors in EAML than classic AML (10 vs 5.6 cm). This pattern may also indicate more invasive EAML biology.

Imaging features differed significantly between the 2 groups. Fat-containing masses consistent with AML were reported more often in patients with classic AML than in those with EAML (66.3% vs 13.3%;  $P < 0.001$ ). Because EAML exhibits variable proportions of adipose tissue, smooth muscle, and blood vessels, its radiologic appearance is heterogeneous and overlaps with other renal tumors.<sup>6,16</sup> Accordingly, radiologists frequently suggested RCC (73.3%) or reported an indeterminate diagnosis (13.3%) in our EAML cases. Thus, radiologic diagnosis of EAML remains challenging, though future work may identify imaging indicators that raise preoperative suspicion.

Indications for surgery differed significantly between groups. In classic AML, the main drivers were risk of bleeding or increasing tumor size (34.7%) and symptoms including prior rupture (27.7%). Nevertheless, one third of classic AML cases underwent surgery due to suspected malignancy. In EAML, suspected malignancy prompted surgery in 86.7% of patients.

The operative approach reflected these indications. Radical nephrectomy was performed significantly more often in cases of EAML, whereas nephron-sparing surgery was favored in classic AML. This pattern suggests potential overtreatment when managing presumed cancer, particularly the use of radical nephrectomy for lesions that are ultimately proven to be AML. In a surgical series of 219 patients, Lane et al reported that 55% had radiographic features prompting surgery for suspected RCC.<sup>2</sup> Given that EAML cannot be reliably diagnosed without immunohistochemistry, percutaneous renal mass biopsy may be necessary for treatment planning in selected indeterminate cases.

In multivariable analysis, larger tumor size ( $\geq 10$  cm) and a radiologic impression of RCC independently predicted EAML. Tumor size  $\geq 10$  cm conferred higher odds of EAML (odds ratio 15.4, 95% CI 2.59–91.98;  $P = 0.003$ ). A radiographic report suggestive of malignancy increased the odds of EAML diagnosis by approximately 47-fold. In a retrospective comparison of classic AML ( $n = 204$ ) and EAML ( $n = 27$ ), younger age, male sex, and larger tumors predicted EAML.<sup>6</sup> That study also reported that 70.4% of EAML cases were interpreted as RCC on computed tomography.<sup>6</sup> Another series of 9 EAML suggested that a lipid-poor mass without calcification should raise clinical suspicion.<sup>17</sup> Thus, large tumor size and radiologic concern for malignancy should

prompt consideration of EAML; however, preoperative tissue diagnosis remains essential when distinguishing malignancy from EAML is required. Higher body mass index appeared protective against EAML in univariable analysis ( $P = 0.024$ ) but lost significance in multivariable models ( $P = 0.069$ ). Obesity has been associated with higher renal neoplasm risk in European cohorts<sup>18,19</sup>, yet its relationship with angiomyolipoma is unclear. Notably, our cohort represents a Southeast Asian population with generally lower body mass index; further studies are needed to clarify this association.

Surgical excision remains the primary treatment for renal EAML. In 2008, Aydin et al reported a benign course in all EAML cases ( $n = 15$ ) over a median follow-up of 5.1 years.<sup>14</sup> In our cohort, surgical outcomes differed between classic AML and EAML. Classic AML showed no progression with 3-year overall survival of 100%, whereas overall survival was 76.9% for EAML, consistent with multicenter series.<sup>7,20</sup>

To understand these poorer outcomes, previous studies identified adverse clinicopathologic parameters. Brimo et al analyzed 40 EAML and proposed 4 features:  $\geq 70\%$  atypical epithelioid cells;  $\geq 2$  mitotic figures per 10 high-power fields; atypical mitotic figures; and necrosis. The presence of 3 or 4 features strongly predicted malignant behavior.<sup>21</sup> Similarly, Nese et al proposed 5 parameters for risk stratification: tuberous sclerosis complex or concurrent classic AML, necrosis, tumor size  $\geq 7$  cm, extrarenal invasion and/or renal vein involvement, and carcinoma-like growth pattern. Low-, intermediate-, and high-risk categories corresponded to 0–1, 2–3, and 4–5 parameters, respectively.<sup>9</sup> In 2015, Lei et al reported that tumor size  $\geq 9$  cm, venous tumor thrombus, epithelioid cells  $\geq 70\%$  or atypical cells  $\geq 60\%$ , and necrosis were associated with progression.<sup>20</sup> In our 15 EAML cases, 12 patients (80%) had at least 1 adverse parameter under the Nese scheme.<sup>9</sup> All low-risk patients remained recurrence-free at last follow-up. Seven of 15 patients (46.7%) were high risk, and 4 of these (57.1%) progressed. Prognostic factor analysis for malignant behavior was not feasible given the small number of events. Due to the rarity of EAML, no consensus criteria for predicting malignancy have been established.

Because evidence remains limited, treatment guidelines for EAML are not well established. Complete surgical resection remains the preferred option; however, some patients experience tumor progression after surgery. Reported recurrence and metastasis rates reach 17% and 49%, respectively, and mortality can reach 33%.<sup>9</sup> In our series, 4 patients (26.7%) developed progression and were therefore classified as having malignant EAML; all

had adverse parameters per Nese et al.<sup>9</sup> One patient with positive margins and nodal involvement developed liver metastasis 3 months postoperatively. Another patient with positive margins but no nodal involvement had local recurrence at 9 months.

In contrast, 2 patients with negative surgical margins had longer progression-free survival at 29 and 36 months. These findings underscore that complete tumor removal remains crucial for survival. In our cohort, none of the 3 patients with metastatic EAML received systemic therapy because of poor performance status, which led to rapid progression and tumor-related death; overall mortality was 20%. In 2023, Zhang et al summarized 46 cases of metastatic EAML, reporting poor prognosis and low 5-year survival.<sup>10</sup> Recent reports suggest benefit from mechanistic target of rapamycin inhibitors, tyrosine kinase inhibitors, and immunotherapy in malignant EAML.<sup>5</sup> Based on our data, patients with high-risk adverse parameters warrant close monitoring and may benefit from intensive adjuvant management to improve outcomes.

Our study has limitations. First, its retrospective design restricted analyses to available clinical data. Second, excluding patients managed with angioembolization or active surveillance may have underestimated the true EAML population.<sup>22</sup> Third, radiologic metrics of intralesional fat (Hounsfield units, chemical shift imaging) were unavailable due to incomplete data. Fourth, the number of EAML cases was small because this was a single-center study. Larger multicenter cohorts with standardized radiologic characterization and comprehensive data collection are needed for more robust conclusions.

## CONCLUSIONS

EAML presents distinct clinical challenges and is associated with less favorable surgical outcomes than classic AML. Larger tumor size and a preoperative radiologic impression of malignancy were significantly associated with an EAML diagnosis. Adverse pathologic features in EAML indicate malignant potential. Active postoperative management and long-term surveillance are therefore essential to optimize patient outcomes.

## DECLARATIONS

### Grants and Funding Information

None.

### Conflict of Interests

The authors declare no conflict of interest.

### Registration number of clinical trial

None.

### Author Contributions

Conceptualization and methodology: N.W., T.T., T.H.; Investigation: N.W., V.W.; Formal analysis: N.W., P.L., K.J.; Visualization and writing—original draft: N.W., T.T.; Writing-review and editing: T.T., E.C., S.J., N.T.; Supervision: T.T.

### Ethics Statement

The study protocol was approved by the Siriraj Institutional Review Board (SIRB) (COA no. Si 605/2025).

### Use of Artificial Intelligence

No artificial intelligence tools or technologies were used in the writing and analysis.

## REFERENCES

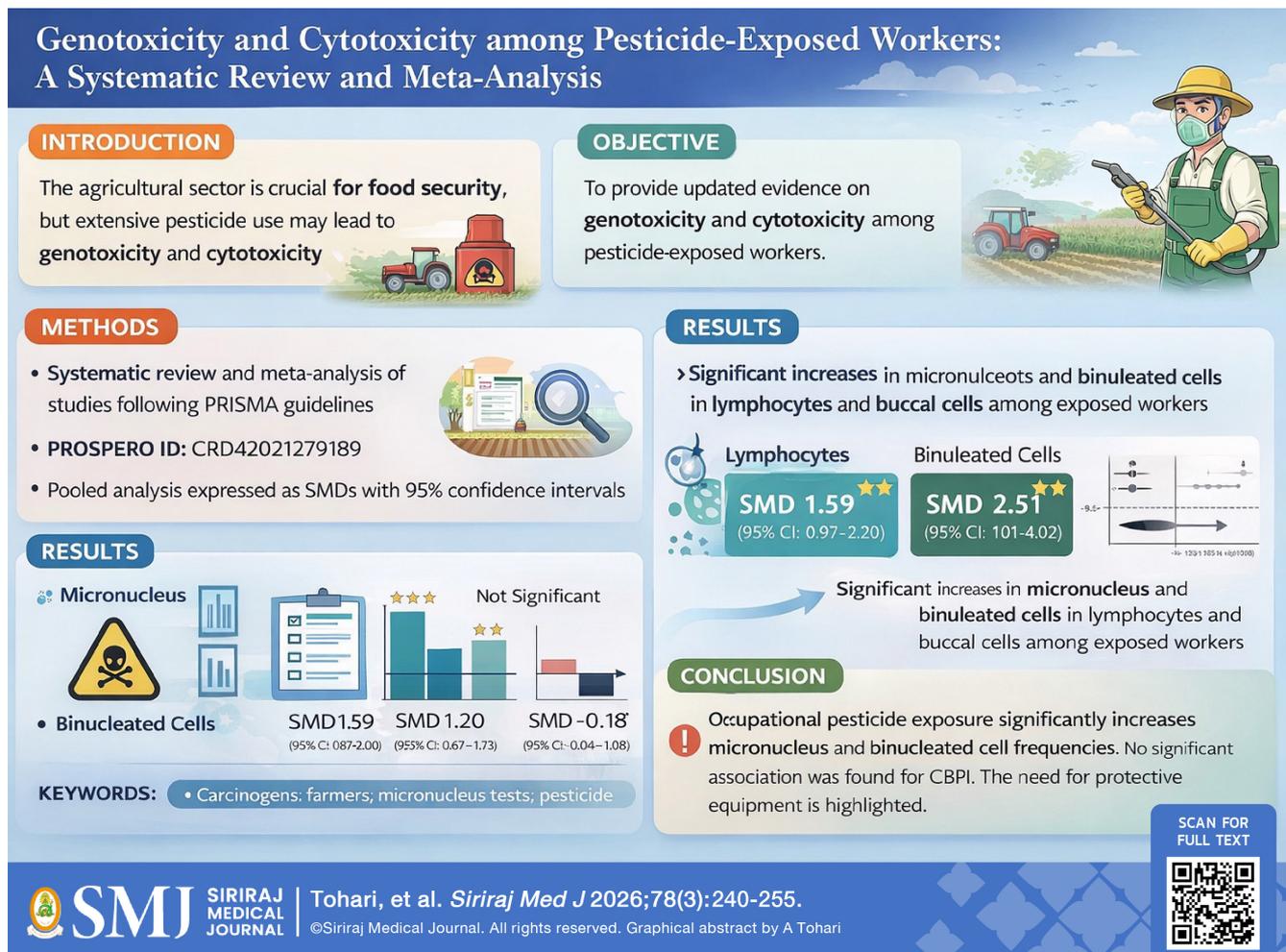
1. Lopez-Beltran A, Scarpelli M, Montironi R, Kirkali Z. 2004 WHO classification of the renal tumors of the adults. *European Urology* [Internet]. 2006 Jan 18;49(5):798–805. Available from: <https://pubmed.ncbi.nlm.nih.gov/16442207/>
2. Lane BR, Aydin H, Danforth TL, Zhou M, Remer EM, Novick AC, et al. Clinical correlates of renal angiomyolipoma subtypes in 209 patients: classic, fat poor, tuberous sclerosis associated and epithelioid. *The Journal of Urology* [Internet]. 2008 Jul 17;180(3):836–43. Available from: <https://doi.org/10.1016/j.juro.2008.05.04>
3. Vos N, Oyen R. Renal angiomyolipoma: The good, the bad, and the ugly. *Journal of the Belgian Society of Radiology* [Internet]. 2018 Jan 1;102(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/30039053/>
4. Chuang CK, Lin HCA, Tasi HY, Lee KH, Kao Y, Chuang FL, et al. Clinical presentations and molecular studies of invasive renal epithelioid angiomyolipoma. *International Urology and Nephrology* [Internet]. 2017 May 25;49(9):1527–36. Available from: <https://pubmed.ncbi.nlm.nih.gov/28547571/>
5. Yang J, Liang C, Yang L. Advancements in the diagnosis and treatment of renal epithelioid angiomyolipoma: A narrative review. *The Kaohsiung Journal of Medical Sciences* [Internet]. 2022 Sep 3;38(10):925–32. Available from: <https://pubmed.ncbi.nlm.nih.gov/36056704/>
6. Lee W, Choi SY, Lee C, Yoo S, You D, Jeong IG, et al. Does epithelioid angiomyolipoma have poorer prognosis, compared with classic angiomyolipoma? *Investigative and Clinical Urology* [Internet]. 2018 Jan 1;59(6):357. Available from: <https://pubmed.ncbi.nlm.nih.gov/30402567/>
7. Delhorme J b., Fontana A, Levy A, Terrier P, Fiore M, Tzanis D, et al. Renal angiomyolipomas: At least two diseases. A series of patients treated at two European institutions. *European Journal of Surgical Oncology* [Internet]. 2016 Dec 5;43(4):831–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/28007324/>
8. Tsai H, Lee K, Ng K, Kao Y, Chuang C. Clinicopathologic analysis of renal epithelioid angiomyolipoma: Consecutively excised 23 cases. *The Kaohsiung Journal of Medical Sciences*

- [Internet]. 2019 Jan 1;35(1):33–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/30844148/>
9. Nese N, Martignoni G, Fletcher CD, Gupta R, Pan CC, Kim H, et al. Pure epithelioid PECOMas (So-Called epithelioid angiomyolipoma) of the kidney. *The American Journal of Surgical Pathology* [Internet]. 2011 Jan 22;35(2):161–76. Available from: <https://pubmed.ncbi.nlm.nih.gov/21263237/>
  10. Zhang J, Wang WJ, Chen LH, Wang N, Wang MW, Liu H, et al. Primary renal malignant epithelioid angiomyolipoma with distant metastasis: a case report and literature review. *Frontiers in Oncology* [Internet]. 2023 Aug 22;13. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10477911/>
  11. Mai KT, Perkins DG, Collins JP. Epithelioid cell variant of renal angiomyolipoma. *Histopathology* [Internet]. 1996 Mar 1;28(3):277–80. Available from: <https://pubmed.ncbi.nlm.nih.gov/8729052/>
  12. Lee KH, Tsai HY, Kao YT, Lin HC, Chou YC, Su SH, et al. Clinical behavior and management of three types of renal angiomyolipomas. *Journal of the Formosan Medical Association* [Internet]. 2018 Mar 15;118(1):162–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/29549981/>
  13. Boorjian SA, Sheinin Y, Crispen PL, Lohse CM, Kwon ED, Leibovich BC. Hormone receptor expression in renal angiomyolipoma: Clinicopathologic correlation. *Urology* [Internet]. 2008 Mar 29;72(4):927–32. Available from: <https://doi.org/10.1016/j.urology.2008.01.067>
  14. Aydin H, Magi-Galluzzi C, Lane BR, Sercia L, Lopez JI, Rini BI, Zhou M. Renal angiomyolipoma: clinicopathologic study of 194 cases with emphasis on the epithelioid histology and tuberous sclerosis association. *Am J Surg Pathol*. 2009 Feb;33(2):289–97.
  15. Swärd J, Henrikson O, Lyrdal D, Peeker R, Lundstam S. Renal angiomyolipoma-patient characteristics and treatment with focus on active surveillance. *Scandinavian Journal of Urology* [Internet]. 2020 Jan 23;54(2):141–6. Available from: <https://doi.org/10.1080/21681805.2020.1716066>
  16. Thiravit S, Teerasamit W, Thiravit P. The different faces of renal angiomyolipomas on radiologic imaging: a pictorial review. *British Journal of Radiology* [Internet]. 2018 Jan 12;91(1084). Available from: <https://doi.org/10.1259/bjr.20170533>
  17. Froemming AT, Boland J, Chevillet J, Takahashi N, Kawashima A. Renal Epithelioid Angiomyolipoma: imaging characteristics in nine cases with Radiologic-Pathologic correlation and review of the literature. *American Journal of Roentgenology* [Internet]. 2013 Jan 23;200(2):W178–86. Available from: <https://doi.org/10.2214/ajr.12.8776>
  18. Overweight as an avoidable cause of cancer in Europe. PubMed [Internet]. 2001 Feb 1; Available from: [https://doi.org/10.1002/1097-0215\(200002\)9999:9999](https://doi.org/10.1002/1097-0215(200002)9999:9999)
  19. Wolk A, Gridley G, Svensson M, Nyren O, McLaughlin JK, Fraumeni JF, et al. A prospective study of obesity and cancer risk (Sweden). *Cancer Causes & Control* [Internet]. 2001 Jan 1;12(1):13–21. Available from: <https://doi.org/10.1023/a:1008995217664>
  20. Lei JH, Liu LR, Wei Q, Song TR, Yang L, Yuan HC, et al. A Four-Year Follow-up Study of Renal Epithelioid Angiomyolipoma: A Multi-Center Experience and Literature review. *Scientific Reports* [Internet]. 2015 May 5;5(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/25939249/>
  21. Brimo F, Robinson B, Guo C, Zhou M, Latour M, Epstein JI. Renal Epithelioid angiomyolipoma with atypia: A series of 40 cases with emphasis on clinicopathologic prognostic indicators of malignancy. *The American Journal of Surgical Pathology* [Internet]. 2010 Apr 20;34(5):715–22. Available from: <https://doi.org/10.1097/pas.0b013e3181d90370>
  22. Chaiyasoot W, Yodying J, Limsiri T. Selective Arterial Embolization of Renal Angiomyolipoma: Efficacy, Tumor Volume Reduction and Complications. *Siriraj Med J* [internet]. 2021 Mar. 31 [cited 2026 Feb. 5];73(5):337–43. available from: <https://he02.tci-thaijo.org/index.php/sirirajmedj/article/view/250279>

# Genotoxicity and Cytotoxicity among Pesticide-Exposed Workers: A Systematic Review and Meta-Analysis

Achmad Ilham Tohari, M.D.<sup>1</sup>, Muhammad Rijal Fahrudin Hidayat, M.D.<sup>1</sup>, Nabil Athoillah, M.D.<sup>1</sup>, Muhammad Yuda Nugraha, M.D.<sup>1</sup>, Elly Nurus Sakinah, M.D. Ph.D.<sup>2</sup>, Supangat, M.D., Ph.D.<sup>2,\*</sup>, Saekhol Bakri, M.D., Ph.D.<sup>3</sup>, Athira Nandakumar, Ph.D.<sup>4</sup>

<sup>1</sup>PANAH Research Center, Faculty of Medicine, University of Jember, Jember, Indonesia, <sup>2</sup>Department of Pharmacology, Faculty of Medicine, University of Jember, Jember, Indonesia, <sup>3</sup>Department of Public Health, Faculty of Medicine, Diponegoro University, Semarang, Indonesia, <sup>4</sup>Department of Epidemiology and Preventive Medicine, Kagoshima University, Kagoshima, Japan.



\*Corresponding author: Supangat

E-mail: drsupangat@unej.ac.id

Received 30 November 2025 Revised 2 February 2026 Accepted 3 February 2026

ORCID ID: <http://orcid.org/0000-0003-4915-9218>

<https://doi.org/10.33192/smj.v78i3.279093>



**ABSTRACT**

**Objective:** To provide updated evidence on genotoxicity and cytotoxicity among workers occupationally exposed to pesticides.

**Materials and Methods:** This systematic review and meta-analysis followed PRISMA guidelines. Studies assessing micronuclei and cytotoxicity biomarkers in occupationally pesticide-exposed workers were included. Pooled analyses used Mantel-Haenszel fixed- and random-effects models, and results were expressed as SMDs with 95% confidence intervals (CIs). The protocol was registered in PROSPERO (CRD42021279189).

**Results:** Micronucleus frequencies were significantly higher in lymphocytes (SMD 1.59; 95% CI 0.97–2.20;  $p < 0.001$ ;  $I^2 = 96\%$ ) and buccal cells (SMD 1.20; 95% CI 0.67–1.73;  $p < 0.00001$ ;  $I^2 = 97\%$ ) among exposed workers. Binucleated cells were also increased in lymphocytes (SMD 2.51; 95% CI 1.01–4.02;  $p < 0.001$ ;  $I^2 = 98\%$ ) and buccal cells (SMD 0.56; 95% CI 0.04–1.08;  $p = 0.03$ ;  $I^2 = 96\%$ ). No significant difference was observed for CBPI (SMD  $-0.18$ ; 95% CI  $-0.90$ – $0.54$ ;  $p = 0.63$ ;  $I^2 = 96\%$ ).

**Conclusion:** Occupational pesticide exposure is associated with increased micronucleus and binucleated cell frequencies, although high heterogeneity and potential confounding factors limit certainty. No significant association was found for CBPI. Subgroup analyses showed no sex-related differences, while concurrent smoking appeared to amplify genotoxic markers. The available evidence supports a protective effect of appropriate personal protective equipment against pesticide-induced genotoxicity.

**Keywords:** Carcinogens; farmers; micronucleus tests; pesticide (Siriraj Med J 2026;78(3):240-255)

**INTRODUCTION**

Food insecurity and undernutrition continue to be major issues in many nations even after numerous steps have been taken to reduce the global hunger crisis.<sup>1</sup> As such, the agricultural sector is crucial to increase food availability and achieving food security.<sup>2</sup> Additionally, as a result of increasing food demand and decreasing crop production, pesticides are often used in agricultural fields to eradicate various insects, pests, weeds, and other unwanted vegetation in exchange for high productivity.<sup>3</sup> There are many different types of pesticides, such as nematicides, herbicides, insecticides, fungicides, and rodenticides. Around two million pesticides are used in the world, of which 47.5%, 29.5%, 17.5%, and 5.5% are related to herbicides, insecticides, fungicides, and other pesticides, respectively.<sup>4</sup>

Agricultural workers are the population that is continuously exposed to many kinds of pesticides. There are four main routes of pesticide exposure, including inhalation, dermal contact, food ingestion, and an environment that is contaminated with pesticides.<sup>5</sup> Pesticides may accumulate within the body and cause many health risks, such as acute poisoning and chronic effects that may lead to genotoxicity, which is the precursor of carcinogenesis.<sup>6</sup> The AGRICOH consortium published two studies that examined the relationship between pesticide exposure and cancer incidence.<sup>7,8</sup> Compared to the general population, a study by Togawa reported the incidence of primary

cancer diagnosis and specific subtypes in occupational agricultural cohorts. They discovered that women were at higher risk for skin melanoma (meta-SIR = 1.18, CI: 1.01-1.38), multiple myeloma (meta-SIR = 1.27, CI: 1.04-1.54), and prostate cancer (meta-SIR = 1.06, CI: 1.01-1.12).<sup>8</sup> In Leon et al.'s study, which focused on non-Hodgkin's lymphoma, researchers found that there was a higher risk of non-Hodgkin's lymphoma among those who had ever used terbufos (meta-HR = 1.18, CI: 1.00-1.39), chronic lymphocytic leukemia/small lymphocytic lymphoma among those who had ever used deltamethrin (meta-HR = 1.48, CI: 1.06-2.07), and diffuse large B-cell lymphoma in those who had ever used glyphosate (meta-HR = 1.36, CI: 1.00-1.85).<sup>7</sup> Several carcinogenesis mechanisms have also been used to demonstrate pesticides' pathways, such as genotoxicity, cytotoxicity, disruption of hormone levels, oxidative stress, inflammation, immune system regulation, and procarcinogen activation.<sup>9-11</sup>

Amongst all the mechanisms above, genotoxicity and cytotoxicity have been proven as a method for screening for the risk of cancer and monitoring the prognosis.<sup>12</sup> In evaluating genotoxicity-cytotoxicity, biomonitoring assessment is considered the best choice because of the ability to measure chromosomal damage from the beginning phase of chemical carcinogenesis.<sup>13</sup> There are many kinds of biomonitoring assessments but this study focused on micronuclei, binucleated cells, and Cytokinesis Block

Proliferation Index (CBPI). For analyzing cytogenetic effects, the micronucleus test provides a sensitive and practical approach.<sup>14</sup> Micronuclei are small extra nuclei that originate from chromosomal fragments that are excluded from the nucleus. Furthermore, chromosomal breakage during cell division is closely linked with these markers.<sup>15</sup> An international cohort study comprising more than 6000 participants globally reported that there was a high risk of cancer in numerous organs (liver, biliary duct, pancreas, lung, stomach, colon-rectum, urogenital) among the medium/high micronuclei subjects. These subjects also had lower cancer-free patients compared to the low micronuclei subjects.<sup>16</sup>

Binucleated cells are cells that have two nuclei. These markers often arise in various malignancies such as acute myeloid leukemia, angiosarcoma, and malignant mesothelioma.<sup>17-19</sup> There are two separate theories on the formation of binucleated cells: “abnormal mitosis” and “cell-cell fusion”.<sup>20</sup> The CBPI demonstrates human lymphocytes’ ability to proliferate in the examined experimental settings as well as the capacity of cells to self-repair.<sup>21</sup> Micronucleus and binucleated cells biomonitoring assessment can be obtained through peripheral blood lymphocytes (using the Cytokinesis Block Micronuclei Assay (CBMN)) and buccal cells (using Buccal Micronucleus Cytome Assay (BMCyt)).<sup>22,23</sup> Meanwhile, CBPI can only be obtained using the CBMN assay.<sup>21</sup>

Pesticides induce MN, binucleated cells, and reduced CBPI via clastogenic mechanisms, such as ROS-mediated DNA breaks, adduct formation, and topoisomerase inhibition, leading to acentric fragments that form MN.<sup>24,25</sup> Aneugenic effects involve tubulin binding, spindle/kinetochore disruption, and interference with kinesin-14, pericentrin, and Aurora Kinase A, causing chromosome loss into MN. Cytokinesis inhibition from ATP depletion, G2/M arrest, and actin/myosin dysregulation yields binucleated cells and lowers CBPI, signaling cytotoxicity.<sup>24,26</sup> These molecular pathways directly link exposure to cytogenetic biomarkers for robust genotoxicity assessment.

However, the published studies regarding these biomonitoring assessments and occupational pesticide exposure remain inconclusive. Studies conducted by Moshou et al. showed an increase in micronuclei both in lymphocytes and buccal cells in the pesticide exposure group compared to the control group (lymphocytes:  $13,67 \pm 6,3197$  vs  $8,88 \pm 3,4783$ ; buccal cells:  $12,46 \pm 5,1439$  vs  $8,50 \pm 3,0374$ ) while a study by Pastor et al. did not show any significant results (lymphocytes:  $12,55 \pm 8,5028$  vs  $13,82 \pm 10,1877$ ; buccal cells:  $1,03 \pm 1,359$  vs  $1,06 \pm 1,4595$ ).<sup>27,28</sup> In addition, many studies showed

significant results whereas others showed vice versa. Thus, the objective of this systematic review and meta-analysis is to provide the latest evidence of micronucleus and cytotoxicity assessment in the population who are occupationally exposed to pesticides.

## MATERIALS AND METHODS

### Study protocol and study design

This study used the Preferred Reporting Items of Systematic Reviews and Meta-Analysis (PRISMA) guidelines<sup>29</sup>, and we practiced the PECOS approach (Populations: Workers exposed to pesticide; Exposure: Any kind of pesticide; Comparison: Populations that are not exposed to pesticide; Outcome: CBPI, micronucleus, and binucleated cells in lymphocytes as well as buccal cells. The study design employed observational studies (including cohort and case-control) to gather papers from the PubMed and Scopus databases. The protocol of this study was registered in The International Prospective Register of Systematic Reviews (PROSPERO) on November 7, 2022, with the registered number CRD42022334503. Deviation from protocol was stated in the supplementary file.

### Database searching and study selection

We searched the papers using the same keywords for all databases (PubMed and Scopus): (Pesticide OR Pesticides) AND (“micronuclei” OR “micronuclei assay” OR “genotoxicity”). There was no year of publication restriction, although a language restriction to English was applied. All papers from databases were imported to Rayyan.ai to execute the study selection process and remove duplicates. Two investigators (AIT and MRFH) performed a two-step study selection; the first step was title and abstract screening, and the second was full manuscript screening. All study selection was accomplished independently. If a conflict arose between the investigators, the decision was made through discussion with the third investigator (SS). The study selection is up to date until January 2024. We did not perform a manual hand search in this study.

### Eligibility criteria

We determined the included studies based on the following inclusion criteria: (1) observational studies such as case-control, cross-sectional, or cohort studies; (2) assessed pesticide exposure and micronucleus, binucleated cells, or CBPI; (3) availability of data on central tendency (mean/median) and dispersion (standard deviation (SD) / standard error (SE) / interquartile range (IQR) / 95% confidence interval (CI)); (4) If there was more than

one paper with the same populations, the most recent papers having complete data was selected. The exclusion criteria were: (1) published studies in a language other than English; (2) using the same sample as the latest study; (3) incomplete data to measure the effect size.

### Risk of bias assessment

After the study selection, we conducted a risk of bias assessment using the Risk of Bias in Non-randomized Studies of Exposures (ROBINS-E) tool.<sup>30</sup> There are seven domains in ROBINS-E, including (1) confounding; (2) selection of participants; (3) exposure classification; (4) post-exposure intervention; (5) missing data; (6) outcome measures; and (7) selection of reported outcomes. Each domain was divided into three levels (“low risk of bias”, “medium risk of bias”, and high risk of bias”) based on the ROBINS-E protocols (supplementary file). The study overall was judged to have low risk of bias if all domains had low risk of bias. If any domain had moderate or high risk of bias, the aggregate risk was categorized as moderate or high, respectively.

### Data extraction

We extracted and combined the following information from included studies: (1) name of first author; (2) year of publication; (3) sample size; (4) ages of participants; (5) total number of each gender in participants; (6) smoking status; (7) alcohol drinking status; (8) usage of personal protective equipment (PPE); (9) duration of exposure; (10) study area; (11) type of occupational exposure; (12) types of plants planted by workers; (13) type of pesticide; (14) effects estimates (mean and standard deviations) of each outcomes. If the available data were SE, the SD data were obtained using the square root of the sample size.<sup>31</sup> If the study reported in median (min–max) or median (IQR), we extracted the information in the same way as in the study. When additional information was required, we contacted the original paper’s corresponding author.

### Statistical analysis

The data were extracted and described narratively and quantitatively. Meta-analysis was performed in Review Manager (RevMan) V5.4 (The Cochrane Community). Standardized mean differences (SMD) with 95% confidence intervals (CI) were calculated as the outcome measures. SMD can be defined as the mean of the exposure group minus the mean of the control group divided by the pooled SD of all groups. If there was a positive SMD, it indicated an increase in outcome measures due to pesticide exposure, and vice versa.<sup>31</sup> The heterogeneity test was calculated using the  $I^2$  test which reported low

heterogeneity if the  $I^2 < 50\%$  and high heterogeneity if  $I^2 \geq 50\%$ .<sup>31</sup> Random effect models were used in this meta-analysis due to high heterogeneity. The significance of the meta-analysis was set at  $p < 0.05$ . Publication bias was assessed using a funnel plot and Egger’s test if at least 10 studies were included in the analysis. The sensitivity analysis was conducted using leave-one-out analysis by excluding each study one by one to show the effects of each study in the pooled analysis.

## RESULTS

### Characteristics of the included study

We found 3592 studies from PubMed and Scopus databases, and then we performed duplicate detection and found that 992 studies were detected as duplicates. All duplicate studies were removed. A total of 2557 studies were removed due to being irrelevant studies or not meeting the inclusion criteria. Approximately 43 papers underwent eligibility assessment; some articles were removed for specific reasons, including the absence of comparison (2 studies), language other than English (6 studies), same sample (3 studies), Non-Occupational Exposure (4 studies), and incomplete data (2 studies). Finally, 26 studies were included in the systematic review and meta-analysis. The PRISMA flow chart is presented in [Fig 1](#). All included studies were published between 2000 and 2023. Most of them were single-center studies, except one study that was a multi-center study conducted in four countries in Europe.<sup>28</sup> Among the included studies, six were conducted in Brazil, six in Turkey, two in Greece, Pakistan, and Spain, and one each in Poland, Hungary, India, Argentina, Iraq and Chile. Based on the outcome, 14 studies reported MN frequency in lymphocytes, 18 studies reported MN frequency in buccal cells, six studies reported binucleated cells in lymphocytes, 11 studies reported binucleated cells in buccal cells, and eight studies that reported CBPI ([Supplementary Table 1](#)).

### Risk of bias assessment

We used the ROBINS-E tools to assess the risk of bias among included studies. The risk of bias decisions for each study and domain are shown in [Supplementary Fig 1](#). Overall, the risk of bias in included studies varied from low to high risk. There were two studies graded as high risk of bias<sup>32,33</sup>; The main reasons were because of high risk of bias in the confounding domain (due to limited confounding data reported) and high risk of bias in outcome measure (due to inconsistent outcome reported); further reasons for the high risk grading can be found in [the supplementary file](#). There was one study that had an overall low risk of bias.<sup>34</sup> Other studies included

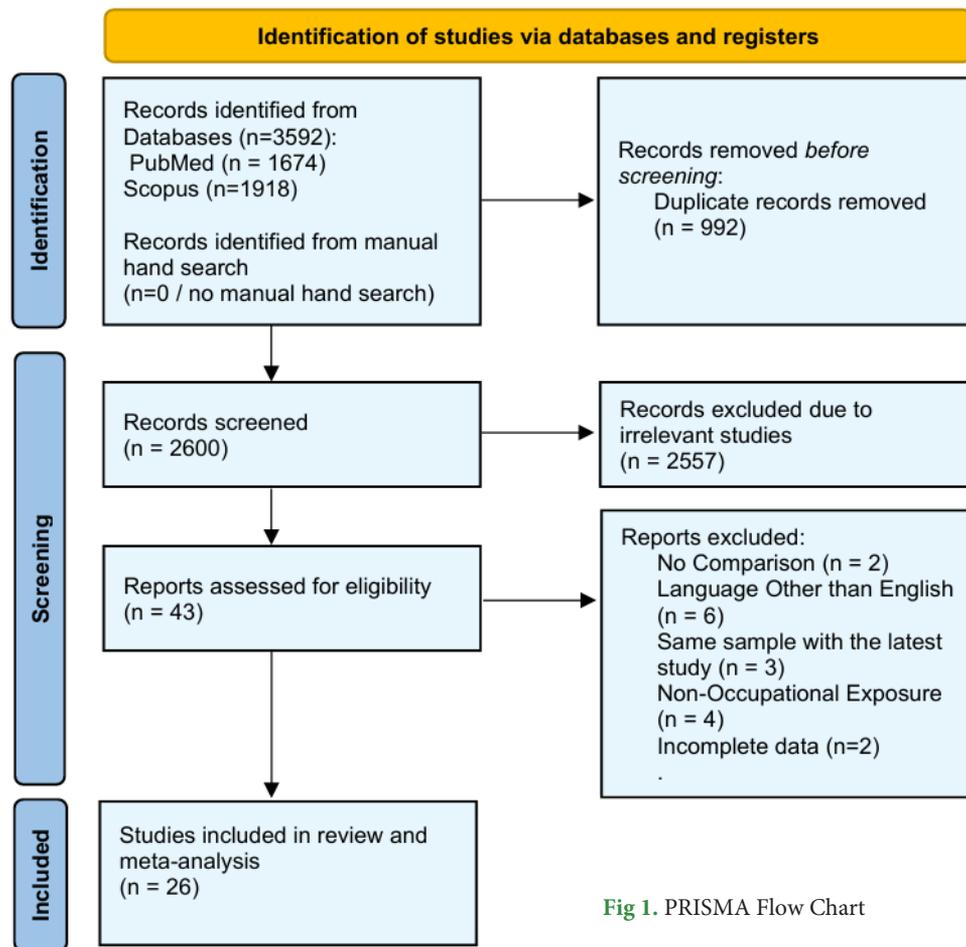


Fig 1. PRISMA Flow Chart

in the analysis were scored as need concern due to at least one domain in ROBINS-E filled as “need concern” in confounding, measurement, selecting participants, or reporting data.<sup>27,28,35-55</sup>

### Characteristics of participants in included studies

Table 1 shows the characteristics of participants. The range of mean age among the participants was 32.83 - 49.15 years for the exposure group and 26.5 - 47/87 years for the control group. All studies reported the proportion of each gender in the study. However, several studies were conducted on one specific gender. Studies by Ali (2008) and Yslas (2016) were only conducted on female agricultural workers.<sup>39,47</sup> Nevertheless, as many as seven studies were conducted on male agricultural workers.<sup>27,33,35,37,40,43,46,55</sup> Almost all studies revealed the smoking habits of the participants, except for three studies.<sup>32,44,54</sup> Two studies were conducted on all non-smoker participants. Regarding alcohol intake, not all studies reported the exact data. Fifteen of 26 studies did not mention information about alcohol intake. Regarding the usage of PPE, two studies from Europe reported that 80% of workers used adequate PPE including gloves, protective clothes, masks, goggles, and boots.<sup>28,35</sup> Only four studies did not mention any information about

PPE.<sup>33,36,47,52</sup> Cumulative exposure time in each study varied from one year to more than 29 years.

### Type of pesticides used in included studies

A list of participants' jobs, types of plants, and pesticides is shown in [Supplementary Table 2](#). Most of the included studies recruited agricultural workers as the exposure group, along with office or non-agricultural workers as the control group. The control group may be from the same area or a different area (another city). Regarding the type of plants, many kinds of vegetables and fruits were identified from all studies, despite three studies from Brazil that focused on tobacco.<sup>34,44,51</sup> Regarding the type of pesticides, we grouped them as insecticides, herbicides, and fungicides. The three pesticides used most often in the included studies were from the insecticides group: organophosphate (16 studies), pyrethroid (12 studies), and carbamate (11 studies). In the herbicide group, Atrazine and 2,4-D (2,4-Dichlorophenoxyacetic acid) were found in three studies. Paraquat, glyphosate, triazine, and cycloxydin were only found in two studies. Any copper derivatives were the most identified fungicides (6 studies), as well as mancozeb and propineb (4 studies). The percentage of pesticides used is also described in [Supplementary Table 2](#). All of the pesticide characteristic data were compiled

**TABLE 1.** Characteristics of participants in included studies.

No	Study Country	Age		n	Control mean (SD)	Gender		Smoking Habit		Alcohol		PPE Exposure	years of exposure Exposure Mean (SD)
		n	Exposure mean (SD)			Exposure (M/F)	Control (M/F)	Exposure (SM/NS/C-Day mean (SD))	Control (SM/NS/C-Day mean (SD))	Exposure	Control		
1	Lucero, 2000 <sup>35</sup> Spain	64	32.83 (9.12)	50	38.56 (9.69)	All Male	All Male	37/ 27/ 12.54 (8.47)	36/ 14/ 14.7 (8.47)	NM	NM	80% of workers used gloves and protective clothes	9.82 (8.08) years
2	Vrhovac, 2002 <sup>36</sup> Croatia	10	39.52 (6.04)	20	31 (7.71)	7 / 3	12 / 8	6/ 4/ 17 (7)	7/ 13/ 8.43 (1.88)	NM	NM	NM	22.25 years average
3	Pastor, 2003 <sup>28</sup> Greece, Spain, Poland, Hungary	244	39.34 (9.84)	231	41.91 (9.73)	198 / 46	194 / 37	116/ 131/ 18.6 (9.66)	105/ 126/ 17.58 (8.23)	95.24 (24.22) g/week mean (SD)	98.62 (137.40) g/week mean (SD)	80% of workers used PPE (gloves, mask. Boots, googles)	13.92 (9.06) years
4	Bhalli, 2006 <sup>37</sup> Pakistan	29	34.17 (2.96)	35	35.20 (3.52)	All Male	All Male	18/ 11/ NM	17/ 18/ NM	NM	NM	All workers did not use appropriate PPE	13.48 (3.84) years
5	Sailaja, 2006 <sup>38</sup> India	54	35.1 (5.20)	54	33.4 (5.50)	42 / 12	43 / 11	33/ 21/ NM	31/ 23/ NM	30 drinkers/ 24 non-drinkers	33 drinkers/ 21 non-drinkers	Lack of PPE used	8.60 (2.50) years
6	Ergene, 2007 <sup>33</sup> Turkey	32	36.21 (6.90)	32	34.56 (5.92)	All Male	All Male	16/ 16/ 20.93 (6.38)	16/ 16/ 16.31 (4.94)	NM	NM	NM	34.56 (10.47) years
7	Ali, 2008 <sup>39</sup> Pakistan	69	37.55 (12.75)	69	37.52 (13.47)	All Female	All Female	All Non-Smokers	All Non-Smokers	NM	NM	No PPE	10.26 (6.14) years
8	Bortoli, 2009 <sup>40</sup> Brazil	29	38.8 (12.2)	37	36.90 (11.10)	All Male	All Male	12/ 17/ 21 (12.40)	17/ 20/ 22.9 (11.50)	21 drinkers/ 8 non-drinkers	25 drinkers/ 12 non-drinkers	9 used / 20 did not use	16.30 (10) years
9	Martinez - valenzuela, 2009 <sup>41</sup> Mexico	70	37.18 (14.07)	70	38.75 (14.27)	45 / 25	49 / 21	42/ 28/ NM	49/ 21/ NM	32 drinkers/ 38 non-drinkers	19 drinkers/ 51 non-drinkers	No PPE	7 years average
10	Coskun, 2011 <sup>42</sup> Turkey	46	42.5 (10.1735)	48	26.50 (7.90)	34 / 12	26 / 22	31/ 15/ NM	10/ 38/ NM	NM	NM	17% of workers used mask and glove	NM
11	Gentile, 2012 <sup>43</sup> Argentina	20	36.25 (12.25)	10	37 (12.25)	All Male	All Male	5/ 15/ 13 (4.47)	0/ 10/ NA	NM	NM	Workers used several PPE: mask, gloves, and googles (25%); gloves and mask (25%); gloves only (25%)	9.93 (11.64) years
12	Da Silva, 2014 <sup>44</sup> Brazil	30	42.1 (10.15)	30	40.17 (13.02)	17 / 13	15 / 15	NM	NM	NM	NM	33% of workers used complete PPE (gloves, mask, glasses, boots). 67 workers used incomplete PPE	29.23 (12.83) years
13	Tumer, 2015 <sup>45</sup> Turkey	58	51.5 (39.8 - 58)a	58	34.5 (22 - 44.30)a	51 / 7	42 / 16	29/ 29/ NM	23/ 35/ NM	NM	NM	13 (22.4%) of workers used PPE	17.19 (8.26) years
14	Martinez- Valenzuela,2016 <sup>46</sup> Mexico	30	37.7 (16.98)	30	35.9 (14.24)	All Male	All Male	10/ 20/ NM	NM	17 alcohol. 13 Non-alcohol drinkers	NM	All pilots used gloves and mask	2 - 6 years

**TABLE 1.** Characteristics of participants in included studies. (Continue)

No	Study Country	Age		Gender		Smoking Habit		Alcohol		PPE		years of exposure Exposure Mean (SD)	
		n	Exposure mean (SD)	n	Control mean (SD)	Exposure (M/F)	Control (M/F)	Exposure (SM/NS/C-Day mean (SD))	Control (SM/NS/C-Day mean (SD))	Exposure	Control		Exposure
15	Yslas, 2016 <sup>47</sup> Mexico	37	35.5 (12.40)	34	31.9 (10.60)	All Female	All Female	2/ 35/ NM	0/ 34/ NM	2 drinkers. 35 non drinkers	3 drinkers. 31 non drinkers	NM	7.70 (8.70) years
16	Cayir, 2017 <sup>48</sup> Turkey	68	38.01 (10.01)	43	33.23 (7.39)	39 / 29	26 / 17	37/ 17/ 21.45 (10.28)	26/ 17/ 23.69 (5.74)	1.96 (5.86) L/Months mean (SD)	4.69 (23.7) L/Months mean (SD)	10% of workers used gloves. 6% of workers used mask	5 - 15 years
17	Kahl, 2018 <sup>34</sup> Brazil	40	45.0 (11.38)	40	45.60 (10.75)	19 / 21	19 / 21	All Non-Smokers	All Non-Smokers	NM	NM	32.50% of workers did not use PPE. 25% used gloves. 10% used complete PPE (mask, gloves, boots, hat, googles)	28.30 (13.28) years
18	Cobanoglu, 2019 <sup>49</sup> Turkey	66	38.7 (11.20)	50	36.3 (9.15)	38 / 28	25 / 25	36/ 30/ 21.9 (10.10)	14/ 36/ 10.3 (5.94)	1.95 (5.94) L/Month mean (SD)	0.80 (2.60) L/Month mean (SD)	Only few workers used masks and gloves	NM
19	Moshou, 2020 <sup>27</sup> Greece	24	40.2 (2.3)	24	33.70 (2.30)	All Male	All Male	12/ 12/ 32.08 (5.41)	12/ 12/ 22.08 (3.23)	NM	NM	12.50% of workers used masks. 8.30% of workers used masks and gloves	19.79 (2.46) years
20	Quintana, 2021 <sup>50</sup> Mexico	54	36.09 (1.60)	26	37.62 (3.10)	52 / 2	8 / 2	17/ 37/ NM	3/ 23/ NM	30 alcohol drinkers. 24 non- alcohol drinkers	2 alcohol drinkers. 24 non- alcohol drinkers	87% of workers used PPE (mask and gloves)	5.36 (0.43) years
21	Landeros, 2022 <sup>52</sup> Chile	30	39.8 (12.7)	30	39.60 (12)	12 / 18	10 / 20	17/ 13/ NM	8/ 22/ NM	NM	NM	NM	9.70 (9.10) years
22	Dalberto, 2022 <sup>51</sup> Brazil	84	39.58 (13.64)	85	39.85 (13.74)	41 / 43	44 / 41	All Non-Smokers	All Non-Smokers	NM	NM	58.13% of workers wear PPE (gloves, mask, clothes)	25.07 (15.72) years
23	Santos, 2022 <sup>53</sup> Brazil	81	49.16 (10.06)	81	47.87 (10.66)	69 / 12	62 / 15	33/ 47/ NM	21/ 56/ NM	49 alcohol drinkers. 32 non- alcohol drinkers	51 alcohol drinkers. 26 non- alcohol drinker	39.70% of workers wear PPE	30 (14) years
24	Alarcon, 2023 <sup>54</sup> Mexico	42	38.8 (13.6)	46	32.10 (13.20)	37 / 5	23 / 23	NM	NM	NM	NM	90% of workers do not use any PPE	15.40 (10.40) years
25	Lucio, 2023 <sup>55</sup> Brazil	23	46.96 (16.70)	27	40.55 (11.26)	All Male	All Male	3/ 20/ NM	0/ 27/ NM	14 alcohol drinkers. 9 non- alcohol drinkers	20 alcohol drinkers. 7 non- alcohol drinkers	43.50% wear mask; 34.80% wear gloves; 34.8% wear boots; 26.1% wear googles; 4.3% wear apron	35.35 (19.59) years
26	Alhamadany, 2023 <sup>32</sup> Iraq	100	15 – 57b	103	15 – 57b	NM	NM	NM	NM	NM	NM	Few workers wear PPE	NM

**Abbreviations:** NM = Not Mentioned; PPE: Personal Protective Equipment; M = Male; F = Female; SM = Smokers; NS = Non-Smokers; C-day = Cigarette per day; a= data presented in median(IQR); b=data presented in range.

using a questionnaire. In the occupational condition, the exposure was usually mixed with all types of pesticides.

### Meta-analysis of Micronucleus, Binucleated Cells, and CBPI

The forest plot of micronucleus frequency, binucleated cells, and CBPI is shown in Fig 2, Fig 3, and Fig 4, respectively. Subgroup analysis based on the region of study was performed. We applied a random effects model for the meta-analysis due to high heterogeneity ( $I^2 > 50\%$ ) between the studies. The meta-analysis of micronucleus frequency in lymphocytes and buccal cells is shown in Fig 2. Regarding micronuclei in lymphocytes, a total of 781 participants were included in the exposure group and 701 in the control group. The pooled SMD showed significant increase of micronuclei in lymphocytes within the exposure group compared to the control group [SMD = 1.59, 95% CI (0.97, 2.20),  $p < 0.00001$ ,  $I^2 = 96\%$ ]. The subgroup analysis revealed that the results were identical in all regions (Latin and South America, Asia,

and Europe). Meta-analysis of micronuclei in buccal cells included 1,127 participants in the exposure group and 1,047 in the control group. The pooled SMD showed a significant increase in micronuclei in buccal cells within the exposure group compared with the control group [SMD = 1.20, 95% CI (0.67, 1.73),  $p < 0.00001$ ,  $I^2 = 97\%$ ]. However, the subgroup analysis in the European region revealed that although the increase in micronuclei was noted, the results were not significant.

Fig 3 reports the meta-analysis of binucleated cells in lymphocytes and buccal cells. A total of 883 participants (461 in the exposure group, and 472 in the control group) were included in the binucleated lymphocytes analysis. Of these, 1,579 participants (815 in the exposure group, 764 in the control group) were included in the binucleated buccal cells analysis. Both of the pooled SMD in lymphocytes and buccal cells showed significant increase of binucleated cells in the exposure group [SMD = 2.51, 95% CI (1.01, 4.02),  $p < 0.001$ ,  $I^2 = 98\%$ ] and [SMD = 0.56, 95% CI (0.04, 1.08),  $p < 0.03$ ,  $I^2 = 96\%$ ], respectively. Nevertheless,

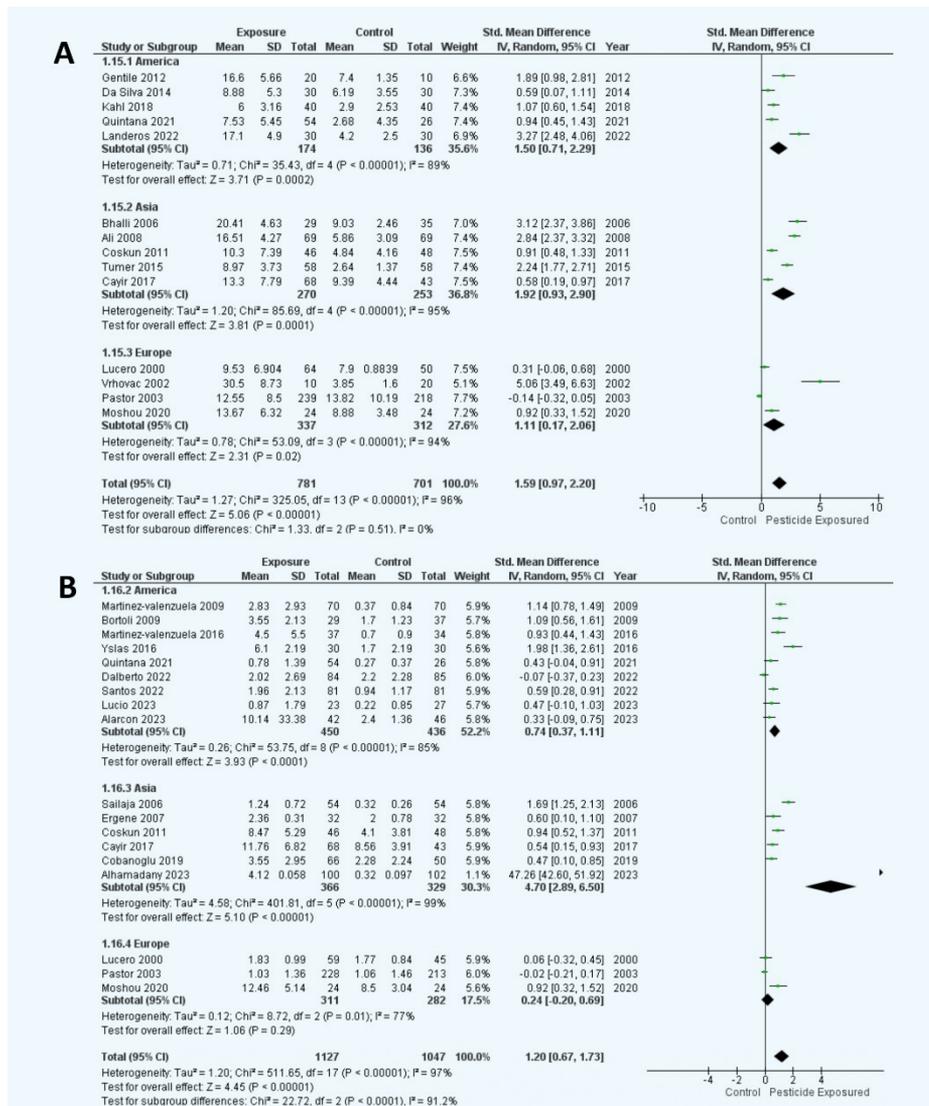


Fig 2. Meta-analysis of Occupational Pesticide Exposure on Micronucleus Frequency. (A) Micronucleus frequency in lymphocytes; (B) Micronucleus frequency in buccal cells.

the regional subgroup analysis showed no significant difference in the European region for both analyses, plus the Asian region for the buccal cells. These results should be interpreted carefully since the number of studies included in these regions was limited. Fig 4 shows the meta-analysis regarding the CBPI value. Based on the

trends of included studies, the CBPI is slightly reduced in the exposure group. However, the meta-analysis was not significant, suggesting that pesticide exposure does not significantly reduce the cell proliferation kinetics [SMD = -0.18, 95% CI (-0.90, 0.54), p = 0.63, I<sup>2</sup> = 96%].

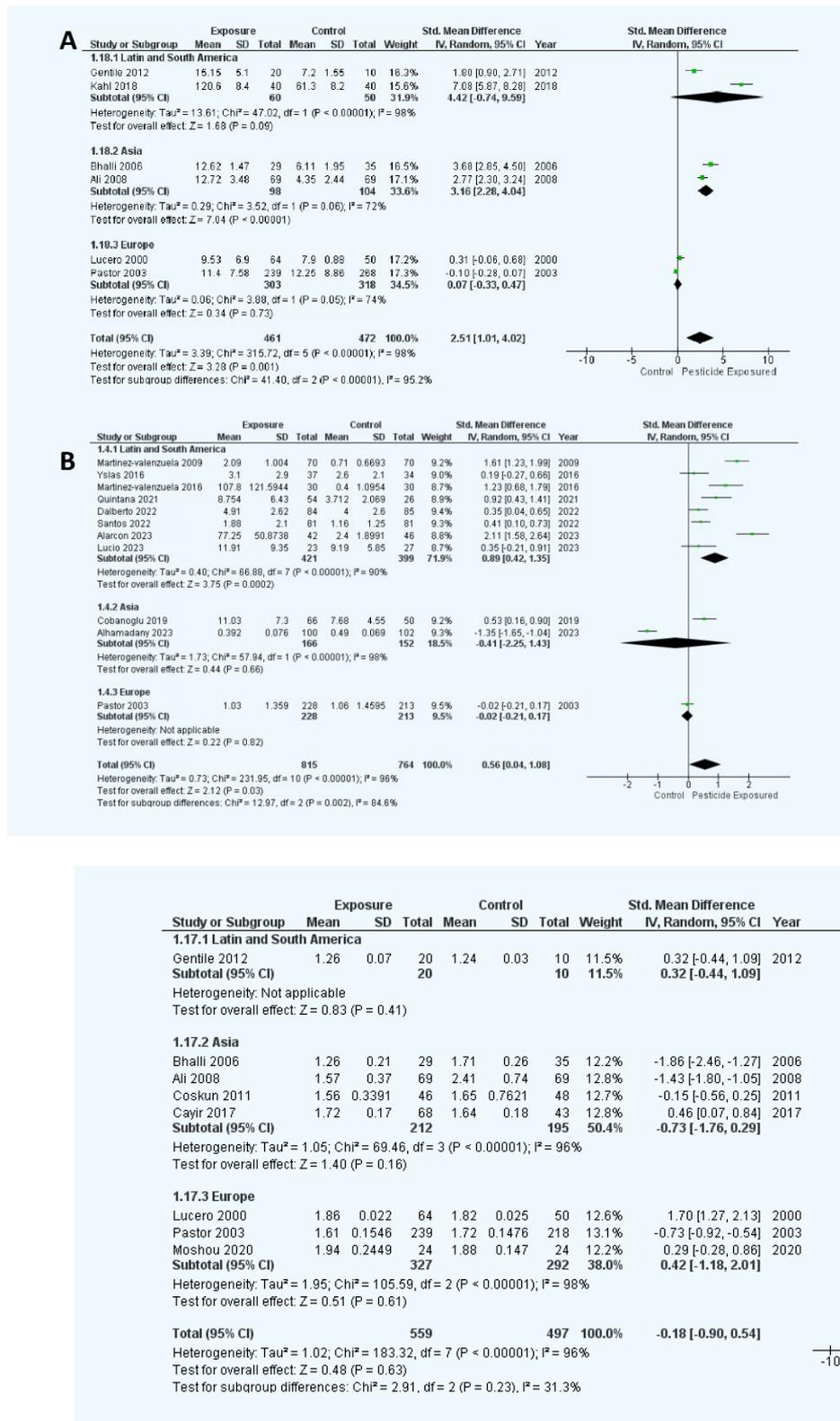


Fig 3. Meta-analysis of Occupational Pesticide Exposure on Binucleated Cells. (A) Binucleated cells in lymphocytes; (B) Binucleated cells in buccal cells.

Fig 4. Meta-analysis of Occupational Pesticide Exposure on CBPI.

**Meta-analysis based on gender in exposure group**

**TABLE 2** shows the summary of the meta-analysis based on the gender in the exposure group. Three analyses used fixed effect model (micronucleus in buccal cells, binucleated cells in buccal, and CBPI) due to low heterogeneity proven by  $I^2=0\%$ . All of the meta-analysis reported that there were no significant difference between each gender in the exposure group, indicating the effects of pesticide exposure were the same among male and female workers in terms of micronucleus, binucleated cells, and CBPI. The forest plot of gender in the exposure group is shown in **Supplementary Fig 2**.

**Meta-analysis based on smoking status in exposure group**

The summary of meta-analysis based on smoking status is shown in **Table 3**. Regarding the smokers

compared to non-smokers in exposed individuals, there was significant increase of micronucleus and binucleated cells in lymphocytes within the smokers in the pesticide exposure group, [Fixed, SMD = 0.55, 95% CI (0.18, 0.92),  $p = 0.004$   $I^2 = 49\%$ ] and [Random, SMD = 3.89, 95% CI (0.45, 7.33),  $p = 0.03$   $I^2 = 95\%$ ] respectively. Meta-analysis on micronucleus in buccal cells and CBPI was not significant. However, the results should be interpreted carefully due to the lack of studies that mentioned biomonitoring assessment among smokers and non-smokers. We did not perform a meta-analysis on binucleated cells in buccal cells because there was only one study that mentioned the data and which showed no significant difference between smokers and non-smokers.<sup>53</sup> The forest plot of smoking status in the exposure group is shown in **Supplementary Fig 3**.

**TABLE 2.** Summary of meta-analysis based on gender in exposure group.

Analysis	Studies	Male/Female	Results*
Micronucleus in Lymphocytes	5	334 / 109	[Random, SMD = 0.14, 95% CI (-0.37, 0.64), $p = 0.60$ $I^2 = 75\%$ ]
Micronucleus in Buccal Cells	4	342 / 105	[Fixed, SMD = 0.13, 95% CI (-0.10, 0.36), $p = 0.27$ $I^2 = 0\%$ ]
Binucleated Cells in Lymphocytes	3	266 / 87	[Random, SMD = -0.32, 95% CI (-0.98, 0.34), $p = 0.34$ $I^2 = 82\%$ ]
Binucleated Cells in Buccal	3	300 / 93	[Fixed, SMD = 0.18, 95% CI (-0.07, 0.42), $p = 0.16$ $I^2 = 0\%$ ]
CBPI	2	232 / 75	[Fixed, SMD = 0.10, 95% CI (-0.16, 0.37), $p = 0.45$ $I^2 = 0\%$ ]

\*(male=SMD positive; female=SMD negative)

**Abbreviation:** CBPI: Cytokinesis Block Proliferation Index

**TABLE 3.** Summary of meta-analysis based on smoking status in exposure group.

Analysis	Studies	S / NS	Results
Micronucleus in Lymphocytes	4	58 / 67	[Fixed, SMD = 0.55, 95% CI (0.18, 0.92), $p = 0.004$ $I^2 = 49\%$ ]
Micronucleus in Buccal Cells	3	44 / 80	[Random, SMD = 0.42, 95% CI (-0.87, 1.71), $p = 0.53$ $I^2 = 90\%$ ]
Binucleated Cells in Lymphocytes	3	32 / 38	[Random, SMD = 3.89, 95% CI (0.45, 7.33), $p = 0.03$ $I^2 = 95\%$ ]
CBPI	3	32 / 38	[Random, SMD = 0.33, 95% CI (-1.83, 2.48), $p = 0.77$ $I^2 = 93\%$ ]

**Abbreviations:** S: Smokers; NS: Non-Smokers; CBPI: Cytokinesis Block Proliferation Index

### Publication bias and sensitivity analysis

Egger's test and funnel plot were performed for micronuclei and buccal binucleated cells (Supplementary Fig 4). The funnel plot for micronuclei both in lymphocytes and buccal cells showed asymmetry along with significant Egger's test while the analysis in buccal binucleated cells showed vice versa. Leave-one-out analysis was used to perform the sensitivity analysis, removing each study individually to demonstrate how each study affected the pooled SMD and 95% CI data. The resume of the sensitivity analysis is presented in Supplementary Fig 5. There was no significant difference reported in the pooled SMD after the sensitivity analysis, suggesting that all of the analyses were reliable.

### DISCUSSION

Biomarkers of genomic instability and genotoxic damage include micronuclei and binucleated cells, which are caused by chromosome alterations or DNA damage. The potential danger is that the damaged cell will continue to divide and disrupt the cell's physiological or metabolic functions if the genotoxic damage is not repaired naturally by cellular repair mechanisms or if the altered cell is not removed. This kind of damage has been suggested to be one of the primary mechanisms of chronic disorders and carcinogenesis.<sup>56</sup> The results of micronucleus in this study suggest a potential increase of micronuclei within the pesticide exposure group with [pooled SMD = 1.59, 95% CI (0.97, 2.20),  $p < 0.00001$ ,  $I^2 = 96\%$ ] and [pooled SMD = 1.20, 95% CI (0.67, 1.73),  $p < 0.00001$ ,  $I^2 = 97\%$ ] for lymphocytes and buccal cells, respectively. Regarding the binucleated cells, the pooled SMD may indicate increase of binucleated cells in the pesticide exposure group for lymphocytes and buccal cells with [SMD = 2.51, 95% CI (1.01, 4.02),  $p < 0.001$ ,  $I^2 = 98\%$ ] and [pooled SMD = 0.56, 95% CI (0.04, 1.108),  $p = 0.03$ ,  $I^2 = 96\%$ ], respectively. However, with only limited studies, the power to detect moderators was insufficient, and we cannot rule out unmeasured confounding (e.g., genetic factors, exact exposure levels, and PPE used, due to differences in reporting across studies).

Pesticide as environmental exposure was believed to have the capacity to delay the cell cycle and decrease cell proliferation.<sup>57,58</sup> Cell proliferation or the ability of the cells to repair themselves is represented in CBPI value.<sup>21</sup> One possible explanation for the cell cycle delay is a long-term low-level exposure that can trigger apoptosis.<sup>57</sup> However, our meta-analysis of CBPI suggests the trend is slightly reduced in the exposure group but not significant [SMD = -0.18, 95% CI (-0.90, 0.54),  $p = 0.63$ ,  $I^2 = 96\%$ ]. The meta-analysis may indicate a negative correlation

between pesticide exposure and CBPI values, although this association did not reach statistical significance due to high heterogeneity and small sample size. It is possible that higher exposure levels could lead to reduced CBPI, but further research is needed to confirm this relationship.

Humans are currently exposed to a variety of genotoxic substances found in the contaminated environment. According to the United States Environmental Protection Agency, pesticides may be classified as carcinogenic or probably carcinogenic for humans.<sup>59</sup> Certain pesticides can cause genotoxicity, endocrine and chromosomal changes, as well as mutations and signaling abnormalities in embryonic or somatic cells.<sup>60</sup> Globally, acute pesticide poisoning and chronic exposure are a major cause of mortality and morbidity, particularly in developing countries where the regulations for pesticides are not well-developed.<sup>61</sup> Thus, to determine the degree of exposure and health risk, biomonitoring assessments have been developed to examine chromosomal or genetic damage resulting from various exposures, such as micronuclei, binucleated cells, comet assay, and chromosome aberrations.<sup>62</sup>

Micronucleus assays for pesticide exposure can be conducted using exfoliated buccal cells as well as peripheral blood cells (lymphocytes, PBL). This assessment may also be helpful in term of diagnosis and prognosis of certain cancer types and other disorders. Currently, micronucleus assays are the most used method for detecting DNA damaging effects brought on by environmental, lifestyle, occupational, and nutritional factors.<sup>63</sup> The included studies used Cytokinesis CBMN and BMCyt for lymphocytes and buccal cell samples, respectively. Micronuclei in binucleated cells are scored in the CBMN assay following cytochalasin B (Cyt-B)-mediated cytokinesis blockage. The selective examination of the micronucleus in cells that have undergone one mitosis is possible by the addition of Cyt-B.<sup>22</sup> Therefore, CBMN not only assesses micronucleus but also binucleated cells and may calculate CBPI. On the other hand, the micronucleus assessment in buccal cells followed a different procedure. After obtaining the samples, the BMCyt assay was carried out in a series of steps. These included creating a single-cell suspension, preparing the slide via cytocentrifugation, fixation, and staining with Feulgen and Light Green for inspection under bright field and fluorescence microscopy.<sup>64</sup>

Regarding the type of pesticides, there were many kinds of pesticides used in the included studies which are presented in Supplementary Table 2. Overall, we divided the type of pesticide into insecticide, herbicide, and fungicide. It has to be noted that not all published studies mentioned the chemical names of pesticides used. Some studies just mentioned "mixed pesticides"

or just mentioned the pesticide group. We arranged the pesticides used based on the information in each study. Organophosphate, carbamate, and pyrethroids are the most common pesticides used in the included studies. According to United States Environmental Protection Agency (USEPA) classification, the most frequent carcinogenic classifications are possible/suggestive carcinogenic for the organophosphate group and probably/likely carcinogenic for the carbamate and pyrethroids group.<sup>65,66</sup> Numerous carcinogenic chemicals have been shown to cause or are likely to cause malignant transformation; however, there is no widely accepted mechanistic basis for classifying chemical carcinogens other than genotoxicity, due to the diversity of biological processes involved.<sup>67</sup>

The results of the meta-analysis regarding the female workers in the exposure group revealed no significant difference. Males and females may respond differently to pesticides because of variations in body mass, fat, hormones, and organs.<sup>68</sup> Some of the instances of the sex-specific effects are reproductive issues, development of the fetus, and organ-specific functions such as the testes, ovaries, uterus, and breast. A pooled analysis of previous studies revealed a substantial correlation between women in the highest quintiles for DDE content (pesticide residue) and their relative risks for breast cancer.<sup>69</sup> A well-conducted research study in the United States examined over 55,000 men who used pesticides. The study discovered a higher risk of prostate cancer among these men, particularly in those with a family history of prostate cancer and in those who used the fumigant methyl bromide.<sup>70</sup>

Regarding smoking status, it is generally accepted that smoking cigarettes produces high concentrations of reactive oxygen species (ROS), and that ROS-induced oxidative DNA damage has been associated with cancer and cytotoxicity.<sup>71,72</sup> According to a study by Kumar, tobacco-specific nitrosamines are strong mutagenic and clastogenic agents that induce chromatid/chromosomal abnormalities which lead to the formation of micronuclei.<sup>73</sup> The detrimental effect of smoking on genotoxicity in the general population is widely known.<sup>74</sup> Therefore, this study provides another point of view of the smoking effects in agricultural workers who have exposure to pesticides. The results of the sub-analysis based on the smoking status in this study revealed that there was an increase in micronuclei and binucleated cells in lymphocytes within the pesticide exposure group. Therefore, in the smokers group the source of exposure was not only from pesticides but also through cigarette smoking. This combination may cause higher increase in genotoxicity. Given this, we may not recommend being smokers while still exposed to pesticides.

We performed subgroup analysis in the main outcome measured which showed a difference between the Europe, Asia, and Latin-south America regions. Almost all studies reported a significant increment except those from Europe (Fig 2 and Fig 3). The subgroup analysis of the Europe region within micronucleus in buccal, binucleated cells in lymphocytes, and binucleated cells in buccal cells showed no significant difference even though the pooled SMD was increased significantly with subgroup SMD as follows: [SMD = 0.24, 95% CI (-0.20, 0.69),  $p = 0.29$ ,  $I^2=77\%$ ], [SMD = 0.07, 95% CI (-0.33, 0.47),  $p = 0.73$ ,  $I^2=74\%$ ], and [SMD = -0.02, 95% CI (-0.21, 0.17),  $p = 0.82$ ], respectively. Furthermore, in the Asia region, the subgroup analysis of binucleated cells within buccal cells also showed no significant difference [SMD = -0.41, 95% CI (-2.25, 1.43),  $p = 0.66$ ,  $I^2=98\%$ ].

The difference in regional subgroup analysis may be due to pesticide restriction, PPE regulation, and other non-occupational exposure that differs in each region. Donley et al. reported that about 322 million pounds worth of the pesticides used in US agriculture were prohibited in Europe, 26 million pounds were banned in Brazil, and 40 million pounds were prohibited in China.<sup>75</sup> These reports indicate that pesticide restriction regulation is stricter in Europe than in other regions, resulting in limited distribution of dangerous pesticides. Another factor is the usage of PPE. In the systematic review, we found that almost all studies of micronuclei reported significant increase in the exposure group. However, two studies from Europe reported no statistically significant results.<sup>28,35</sup> These findings arise due to the usage of PPE by around 80% of workers in both studies. It is essential to wear adequate PPE during the exposure time including mask, gloves, goggles, boots, hat, and protective clothes.<sup>76</sup> Thus, the usage of inadequate PPE may cause weak protection against pesticide exposure, as reported by a study in Mexico where the PPE usage was around 87% (mask and gloves) but the micronucleus remains significantly increased.<sup>50</sup> In agriculture, where the risk of harm is significantly elevated, implementing effective risk mitigation techniques is essential for farmers, and the proper use of PPE is a critical measure. The Occupational Safety and Health Administration (OSHA) requires farms with 11 or more employees to thoroughly evaluate existing hazards, provide appropriate and effective PPE, and implement extensive training on its proper use.<sup>77</sup> The necessary PPE varies depending on the specific hazards in the agricultural workplace. Chemical personal protective equipment is crucial when handling pesticides to safeguard against their toxic effects. Referring to the Safety Data Sheet (SDS) is essential for guidance on suitable

attire, ocular protection, and breathing precautions. Suggested PPE may include caps, safety goggles, chemical-resistant gloves, and full-length shirts and trousers, effectively safeguarding agricultural workers from direct exposure to hazardous chemicals. Respiratory dangers in agricultural activities require the use of respirators to prevent inhalation of airborne hazards. Selecting the appropriate respirator is crucial, especially when hazardous substances are aerosolized and present an inhalation hazard to workers.<sup>78</sup> In addition, we must consider sources of pesticide exposure beyond occupational exposure. As mentioned before, there are four ways of pesticide exposure, including food ingestion, inhalation, dermal contact, and contaminated environment.<sup>5</sup> A study in Brazil reported that the frequency of micronuclei among the control group from the same area showed a slight increase compared to occupationally exposed workers although there was no statistical significance.<sup>51</sup> This study may indicate that there is another source of exposure that affects the control group, it may be from food or environment since all of the samples were from the same area. Based on the results of this study, we encourage the government to pay more attention to policies that require workers to use PPE to protect themselves, and to initiate a regular biomonitoring strategy using micronuclei screening to prevent long-term effects of pesticides and the development of cancer. Workers with an increased number of micronuclei or binucleated cells should be advised to wear more PPE and reduce their pesticide exposure.

This study has several limitations. First, the meta-analysis should be interpreted carefully since the total number of studies in each region may be different, confounding factors that may influence the results were different in each study included, we hardly found tobacco chewing status, which may affect micronuclei in buccal cells, among the included studies, and high heterogeneity was reported even though the leave-one-out analysis showed no difference. Moreover, the Healthy Worker Effect may distort this cross-sectional pesticide studies by favoring healthier current employees, perhaps leading to an underestimation of genotoxic risks, such as micronuclei development. Workers most susceptible to pesticides—exhibiting early cytogenetic damage, cytotoxicity, or symptoms—frequently leave the workforce owing to illness, job transition, or disability, resulting in a “survivor” population with diminished biomarker levels.<sup>79</sup> Second, several studies have a small sample size included in the analysis. Third, it is difficult to state which pesticide has more influence on micronuclei and cytotoxicity due to

the unavailability of data in published studies. It has to be noted that many factors may influence micronuclei and binucleated assessments such as age, smoking status, use of PPE, exposure time, and residence. Fourth, the funnel plot and Egger’s test in micronucleus suggested there was a publication bias indicating that studies with significant positive results were more likely to be published than studies with negative or not significant results.

## CONCLUSION

This systematic review and meta-analysis suggest that pesticide exposure may be associated with elevated frequencies of MN and binucleated cells. Nevertheless, significant heterogeneity and possible confounding variables undermine the reliability of this evidence. No notable correlation was detected for CBPI values. Further analyses indicated no substantial disparities in micronucleus, binucleated cell, or CBPI effects between males and females. Nonetheless, simultaneous smoking seems to raise the frequencies of micronuclei and binucleated cells in exposed individuals. The results indicate that PPE provides protection against genotoxicity generated by pesticides. We urge the government to prioritize legislation mandating the use of PPE for worker safety and to implement a systematic biomonitoring plan that uses MN screening to mitigate the long-term effects of pesticides and the onset of cancer. Workers exhibiting a higher frequency of MN or binucleated cells should be counseled to improve their use of personal protective equipment and minimize pesticide exposure.

## Data Availability Statement

All data are available from the corresponding author and can be shared on reasonable request.

## ACKNOWLEDGMENTS

We thank Prof. Chihaya Koriyama, M.D. (Department of Epidemiology and Preventive Medicine, Kagoshima University, Japan) and Ancah Caesarina Novi Marchianti, M.D., Ph.D. (Department of Public Health, University of Jember, Indonesia) who reviewed and gave valuable comments during the preparation of this manuscript.

## DECLARATIONS

### Grants and Funding Information

This systematic review and meta-analysis did not receive any specific funding.

### Conflict of Interests

All authors declared no conflict of interest.

**Registration Number of Clinical Trial**

Not applicable.

**Author Contributions**

Conceptualization and methodology, S and AIT ; Formal analysis, AIT, NA, and MRFH ; Visualization, MRFH and SB ; Supervision, ENS, AN, and S. Writing original draft AIT, NA, MRFH, SB, and AN; All authors have read and agreed to the final version of the manuscript.

**Use of Artificial Intelligence**

The author declares no use of any artificial intelligence tool.

**Ethics Approval and Consent to Participate**

Not applicable.

**Consent for Publication**

Not applicable.

**REFERENCES**

- Sibhatu KT, Qaim M. Rural food security, subsistence agriculture, and seasonality. *PLoS One*. 2017;12:e0186406.
- Viana CM, Freire D, Abrantes P, Rocha J, Pereira P. Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. *Sci Total Environ*. 2022;806:150718.
- Zhao W, Teng M, Zhang J, Wang K, Zhang J, Xu Y, et al. Insights into the mechanisms of organic pollutant toxicity to earthworms: Advances and perspectives. *Environ Pollut*. 2022;303:119120.
- Sharma A, Kumar V, Shahzad B, Tanveer M, Sidhu GPS, Handa N, et al. Worldwide pesticide usage and its impacts on ecosystem. *SN Appl Sci*. 2019;1:1446.
- Qi S-Y, Xu X-L, Ma W-Z, Deng S-L, Lian Z-X, Yu K. Effects of Organochlorine Pesticide Residues in Maternal Body on Infants. *Front Endocrinol (Lausanne)* 2022;13:890307.
- INCHEM. WHO-JMPR Toxicological Monographs-Pesticide Residues in Food. *InchemOrg*, 2022. [cited 2024 Feb 11]. Available from: <https://inchem.org/pages/jmpr.html>
- Leon ME, Schinasi LH, Lebailly P, Beane Freeman LE, Nordby K-C, Ferro G, et al. Pesticide use and risk of non-Hodgkin lymphoid malignancies in agricultural cohorts from France, Norway and the USA: a pooled analysis from the AGRICOH consortium. *Int J Epidemiol*. 2019;48:1519–35.
- Togawa K, Leon ME, Lebailly P, Beane Freeman LE, Nordby K-C, Baldi I, et al. Cancer incidence in agricultural workers: Findings from an international consortium of agricultural cohort studies (AGRICOH). *Environ Int*. 2021;157:106825.
- Eastmond D, Balakrishnan S. Hayes' Handbook of Pesticide Toxicology. 3rd ed. New York: Academic Press; 2010.
- Falzone L, Marconi A, Loreto C, Franco S, Spandidos DA, Libra M. Occupational exposure to carcinogens: Benzene, pesticides and fibers (Review). *Mol Med Rep*. 2016;14:4467–74.
- Alavanja MCR, Bonner MR. Occupational pesticide exposures and cancer risk: a review. *J Toxicol Environ Health B Crit Rev*. 2012;15:238–63.
- Fenech M, Holland N, Kirsch-Volders M, Knudsen LE, Wagner K-H, Stopper H, et al. Micronuclei and disease – Report of HUMN project workshop at Rennes 2019 EEMGS conference. *Mutat Res Genet Toxicol Environ Mutagen*. 2020;850–851:503133.
- Hartwig A, Arand M, Epe B, Guth S, Jahnke G, Lampen A, et al. Mode of action-based risk assessment of genotoxic carcinogens. *Arch Toxicol*. 2020;94:1787–877.
- de Souza DV, dos Anjos Rosario B, Takeshita WM, de Barros Viana M, Nagaoka MR, dos Santos JN, et al. Is micronucleus assay in oral exfoliated cells a suitable biomarker for predicting cancer risk in individuals with oral potentially malignant disorders? A systematic review with meta-analysis. *Pathol Res Pract*. 2022;232:153828.
- Fenech M. The in vitro micronucleus technique. *Mutat Res*. 2000;455(1-2):81–95.
- Bonassi S, Znaor A, Ceppi M, Lando C, Chang WP, Holland N, et al. An increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of cancer in humans. *Carcinogenesis*. 2006;28(3):625–31.
- Stoll LM, Duffield AS, Johnson MW, Ali SZ. Acute myeloid leukemia with myelodysplasia-related changes with erythroid differentiation involving pleural fluid: A case report and brief cytopathologic review. *Diagn Cytopathol*. 2011;39(6):451–4.
- Minimo C, Zakowski M, Lin O. Cytologic findings of malignant vascular neoplasms: A study of twenty-four cases. *Diagn Cytopathol*. 2002;26(6):349–55.
- Kimura N, Dota K, Araya Y, Ishidate T, Ishizaka M. Scoring system for differential diagnosis of malignant mesothelioma and reactive mesothelial cells on cytology specimens. *Diagn Cytopathol*. 2009;37(12):885–90.
- Nishimura K, Watanabe S, Hayashida R, Sugishima S, Iwasaka T, Kaku T. Binucleated HeLa cells are formed by cytokinesis failure in starvation and keep the potential of proliferation. *Cytotechnology*. 2016;68(4):1123–30.
- Ruiz-Ruiz B, Arellano-García ME, Radilla-Chávez P, Salas-Vargas DS, Toledano-Magaña Y, Casillas-Figueroa F, et al. Cytokinesis-Block Micronucleus Assay Using Human Lymphocytes as a Sensitive Tool for Cytotoxicity/Genotoxicity Evaluation of AgNPs. *ACS Omega*. 2020;5(21):12005–15.
- Fenech M. Cytokinesis-block micronucleus cytome assay. *Nat Protoc*. 2007;2(5):1084–104.
- Thomas P, Holland N, Bolognesi C, Kirsch-Volders M, Bonassi S, Zeiger E, et al. Buccal micronucleus cytome assay. *Nat Protoc*. 2009;4(6):825–37.
- Yang L, Baumann C, De La Fuente R, Viveiros MM. Mechanisms underlying disruption of oocyte spindle stability by bisphenol compounds. *Reproduction*. 2020;159(4):383–96.
- Cobanoglu H, Cayir A. Assessment of the genotoxic potential of tetrachlorvinphos insecticide by cytokinesis-block micronucleus and sister chromatid exchange assays. *Hum Exp Toxicol*. 2021;40(12 Suppl):S158–63.
- Ozkan D, Yüzbaşıoğlu D, Unal F, Yılmaz S, Aksoy H. Evaluation of the cytogenetic damage induced by the organophosphorous insecticide acephate. *Cytotechnology*. 2009;59(2):73–80.
- Moshou H, Karakitsou A, Yfanti F, Hela D, Vlastos D, Paschalidou AK, et al. Assessment of genetic effects and pesticide exposure of farmers in NW Greece. *Environ Res*. 2020;186:109558.
- Pastor S, Creus A, Parrón T, Cebulska-Wasilewska A, Siffel C, Piperakis S, et al. Biomonitoring of four European populations occupationally exposed to pesticides: use of micronuclei as

- biomarkers. *Mutagenesis*. 2003;18(3):249–58.
29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62(10):1006–12.
  30. Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, et al. A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE. *Environ Int*. 2019;122:168–84.
  31. Higgins T, Green S, Thomas J, Chandler J, Cumpston M, Li T, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley; 2011. Available from: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook>
  32. Alhamadany AYM, Khalaf SD, Alkateb YNM. Genotoxicity and genomic instability in oral epithelial cells of agricultural workers exposed to pesticides using micronucleus and comet assay in Nineveh, Iraq. *Journal of Applied and Natural Science*. 2023;15:473–9.
  33. Ergene S, Çelik A, Çavaş T, Kaya F. Genotoxic biomonitoring study of population residing in pesticide contaminated regions in Göksu Delta: Micronucleus, chromosomal aberrations and sister chromatid exchanges. *Environ Int*. 2007;33(7):877–85.
  34. Kahl VFS, Dhillon V, Fenech M, De Souza MR, Da Silva FN, Marroni NAP, et al. Occupational exposure to pesticides in tobacco fields: The integrated evaluation of nutritional intake and susceptibility on genomic and epigenetic instability. *Oxid Med Cell Longev*. 2018;2018:7017423.
  35. Lucero L, Pastor S, Suarez S, Durban R, Gomez C, Parron T, et al. Cytogenetic biomonitoring of Spanish greenhouse workers exposed to pesticides: micronuclei analysis in peripheral blood lymphocytes and buccal epithelial cells. *Mutation Research*. 2000;464(2):255–62.
  36. Garaj-Vrhovac V, Zeljezic D. Assessment of genome damage in a population of Croatian workers employed in pesticide production by chromosomal aberration analysis, micronucleus assay and Comet assay. *J Appl Toxicol*. 2002;22(4):249–55.
  37. Bhalli JA, Khan QM, Haq MA, Khalid AM, Nasim A. Cytogenetic analysis of Pakistani individuals occupationally exposed to pesticides in a pesticide production industry. *Mutagenesis*. 2006;21(2):143–8.
  38. Sailaja N, Chandrasekhar M, Rekhadevi PV, Mahboob M, Rahman MF, Vuyyuri SB, et al. Genotoxic evaluation of workers employed in pesticide production. *Mutat Res*. 2006;609(1):74–80.
  39. Ali T, Bhalli JA, Rana SM, Khan QM. Cytogenetic damage in female pakistani agricultural workers exposed to pesticides. *Environ Mol Mutagen*. 2008;49(5):374–80.
  40. Bortoli GM de, Azevedo MB de, Silva LB da. Cytogenetic biomonitoring of Brazilian workers exposed to pesticides: Micronucleus analysis in buccal epithelial cells of soybean growers. *Mutat Res*. 2009;675(1-2):1–4. H
  41. Martínez-Valenzuela C, Gómez-Arroyo S, Villalobos-Pietrini R, Waliszewski S, Calderón-Segura ME, Félix-Gastélum R, et al. Genotoxic biomonitoring of agricultural workers exposed to pesticides in the north of Sinaloa State, Mexico. *Environ Int*. 2009;35(8):1155–9.
  42. Coskun M, Coskun M, Cayir A, Ozdemir O. Frequencies of micronuclei (MNI), nucleoplasmic bridges (NPBs), and nuclear buds (NBUDs) in farmers exposed to pesticides in Çanakkale, Turkey. *Environ Int*. 2011;37(1):93–6.
  43. Gentile N, Mañas F, Bosch B, Peralta L, Gorla N, Aiassa D. Micronucleus assay as a biomarker of genotoxicity in the occupational exposure to agrochemicals in rural workers. *Bull Environ Contam Toxicol*. 2012;88(6):816–22.
  44. Da Silva FR, Kvitko K, Rohr P, Abreu MB, Thiesen F V, Da Silva J. Genotoxic assessment in tobacco farmers at different crop times. *Sci Total Environ*. 2014;490:334–41.
  45. Tumer TB, Savranoglu S, Atmaca P, Terzioglu G, Sen A, Arslan S. Modulatory role of GSTM1 null genotype on the frequency of micronuclei in pesticide-exposed agricultural workers. *Toxicol Ind Health*. 2015;32(12):1942–51.
  46. Martínez-Valenzuela C, Waliszewski SM, Amador-Muñoz O, Meza E, Calderón-Segura ME, Zenteno E, et al. Aerial pesticide application causes DNA damage in pilots from Sinaloa, Mexico. *Environ Sci Pollut Res Int*. 2016;24(3):2412–20.
  47. Castañeda-Yslas IJ, Arellano-García ME, García-Zarate MA, Ruíz-Ruiz B, Zavala-Cerna MG, Torres-Bugarín O. Biomonitoring with Micronuclei Test in Buccal Cells of Female Farmers and Children Exposed to Pesticides of Maneadero Agricultural Valley, Baja California, Mexico. *J Toxicol*. 2016;2016:7934257.
  48. [Çayir A, Coskun M, Coskun M, Cobanoglu H. DNA damage and circulating cell free DNA in greenhouse workers exposed to pesticides. *Environ Mol Mutagen*. 2018;59(2):161–9.
  49. Cobanoglu H, Coskun M, Coskun M, Çayir A. Results of buccal micronucleus cytome assay in pesticide-exposed and non-exposed group. *Environ Sci Pollut Res Int*. 2019;26(19):19676–83.
  50. Valencia-Quintana R, López-Durán RM, Milić M, Bonassi S, Ochoa-Ocaña MaA, Uriostegui-Acosta MO, et al. Assessment of Cytogenetic Damage and Cholinesterases' Activity in Workers Occupationally Exposed to Pesticides in Zamora-Jacona, Michoacan, Mexico. *Int J Environ Res Public Health*. 2021;18(12):6269.
  51. Dalberto D, Alves J, Garcia ALH, de Souza MR, Abella AP, Thiesen F V, et al. Exposure in the tobacco fields: Genetic damage and oxidative stress in tobacco farmers occupationally exposed during harvest and grading seasons. *Mutat Res Genet Toxicol Environ Mutagen*. 2022;878:503485.
  52. Landeros N, Duk S, Márquez C, Inzunza B, Acuña-Rodríguez IS, Zúñiga-Venegas LA. Genotoxicity and Reproductive Risk in Workers Exposed to Pesticides in Rural Areas of Curicó, Chile: A Pilot Study. *Int J Environ Res Public Health*. 2022;19(24):16608.
  53. dos Santos IC, da Silva JT, Rohr P, Lengert A van H, de Lima MA, Kahl VFS, et al. Genomic instability evaluation by BMCyt and telomere length in Brazilian family farmers exposed to pesticides. *Mutat Res Genet Toxicol Environ Mutagen*. 2022;878:503479.
  54. Sánchez-Alarcón J, Milić M, Bonassi S, Gómez-Arroyo S, Cortés-Eslava J, Flores-Márquez AR, et al. Occupational exposure to pesticides: DNA damage in horticulturist from Nativitas, Tlaxcala in Mexico. *Environ Toxicol Pharmacol*. 2023;100:104141.
  55. Lucio FT, Almeida IV, Buzo MG, Vicentini VEP. Genetic instability in farmers using pesticides: A study in Brazil with analysis combining alkaline comet and micronucleus assays. *Mutat Res Genet Toxicol Environ Mutagen*. 2023;886:503587.
  56. Mostafalou S, Abdollahi M. Pesticides and human chronic diseases: evidences, mechanisms, and perspectives. *Toxicol Appl Pharmacol*. 2013;268(2):157–77.
  57. Kirsch-Volders M, Fenech M. Inclusion of micronuclei in non-divided mononuclear lymphocytes and necrosis/apoptosis may provide a more comprehensive cytokinesis block micronucleus assay for biomonitoring purposes. *Mutagenesis*. 2001;16(1):51–8.
  58. Alwaeli AM, Al-Marayaty SS, Al-Khalisy MH, Assi MH. Morphologic Features and Morphometric Measurements of

- Human Oocytes That Failed to Cleave after Intracytoplasmic Sperm Injection (ICSI). *Siriraj Med J.* 2026;78(1):1–10.
59. [USEPA. Chemicals Evaluated for Carcinogenic Potential: Office of Pesticides Programs. 2022. Available from: [https://npic.orst.edu/chemicals\\_evaluated.pdf](https://npic.orst.edu/chemicals_evaluated.pdf)
60. [Sasikala S, Minu Jenifer M, Velavan K, Sakthivel M, Sivasamy R, Fenwick Antony ER. Predicting the relationship between pesticide genotoxicity and breast cancer risk in South Indian women in in vitro and in vivo experiments. *Sci Rep.* 2023;13(1):9712.
61. Boedeker W, Watts M, Clausing P, Marquez E. The global distribution of acute unintentional pesticide poisoning: estimations based on a systematic review. *BMC Public Health.* 2020;20(1):1875.
62. Sommer S, Buraczewska I, Kruszewski M. Micronucleus Assay: The State of Art, and Future Directions. *Int J Mol Sci.* 2020;21(4):1534.
63. [Nersesyan A, Kundi M, Fenech M, Stopper H, da Silva J, Bolognesi C, et al. Recommendations and quality criteria for micronucleus studies with humans. *Mutat Res Rev Mutat Res.* 2022;789:108410.
64. Thomas P, Holland N, Bolognesi C, Kirsch-Volders M, Bonassi S, Zeiger E, et al. Buccal micronucleus cytome assay. *Nat Protoc.* 2009;4(6):825–37.
65. Schwingl PJ, Lunn RM, Mehta SS. A tiered approach to prioritizing registered pesticides for potential cancer hazard evaluations: implications for decision making. *Environmental Health* 2021; 20(1):13.
66. USEPA. USEPA Pesticide Chemical Search Database. *OrdsPubEpaGov*, 2023. <https://ordspub.epa.gov/ords/pesticides/?p=CHEMICALSEARCH:1>
67. Stewart B. Mechanisms of carcinogenesis: from initiation and promotion to hallmarks. Lyon: International Agency for Research on Cancer Scientific Publications; 2019.
68. D'Archivio M, Coppola L, Masella R, Tammaro A, La Rocca C. Sex and Gender Differences on the Impact of Metabolism-Disrupting Chemicals on Obesity: A Systematic Review. *Nutrients.* 2024;16(2):181.
69. [Ledda C, Bracci M, Lovreglio P, Senia P, Larrosa M, Martinez-Jarreta B, et al. Pesticide exposure and gender discrepancy in breast cancer. *Eur Rev Med Pharmacol Sci.* 2021;25(7):2898–915.
70. Alavanja MCR, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF, et al. Use of agricultural pesticides and prostate cancer risk in the Agricultural Health Study cohort. *Am J Epidemiol.* 2003;157(9):800–14.
71. Chen Z, Wang D, Liu X, Pei W, Li J, Cao Y, et al. Oxidative DNA damage is involved in cigarette smoke-induced lung injury in rats. *Environ Health Prev Med.* 2015;20(5):318–24.
72. Soi-ampornkul R, Ghimire S, Thangnipon W, Suwanna N, Vatanashevanopakorn C. Curcumin Attenuates Hydrogen Peroxide-Induced Cytotoxicity in Human Neuroblastoma SK-N-SH Cells. *Siriraj Med J.* 2018;70(3):184–90.
73. Kumar V, Rao NN, Nair NS. Micronuclei in oral squamous cell carcinoma. A marker of genotoxic damage. *Indian J Dent Res.* 2000;11(3):101–6.
74. Simanjuntak AM, Putra MRE, Amalia NP, Hutapea A, Suyanto S, Siregar IE. Lung and Airway Disease Caused by E-Cigarette (Vape): A Systematic Review. *Siriraj Med J.* 2024;76(5):325–32.
75. Donley N. The USA lags behind other agricultural nations in banning harmful pesticides. *Environ Health.* 2019;18(1):44.
76. FAO, WHO. International Code of Conduct on Pesticide Management: Guidelines on licensing of public health pest control operators. Rome: WHO Press; 2015.
77. OSHA. Agricultural operations - hazards and controls. Occupational Safety and Health Administration, United States Department of Labor, 2022. [cited 2026 Jan 19]. Available from: <https://www.osha.gov/agricultural-operations/hazards>
78. USDA. Personal Protective Equipments. United States Department of Agriculture Agricultural Research Service, 2016. [cited 2026 Jan 19]. Available from: <https://www.ars.usda.gov/northeast-area/docs/safety-health-and-environmental-training/personal-protective-equipment/>
79. Chowdhury R, Shah D, Payal AR. Healthy Worker Effect Phenomenon: Revisited with Emphasis on Statistical Methods - A Review. *Indian J Occup Environ Med.* 2017;21(1):2–8.