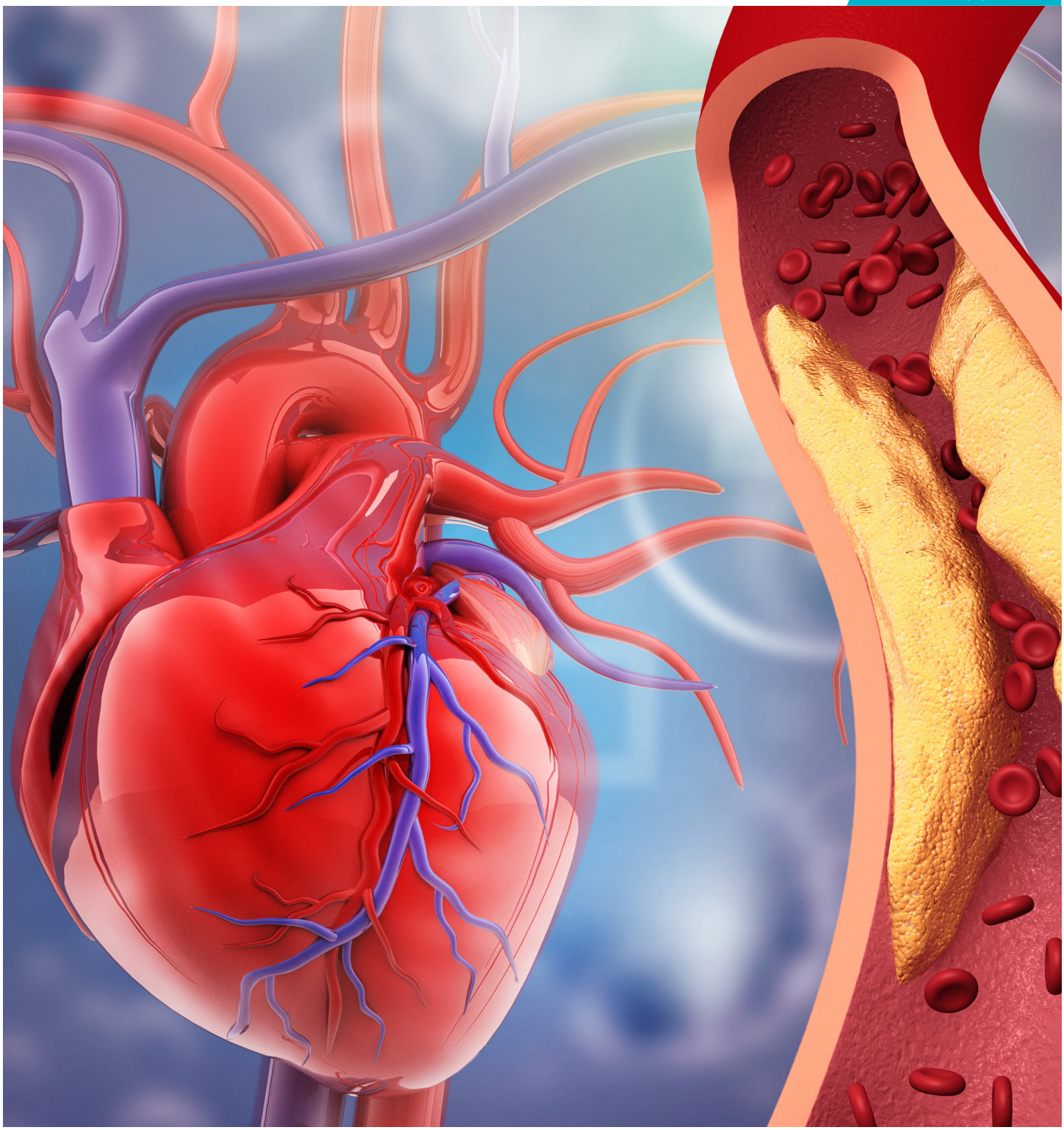
JCMK

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN

EFQM
Recognised for exellence
5 star

Online ISSN 2313-1519
Print ISSN 1812-2892
www.clinmedkaz.org

№23 (3) 2026



Psychometric Validation of the Kazakh Version of the 9-Item European Heart Failure Self-Care Behaviour Scale

> See page 17 and 23

Depression Prevalence in Urban Indonesia: a Preliminary Study of the Makassar General Population

> See page 62 and 68

Young Patient Presenting with ST-elevation Myocardial Infarction (STEMI) Following Alcohol Mixed with Energy Drink (AMED)

> See page 105 and 118

JCMK

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN

Online ISSN 2313-1519
Print ISSN 1812-2892
№23 (3) 2026
Published since 2004.

EDITORIAL

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN

Scientific and practical journal

AIMS AND SCOPE OF THE JOURNAL

Journal "Clinical Medicine of Kazakhstan" (ISSN 1812-2892) is a multi-field dedicated peer-reviewed medical journal. The main thematic scope – publication of materials on medical science and practice, education and healthcare organization. Joint Stock Company "National Scientific Medical Center" publishes the journal bimonthly in a year (in February, April, June, August, October, and December).

All articles sent to editors undergo double-blind review. Manuscripts are judged by two experts exclusively on the basis of their contribution to initial data, ideas and their presentations. Editors accept articles for consideration and publication at no cost. Detailed information is available in the section Information for authors at the end of this material.

The Journal of "Clinical Medicine of Kazakhstan" to the full extent is wedded to initiative of open access and ready to provide free access to full texts of articles, as soon as they will be published and available in the Internet (www.clinmedkaz.org).

Journal was registered in the Ministry of Information of the RK on 05.04.2004 and currently included to the list of Publications, approved by the Committee for Control of Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan for publication of the main outcomes of scientific activity.

The journal is indexed in such international scientific-information bases as Scopus, Index Copernicus International, Google Scholar, CrossRef. DOAJ.

JCMK
JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN

Editorial Office: Journal of Clinical Medicine of Kazakhstan
Ave Abylay-khan #42, 010009, Astana, Kazakhstan. Phone: +7(7172) 577411
Fax: +7(7172) 232927 E-mail: editor@clinmedkaz.org Web Adress: www.clinmedkaz.org

NATIONAL SCIENTIFIC MEDICAL CENTER JSC, ASTANA CITY, REPUBLIC OF KAZAKHSTAN

TREATMENT OF TUMORS AND PARASITIC CYSTS USING HIGH-INTENSITY FOCUSED ULTRASOUND

The latest HIFU installation- JC therapy has been in use in NSMC since January 2010. This installation allows noncontact complete eliminating the endogenic tumor using high intensity focused ultrasound without cutting the tissues and injury of not affected organs.

NSMC successfully treats: mammary and alvus fibroadenoma; breast cancer; liver tumors and cysts (primary liver cancer, liver metastases, echinococcosis, alveococcosis); benign and malignant pancreatic tumors; benign and malignant tumors of kidneys; osteogenic and myelosarkoma of extremities (soft tissues and bones cancer of the extremities).

HIFU-therapy of echinococcosis and alveococcosis developed in the clinic is the one and only in the world and is an alternative to surgical treatment of this disease, causing the economic feasibility.

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/10.23950/jcmk/18692

3-Three Year BAROS Score Outcomes for One anastomosis Gastric Bypass and FundoRing One anastomosis Gastric Bypass

Oral Ospanov[1], Galymjan Duysenov[2], Venera Rakhmetova[3], Bakhtiyar Yelembayev[4], Zhanbolat Dildabekov[4]

[1]Research and Education Center for Surgery named after Professor Tsoi G. V., Astana medical university, Astana, Kazakhstan
[2]Surgery Center of Professor Oral Ospanov, Astana, Kazakhstan
[3]Departments of Internal Medicine with Courses in Nephrology, Hematology, Allergology, and Immunology, Astana medical university, Astana, Kazakhstan
[4]Public Association "Society of Bariatric and Metabolic Surgeons of Kazakhstan", Astana, Kazakhstan

Received: 2026-03-04.
Accepted: 2026-05-31.

J Clin Med Kaz 2026; 23(3): 4-10

Corresponding author:
Galymjan Duysenov.
E-mail: kazbareo@gmail.com.
ORCID: 0009-0004-5701-5788.

Abstract

Objective: To compare the 3-year quality of life, weight loss, comorbidity remission, and complication rates between FundoRing One-Anastomosis Gastric Bypass (f-OAGB) and standard OAGB (s-OAGB) in patients with obesity.

Background: While OAGB is an effective metabolic procedure, it is associated with specific complications like bile reflux and marginal ulcers. The FundoRing modification, involving total fundoplication of the excluded stomach, may mitigate these risks. Long-term, patient-centered outcomes measured by the Bariatric Analysis and Reporting Outcome System (BAROS) are crucial.

Methods: In this single-center, prospective randomized trial, 1000 patients (BMI 30-50 kg/m²) were assigned 1:1 to f-OAGB (n=500) or s-OAGB (n=500) between January 2021 and December 2024. A detailed CONSORT flow diagram guided the study. Randomization was performed by a statistician using sequentially numbered opaque envelopes. Outcome assessors were blinded. The primary outcome was the total BAROS score at 3 years. Safety was assessed using structured criteria, including 30-day and late complications (Clavien-Dindo classification).

Results: At 3 years, follow-up was 83.0% (415/500) for f-OAGB and 87.0% (435/500) for s-OAGB (reasons for attrition: lost to follow-up, protocol violations, withdrawal of consent). The f-OAGB group demonstrated a significantly higher total BAROS score (5.59±1.13 vs. 4.57±1.09; mean difference 1.02, 95% CI 0.87–1.17; p<0.001), indicating a "very good" vs. "good" outcome. f-OAGB resulted in superior %EWL (84.3±12.1% vs. 76.8±14.3%; p<0.001), greater improvement in medical conditions, and better quality of life. No intraoperative, 30-day, or late mortality occurred in either group during the 3-year follow-up period (overall mortality 0%). f-OAGB was associated with significantly lower rates of bile reflux (0.2% vs. 9.4% at 3 years, p<0.001), marginal ulcers (0.2% vs. 3.4% at 3 years, p<0.001), and dumping syndrome (13.7% vs. 36.5% at 3 years, p<0.001).

Conclusions: At 3 years post-surgery, FundoRing OAGB is superior to standard OAGB in achieving better overall health outcomes as measured by the BAROS score. It provides enhanced weight loss, greater comorbidity resolution, and superior quality of life, while significantly reducing the risk of bile reflux, marginal ulceration, and dumping syndrome. F-OAGB represents a safe and effective technical advancement in metabolic surgery.

Keywords: OAGB; FundoRing; BAROS; Quality of Life; Metabolic Surgery

Introduction

Today obesity is a major global health challenge with important clinical implications [1]. The global prevalence of obesity has reached pandemic proportions, establishing it as a major public health crisis [2]. Bariatric and metabolic surgery (MBS) is the most effective and durable intervention for severe obesity, leading to significant and sustained weight loss and remission of associated medical conditions [3]. Modern technologies, combined with the surgeon's skills, have made it so safe that it allows for discharge on the day of surgery without increasing risks for the patient [4]. Among the various MBS procedures, One-Anastomosis Gastric Bypass (OAGB) has gained widespread acceptance due to its technical simplicity and favorable metabolic outcomes compared to Roux-en-Y Gastric Bypass (RYGB) [5]. However, OAGB is not without its drawbacks. Key concerns include the potential for bile reflux esophagitis and a notable incidence of marginal ulcers at the gastrojejunal anastomosis.

To address these complications, technical modifications have been proposed. One such innovation is the FundoRing OAGB (f-OAGB), which incorporates a total fundoplication of the bypassed (excluded) stomach. This maneuver aims to create a physiologic antireflux barrier, potentially mitigating the risk of postoperative bile reflux [6] and preventing dilatation [7]. While early data on the feasibility of f-OAGB exist, its long-term impact on patient-centered outcomes, quality of life (QoL), and specific complication profiles has not been rigorously compared to the standard OAGB (s-OAGB) technique in a large, randomized setting.

The Bariatric Analysis and Reporting Outcome System (BAROS) [8] provides a validated, multidimensional tool for evaluating the results of bariatric surgery. By integrating weight loss, improvement in co-morbidities, and QoL, while deducting points for complications and reoperations, BAROS offers a holistic and clinically meaningful assessment of surgical success. Previous studies [9] have utilized BAROS to compare different procedures, highlighting its utility in guiding clinical decision-making.

Therefore, the aim of this prospective, randomized controlled trial was to investigate and compare the three-year BAROS scores of patients with obesity undergoing f-OAGB versus s-OAGB. We hypothesized that the FundoRing modification would result in superior overall outcomes, primarily driven by a reduction in procedure-specific complications and a consequent improvement in QoL. A secondary aim was to perform a structured safety analysis using the Clavien-Dindo classification.

Methods

Study Design and Participants

This was a single-center, parallel-group, prospective randomized controlled trial conducted at a university-affiliated bariatric center (Astana Medical University, Kazakhstan) between January 2021 and December 2024. The study protocol was approved by the Institutional Ethics Committee of Astana Medical University (Session No. 12, Decision No. 24, 2023) and was conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent.

Inclusion criteria were: age 18-60 years; body mass index (BMI) 30-50 kg/m²; failure of previous non-surgical weight loss attempts. Both healthy obese patients and those with stable chronic diseases (type 2 diabetes, hypertension, dyslipidemia, non-alcoholic fatty liver disease) were included if they met the BMI criteria.

Exclusion criteria were: super-obesity (BMI >50 kg/m²); age >60 years or <18 years; severe reflux esophagitis (Grade C or D by Los Angeles classification); previous bariatric or major upper gastrointestinal surgery; large hiatal hernia (>5 cm); active malignancy or pregnancy; inability to provide informed consent or adhere to follow-up.

Randomization and Masking

Patients who met the inclusion criteria and did not meet the exclusion criteria were consecutively included and randomized into one of the two study arms by the study statistician, who participated only in the randomization and statistical analysis. The randomization list was maintained in a strictly confidential manner, and allocation concealment was ensured using sequentially numbered, identical, opaque, sealed envelopes. The intervention was randomly assigned to each patient during the pre-procedure visit by a nurse who did not participate in patient enrollment or assessment. Patients and the operating surgeon were not blinded due to the nature of the procedure. However, outcome assessors (endoscopists, laboratory technicians, and BAROS questionnaire administrators) were blinded to group allocation.

Surgical Techniques

All procedures were performed by a single, experienced metabolic and bariatric surgeon. The control procedure was standard laparoscopic OAGB (s-OAGB) [10]. The comparator procedure was FundoRing OAGB (f-OAGB) incorporating total fundoplication of the OAGB-excluded stomach [11].

Postoperative Management and Follow-up

A standardized enhanced recovery after surgery (ERAS) protocol [12] was followed for all patients. Patients were discharged on a liquid diet with daily proton pump inhibitor (PPI) therapy (pantoprazole 40 mg) prescribed for the first 2-6 months postoperatively. Follow-up visits were scheduled at 1, 3, 6, 12, 24, and 36 months. At each visit, data on weight, comorbidity status, and complications were collected and recorded in a prospective database. A CONSORT flow diagram (Figure 1) was used to track patients through enrollment, allocation, follow-up, and analysis. Reasons for attrition were documented.

Outcomes and Definitions

The primary outcome was the total BAROS score at 3 years post-surgery. The BAROS questionnaire was administered to all patients completing the 3-year follow-up. Scores were calculated based on three main domains: percentage of excess weight loss (%EWL), change in obesity-associated medical conditions, and disease-specific QoL (Moorehead-Ardelt Quality of Life Questionnaire II). Points were deducted for complications and reoperations. Final scores were categorized as: failure (≤1 point), fair (>1 to 3 points), good (>3 to 5 points), very good (>5 to 7 points), and excellent (>7 points) [8].

Secondary outcomes included:

- **Weight Loss:** %EWL was calculated using the formula: [(preoperative weight – weight at follow-up) / (preoperative weight – ideal body weight)] × 100, where ideal body weight corresponds to a BMI of 25 kg/m². Additional measures included BMI, total weight loss (%), and proportion of patients achieving BMI <30 kg/m². Body composition (fat mass, lean mass) was assessed by bioelectrical impedance analysis.
- **Comorbidity Outcomes:** Remission of type 2 diabetes (T2D), hypertension, and dyslipidemia was defined as achieving normal glycemic, blood pressure, and lipid parameters

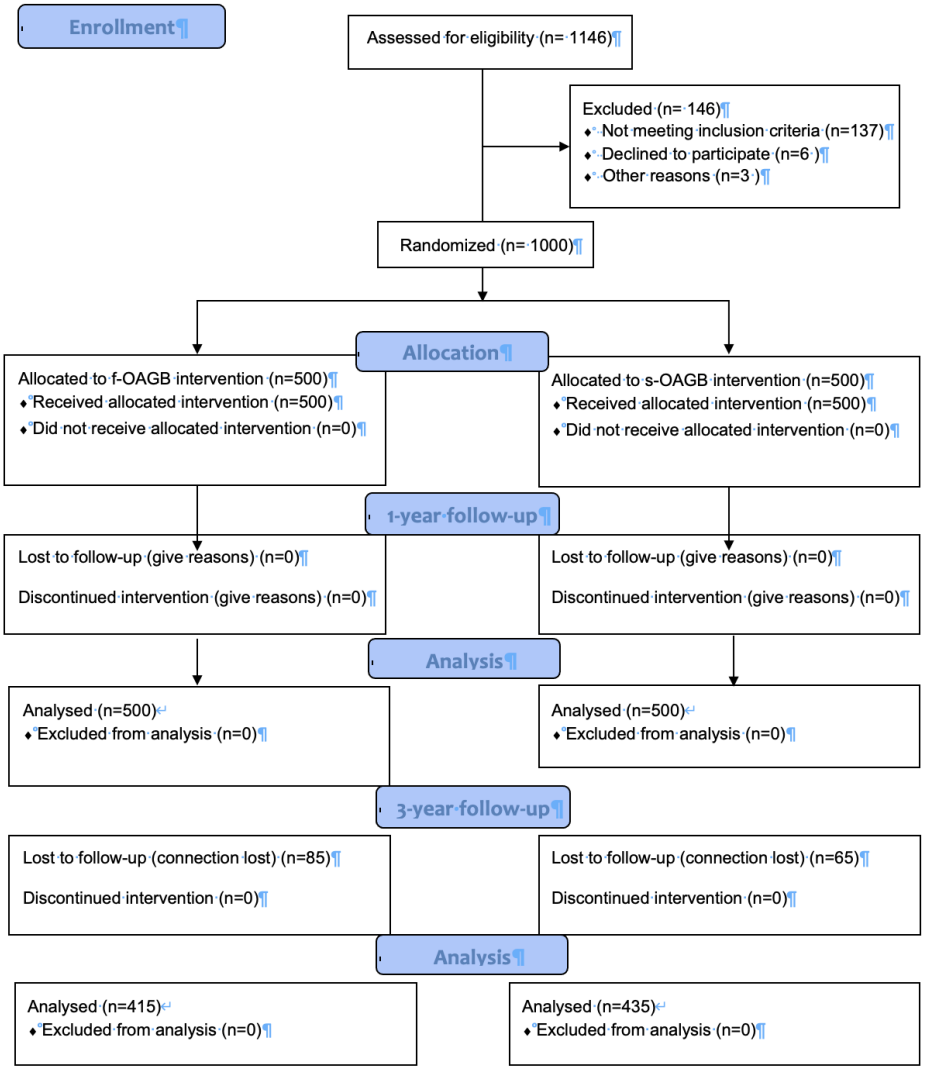


Figure 1 – CONSORT flow diagram

(HbA1c <6.0%, blood pressure <130/85 mmHg, LDL <100 mg/dL, respectively) without the use of related medications. Improvement was defined as a reduction in medication dosage or number of medications required to control the condition. Changes in glycemic control (HbA1c, HOMA-IR), lipid profile (LDL-C, HDL-C, total cholesterol, triglycerides, TC/HDL-C ratio), and blood pressure were recorded at 3, 6, 12, 24, and 36 months.

- **Complications:** Specific complications recorded included dumping syndrome (diagnosed using the Sigstad's clinical score), marginal ulcer (gastrojejunal anastomosis ulcer, GEA) diagnosed by upper endoscopy, bile reflux esophagitis (diagnosed by upper endoscopy with or without symptoms), food intolerance, protein malnutrition, anemia, and vitamin deficiencies.
- **Safety Reporting (Structured):** Adverse events were classified by timing: intraoperative, early postoperative (≤30 days), and late postoperative (>30 days to 36 months). Severity was graded using the Clavien-Dindo classification. Mortality was reported separately for intraoperative, 30-day, and late periods.

Statistical Analysis

Sample size calculation was based on detecting a clinically meaningful difference of 0.5 points in the total BAROS score between groups. Assuming a standard deviation of 2.5, a two-sided alpha of 0.05, and a power of 80%, we estimated a need for 394 patients per group. To account for an anticipated loss to follow-up of 20%, we aimed to recruit 500 patients per group (N=1000).

Statistical analyses were performed using Microsoft Excel for Mac StatPlus MacPro (AnalystSoft Inc., Walnut, CA, USA). Normality of continuous data was assessed using the Kolmogorov-Smirnov test. Continuous variables are presented as mean ± standard deviation (SD) and were compared between groups using an independent samples t-test (or Mann-Whitney U test for non-normal data). The independent samples t-test was appropriate given the large sample size and normal distribution of the data. Categorical variables are presented as counts (n) and percentages (%) and were compared using the Chi-square (χ^2) test or Fisher's exact test. Within-group comparisons were performed using a paired t-test. A two-tailed p-value <0.05 was considered statistically significant. Mean differences with 95% confidence intervals (CIs) were calculated for the primary and key secondary outcomes.

Results

Patient Flow and Baseline Characteristics

Between January 2021 and December 2024, 1146 patients were assessed for eligibility. Of these, 146 were excluded (Figure 1, CONSORT diagram). A total of 1000 patients were randomized: 500 to f-OAGB and 500 to s-OAGB. At the

3-year follow-up mark, 850 patients (85.0%) completed the full evaluation: 415 (83.0%) in the f-OAGB group and 435 (87.0%) in the s-OAGB group. Reasons for attrition in the f-OAGB group were: lost to follow-up (n=31, 6.2%), protocol violation (n=18, 3.6%), withdrawal of consent (n=10, 2.0%), and other reasons (n=26, 5.2%). In the s-OAGB group: lost to follow-up (n=20, 4.0%), protocol violation (n=16, 3.2%), withdrawal of consent (n=12, 2.4%), and other reasons (n=17, 3.4%).

Baseline demographic and clinical characteristics were well-balanced between the two groups, with no statistically significant differences observed (Table 1). Detailed chronic disease status was recorded, including non-alcoholic fatty liver disease (NAFLD) and allergies, which are common in this patient population.

BAROS Outcomes (Primary Outcome)

At 3 years post-surgery, patients in the f-OAGB group had a significantly higher total BAROS score compared to the s-OAGB group (5.59 ± 1.13 vs. 4.57 ± 1.09; mean difference 1.02, 95% CI 0.87–1.17; p<0.001). This corresponds to a "very good" overall outcome for the f-OAGB group versus a "good" outcome for the s-OAGB group.

Analysis of the individual BAROS domains revealed that f-OAGB was superior across all metrics (Table 2). The distribution of BAROS outcome categories also significantly favored the f-OAGB group, with a higher proportion achieving "excellent" and "very good" results.

Table 1 Baseline Demographic and Clinical Characteristics of the Study Population

Characteristic	f-OAGB (n=500)	s-OAGB (n=500)	Test Statistic	p-value
Age (years), mean ± SD	37.0 ± 7.7	40.0 ± 9.2	t=0.45	0.6
Female sex, n (%)	421 (84.2)	407 (81.4)	χ^2=1.37	0.2
Preoperative BMI (kg/m^2), mean ± SD	40.04 ± 5.49	40.49 ± 5.72	t=1.27	0.2
Obesity-Associated Conditions, n (%)				
Dyslipidemia	355 (71.0)	345 (69.0)	χ^2=0.06	0.8
Type 2 Diabetes	105 (21.0)	98 (19.6)	χ^2=0.30	0.58
Arterial Hypertension	334 (66.8)	328 (65.6)	χ^2=0.24	0.62
NAFLD	241 (48.2)	236 (47.2)	χ^2=0.10	0.75
Any allergy	48 (9.6)	52 (10.4)	χ^2=0.17	0.68

BMI, body mass index; NAFLD, non-alcoholic fatty liver disease; SD, standard deviation; f-OAGB, FundoRing one-anastomosis gastric bypass; s-OAGB, standard one-anastomosis gastric bypass.

Weight Loss, Body Composition, and Metabolic Outcomes

Consistent with the BAROS domain scores, f-OAGB resulted in significantly greater weight loss and better metabolic control at 3 years compared to s-OAGB (Table 3). Patients in the f-OAGB group achieved a mean %EWL of 84.3% vs. 76.8% in the s-OAGB group (p<0.001). A higher proportion of patients in the f-OAGB group achieved a BMI <30 kg/m² (94.2% vs. 86.7%, p=0.002). Body composition analysis showed that f-OAGB patients lost more fat mass while preserving lean mass similarly to the s-OAGB group.

Remission rates for T2D, hypertension, and dyslipidemia were consistently higher in the f-OAGB group, with marked improvements in glycemic control and lipid profile (Table 3, see the next page).

Complications and Safety (Structured Analysis)

There was no intraoperative mortality in either group. No 30-day mortality occurred. No late deaths were observed during the 3-year follow-up period in either group (overall mortality 0%)

The incidence of key complications is summarized in Table 4 (see the next page). Using the Clavien-Dindo classification, most complications were Grade I-II. Grade IIIb complications (requiring surgical or endoscopic intervention) occurred in 0.5% of the f-OAGB group (1 case of anastomotic stricture requiring endoscopic dilation) vs. 2.1% of the s-OAGB group (5 cases of bile reflux requiring endoscopic treatment, 4 marginal ulcers requiring endoscopic therapy and prolonged PPI; p=0.04).

Bile Reflux: f-OAGB was associated with a dramatic and significant reduction in the incidence of bile reflux at all time points. At 3 years, only 1 patient (0.2%) in the f-OAGB group presented with bile reflux compared to 41 patients (9.4%) in the s-OAGB group (p<0.001).

Marginal Ulcer: The rate of marginal ulceration was also significantly lower in the f-OAGB group. At 3 years, only 1 patient (0.2%) in the f-OAGB group was diagnosed with a marginal ulcer versus 15 patients (3.4%) in the s-OAGB group (p<0.001).

Table 2 Three-Year BAROS Outcomes

BAROS Component	f-OAGB (n=415)	s-OAGB (n=435)	Mean Difference (95% CI)	p-value
Total BAROS Score, mean ± SD	**5.59 ± 1.13**	**4.57 ± 1.09**	**1.02 (0.87 to 1.17)**	**<0.001**
Weight Loss Domain Score, mean ± SD	2.36 ± 0.64	2.19 ± 0.60	0.17 (0.09 to 0.25)	<0.001
Medical Conditions Score, mean ± SD	1.34 ± 0.67	0.96 ± 0.69	0.38 (0.29 to 0.47)	<0.001
Quality of Life Score, mean ± SD	1.87 ± 0.71	1.41 ± 0.55	0.46 (0.37 to 0.55)	<0.001

BAROS, Bariatric Analysis and Reporting Outcome System; CI, confidence interval; SD, standard deviation; f-OAGB, FundoRing one-anastomosis gastric bypass; s-OAGB, standard one-anastomosis gastric bypass.

Table 3 Clinical Outcomes at 3-Year Follow-up

Outcome	FundoRing OAGB (n=415)	Standard OAGB (n=435)	Difference (95% CI)	p-value
Primary Effectiveness Measure				
Patients achieving BMI <30 kg/m^2, n (%)	391 (94.2)	377 (86.7)	7.5% (3.7 to 11.3%)	0.002
Secondary Anthropometric Measures				
BMI at 3 years (kg/m^2)	25.92 ± 2.03	27.53 ± 3.25	-1.61 (-2.00 to -1.22)	<0.001
Change in BMI from baseline	-14.12 ± 5.83	-12.96 ± 6.01	-1.16 (-1.98 to -0.34)	0.006
Excess weight loss (%)	84.3 ± 12.1	76.8 ± 14.3	7.5 (5.6 to 9.4)	<0.001
Total weight loss (%)	35.2 ± 5.1	32.1 ± 5.8	3.1 (2.3 to 3.9)	<0.001
Lipid Profile (mg/dL)				
LDL-C	73.06 ± 9.2	77.18 ± 14.3	-4.12 (-5.75 to -2.49)	0.001
HDL-C	65.6 ± 4.9	64.8 ± 4.8	0.8 (0.2 to 1.4)	0.015
Total Cholesterol	140.46 ± 23.21	154.24 ± 26.71	-13.78 (-17.15 to -10.41)	<0.001
Triglycerides	88.03 ± 3.63	92.73 ± 3.72	-4.70 (-5.20 to -4.20)	<0.001
TC/HDL-C ratio	2.74 ± 0.69	2.93 ± 0.67	-0.19 (-0.28 to -0.10)	<0.001
Glycemic Control				
HbA1c (%)	5.25 ± 1.4	5.5 ± 1.5	-0.25 (-0.42 to -0.08)	0.004
HOMA-IR	1.8 ± 1.0	2.23 ± 1.24	-0.43 (-0.58 to -0.28)	<0.001
Type 2 Diabetes remission, n (%)	104/105 (99.0)	96/98 (97.9)	1.1% (-2.4 to 4.6%)	0.52
Blood Pressure				
Arterial hypertension resolved, n (%)	331/334 (99.1)	324/328 (98.7)	0.4% (-1.4 to 2.2%)	0.68

BMI, body mass index; CI, confidence interval; HbA1c, glycated hemoglobin; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, homeostatic model assessment for insulin resistance; LDL-C, low-density lipoprotein cholesterol; TC/HDL-C, total cholesterol to HDL-C ratio.

Table 4 Late Postoperative Complications at 3-Year Follow-up

Complication	FundoRing OAGB (n=415)	Standard OAGB (n=435)	p-value
Dumping syndrome, n (%)	57 (13.7)	159 (36.5)	<0.001
Marginal ulcer (GEA), n (%)	1 (0.2)	15 (3.4)	<0.001
Bile reflux esophagitis, n (%)	1 (0.2)	41 (9.4)	<0.001
Food intolerance, n (%)	2 (0.5)	3 (0.7)	0.69
Protein malnutrition, n (%)	19 (4.6)	16 (3.7)	0.5
Anemia, n (%)	88 (21.2)	83 (19.1)	0.44
Thiamine (B1) deficiency, n (%)	18 (4.3)	17 (3.9)	0.75

GEA, gastrojejunal anastomosis; f-OAGB, FundoRing one-anastomosis gastric bypass; s-OAGB, standard one-anastomosis gastric bypass.

Dumping Syndrome: The incidence of dumping syndrome was significantly lower in the f-OAGB group at both the 1-year and 3-year follow-up (13.7% vs. 36.5% at 3 years, p<0.001).

Food Intolerance: Early postoperative food intolerance (within the first year) was more common in the f-OAGB group (5.4% vs. 2.4%, p=0.007). However, this difference resolved over time, and by the 3-year mark, rates of food intolerance were similar between the groups (p=0.69).

Nutritional Complications: Rates of protein malnutrition, anemia, and thiamine deficiency were similar between groups at 3 years, indicating that the FundoRing modification did not increase the risk of nutritional deficiencies.

Discussion

This prospective randomized trial demonstrates that the FundoRing modification of OAGB (f-OAGB) yields significantly better 3-year outcomes compared to the standard OAGB (s-OAGB) in patients with moderate to severe obesity. The primary endpoint, the total BAROS score, was more than one full point higher in the f-OAGB group, representing a clinically meaningful shift from a "good" to a "very good" overall result. This superior performance was driven by enhanced weight loss (84.3% EWL vs. 76.8%), improved remission of medical conditions, better QoL, and a marked reduction in key procedure-specific complications such as bile reflux, marginal ulcers, and dumping syndrome.

The most striking finding of our study is the profound reduction in bile reflux and marginal ulcers with the f-OAGB technique. Bile reflux is a well-documented concern after OAGB, with reported rates varying widely but often affecting a significant minority of patients. The FundoRing procedure, by creating a 360° fundoplication of the bypassed stomach around the anastomosis, appears to effectively recreate an antireflux mechanism. This anatomical barrier likely prevents the ascent of bile into the esophagus and potentially protects the vulnerable gastrojejunal anastomosis from constant exposure to biliopancreatic juices, thereby reducing the risk of marginal ulcer formation and preventing dilatation of the gastric pouch [13]. The near-elimination of these complications in our study is a significant safety advantage.

The higher BAROS QoL scores in the f-OAGB group

are likely a direct consequence of this improved safety profile. Chronic symptoms of bile reflux (heartburn, regurgitation, esophagitis) and dumping syndrome can severely impact a patient's daily life, dietary choices, and overall satisfaction with surgery. By minimizing these adverse effects, f-OAGB allows patients to enjoy the full metabolic and weight loss benefits of the procedure without the burden of debilitating symptoms. The early increase in food intolerance in the f-OAGB group is not unexpected, as the fundoplication may create a degree of early restriction or altered gastric emptying that resolves with time and dietary adaptation.

The structured safety analysis revealed that f-OAGB not only reduces common complications but also lowers the rate of Grade IIIb events requiring re-intervention (0.5% vs. 2.1%, p=0.04).

Patient characteristics, including the presence of chronic diseases such as NAFLD and allergies, were well-balanced between groups, supporting the internal validity of the comparison. The high rates of T2D (99.0% vs. 97.9%) and hypertension (99.1% vs. 98.7%) remission in both groups reflect the excellent metabolic efficacy of OAGB-based procedures.

The observed differences in BAROS scores, while showing some overlap in standard deviations, are statistically and clinically significant, as confirmed by non-overlapping 95% confidence intervals for the mean differences (1.02, 95% CI 0.87–1.17). The independent samples t-test was appropriate given the large sample size (n=850 at follow-up) and the normal distribution of the data.

Our findings align with and extend the literature on OAGB outcomes. While studies like that of Madani et al. [9] have shown excellent 5-year BAROS results for standard OAGB, our data suggest that the FundoRing modification can further elevate these outcomes. The %EWL and comorbidity remission rates in our s-OAGB group are comparable to large published series, lending external validity to our findings. However, the incremental benefit of f-OAGB over s-OAGB, particularly in complication reduction, positions it as a potentially superior first-line option when OAGB is contemplated.

Strengths and Limitations

This study has several notable strengths. First, its **randomized controlled design** represents the highest level of evidence for comparing surgical techniques. Second, the **large sample size** (N=1000) and high 3-year follow-up rate (85%) provide robust and reliable data. Third, the use of the **BAROS score as a primary outcome** ensures that our assessment is patient-centered and holistic, encompassing not just weight loss but also QoL and the impact of complications. Fourth, all procedures were performed by a single high-volume surgeon, minimizing technical variability as a confounding factor. Fifth, the structured safety analysis using Clavien-Dindo classification and detailed reporting of attrition reasons enhance transparency and reproducibility.

However, the study also has limitations. As **a single-center study** from a high-volume expert center, the generalizability of these results to other settings and surgeons with varying experience levels may be limited. **The follow-up of 3 years,** while adequate for assessing mid-term outcomes, is insufficient to capture very long-term complications or weight regain. We also excluded certain high-risk groups, such as patients with **super-obesity (BMI >50 kg/m²)** and those with **severe preoperative reflux (LA Grade C/D),** who might be at higher risk for complications or may not be ideal candidates for this procedure. Finally, while outcome assessors were blinded, **patients and the operating surgeon were not blinded due to the nature of the intervention.**

Future Directions and Implications

The results of this trial have important implications for clinical practice. For surgeons performing OAGB, adopting the FundoRing modification may offer a way to significantly improve patient outcomes and reduce the burden of troublesome complications. For patients, it provides a surgical option with a more favorable risk-benefit profile. Future research should focus **on longer-term follow-up (5-10 years)** of this cohort to confirm the durability of weight loss and the persistence of the antireflux effect. Furthermore, **multicenter trials** are needed to assess the reproducibility of these results across different surgical teams and patient populations. Comparative studies with RYGB and **RYGB with fundoplication** would also help position f-OAGB within the broader armamentarium of metabolic procedures.

Conclusion

Based on the 3-year results of this randomized trial, FundoRing OAGB demonstrates superior outcomes compared to standard OAGB for patients with moderate to severe obesity. The f-OAGB technique results in higher BAROS scores (5.59 vs. 4.57), greater weight loss (84.3% EWL vs. 76.8%), better comorbidity resolution, and enhanced quality of life. These benefits are accompanied by a significant reduction in bile reflux (0.2% vs. 9.4%), marginal ulcers (0.2% vs. 3.4%), and dumping syndrome (13.7% vs. 36.5%), with no increase in major safety events or mortality. F-OAGB represents a safe and effective technical advancement in metabolic surgery.

Author Contributions: Conceptualization, O. O.; investigation, G. D.; verification, V. R., B. Y., Zh. D.; writing – original draft preparation, G. D., B. Y., Zh. D.; writing – review and editing, O. O., G. D. The authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: The authors gratefully acknowledge the international observers for their oversight and all participating experts for their invaluable contribution.

Funding: The work was carried out within the framework of grant funding from the Ministry of Education and Science of the Republic of Kazakhstan for 2024-2026, No. AR23490186.

Data availability statement: All raw data underlying the findings of this study can be obtained from the corresponding author upon reasonable request.

Artificial Intelligence (AI) Disclosure Statement: Artificial intelligence was used to translate and edit the text to improve its readability in English, ensure efficient work under human supervision.

References

1. Zhou XD, Chen QF, Yang W, Zuluaga M, Targher G, Byrne CD, Valenti L, Luo F, Katsouras CS, Thaher O, Misra A, Ataya K, Oviedo RJ, Pik-Shan Kong A, Alswat K, Lonardo A, Wong YJ, Abu-Abeid A, Al Momani H, Ali A, Molina GA, Szepietowski O, Jumaev NA, Kızılkaya MC, Viveiros O, Toro-Huamanchumo CJ, Yen Kok KY, Ospanov O, Abbas SI, Robertson AG, Fouad Y, Mantzoros CS, Zhang H, Méndez-Sánchez N, Sookoian S, Chan WK, Treeprasertsuk S, Adams L, Ocama P, Ryan JD, Perera N, Sharara AI, Al-Busafi SA, Opio CK, Garcia M, Lim-Loo MC, Ruiz-Úcar E, Prasad A, Casajoana A, Abdelbaki TN, Zheng MH. Burden of disease attributable to high body mass index: an analysis of data from the Global Burden of Disease Study 2021. *EClinicalMedicine*. 2024; 76: 102848. https://doi.org/10.1016/j.eclinm.2024.102848. Erratum in: *EClinicalMedicine*. 2024; 78: 102958. https://doi.org/10.1016/j.eclinm.2024.102958.
2. Noel P, Cazeres C, Lutfi RE, Nocca D. A Comprehensive Review of Sleeve Gastrectomy, Roux-en-Y Gastric Bypass, One-Anastomosis Gastric Bypass, Duodenal Switch, and SADI-S: Very Long-Term Outcomes at 10 Years and Beyond. *J Laparoendosc Adv Surg Tech A*. 2026: 10926429261419379. https://doi.org/10.1177/10926429261419379.
3. Gulinac M, Miteva DG, Peshevska-Sekulovska M, Novakov IP, Antovic S, Peruhova M, Snegarova V, Kabakchieva P, Assyov Y, Vasilev G, Sekulovski M, Lazova S, Tomov L, Velikova T. Long-term effectiveness, outcomes and complications of bariatric surgery. *World J Clin Cases*. 2023; 11(19): 4504-4512. https://doi.org/10.12998/wjcc.v11.i19.4504.
4. Kermansaravi M, Cohen RV, Shikora SA, Chiappetta S, Parmar C, Mahawar K, Aminian A, Dillemans B, Monami M, Belluzzi A, Nimeri A, Angrisani L, Di Lorenzo N, Prager G, Petry TB, Kassir R, De Luca M. Same-day Discharge Metabolic and Bariatric Surgery: a GRADE-based International Federation for the Surgery and Other Therapies for Obesity (IFSO) Position Statement. *Obes Surg*. 2026. https://doi.org/10.1007/s11695-026-08555-y.
5. Maselli DB, Hoff AC, Kucera A, Waseem A, Wooley C, Donnangelo LL, Coan B, McGowan CE. Endoscopic revision of one-anastomosis gastric bypass (ER-OAGB) for weight recurrence: a case series of 17 adults. *Ther Adv Gastrointest Endosc*. 2023; 16: 26317745231210120. https://doi.org/10.1177/26317745231210120.
6. Ospanov O, Zharov N, Yelembayev B, Duysenov G, Volchkova I, Sultanov K, Mustafin A. A Three-Arm Randomized Controlled Trial of Primary One-Anastomosis Gastric Bypass: With FundoRing or Nissen Fundoplications vs. without Fundoplication for the Treatment of Obesity and Gastroesophageal Reflux Disease. *Medicina (Kaunas)*. 2024; 60(3): 405. https://doi.org/10.3390/medicina60030405.
7. Ospanov O. The Surgical Technique of Primary Modified Fundoplication Using the Excluded Stomach with Simultaneous Gastric Bypass. *Obes Surg*. 2023; 33(4): 1311-1313. https://doi.org/10.1007/s11695-023-06505-6.
8. Oria HE, Moorehead MK. Updated bariatric analysis and reporting outcome system (BAROS). Surg Obes Relat Dis. 2009; 5(1): 60-66.
9. Madani S, Shahsavan M, Pazouki A, Setarehdan SA, Yarigholi F, Eghbali F, Shahmiri SS, Kermansaravi M. Five-Year BAROS Score Outcomes for Roux-en-Y Gastric Bypass, One Anastomosis Gastric Bypass, and Sleeve Gastrectomy: a Comparative Study. *Obes Surg*. 2024; 34(2): 487-493. https://doi.org/10.1007/s11695-023-07015-1.
10. . Carbajo MA, Luque-de-León E, Jiménez JM, Ortiz-de-Solórzano J, Pérez-Miranda M, Castro-Alija MJ. Laparoscopic One-Anastomosis Gastric Bypass: Technique, Results, and Long-Term Follow-Up in 1200 Patients. *Obes Surg*. 2017; 27: 1153–1167. https://doi.org/10.1007/s11695-016-2428-1.
11. Ospanov O. Standardization of surgical technique of the fundoring method for laparoscopic gastric bypass. *Sci Rep*. 2025; 15(1): 44941. https://doi.org/10.1038/s41598-025-29153-5.
12. Marchetti M, Pan TL, Delfrati S, Gill S, Yates E, Meschini T, Vilches JC, Santía MC, Ramirez PT, Nelson G. Enhanced Recovery After Surgery (ERAS) in gynecologic surgery: hot topic debates at the 2025 ERAS World Congress. *Int J Gynecol Cancer*. 2026; 36(4): 104552. https://doi.org/10.1016/j.ijgc.2026.104552.
13. Ospanov O, Yelembayev B, Buchwald JN, Sultanov K, Rakhmetov M, Duysenov G, Tuleuov B. Comparative endoscopic ultrasound assessment of the gastric pouch after FundoRing-OAGB and OAGB. *BMC Surg*. 2025; 26(1): 23. https://doi.org/10.1186/s12893-025-03386-7.

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/10.23950/jcmk/18500

Identification of Hub Genes Associated with Bile Duct Cancer Using Integrated Gene Expression Data with Protein-Protein Interaction Network

Narmadha Ramasamy[1], Pogala Hema Vardhan[1], Kannan Muthu[1]

[1]Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India

Received: 2025-11-01.
Accepted: 2026-04-16.

J Clin Med Kaz 2026; 23(3): 11-23

Corresponding author:
Kannan Muthu.
E-mail: kannan@bicpu.edu.in.
ORCID: 0000-0002-9250-2379.

ABSTRACT

Introduction: Bile duct cancer (cholangiocarcinoma) is an uncommon but highly aggressive malignancy characterized by late diagnosis and poor prognosis. Molecular profiling is essential to understand its pathogenesis and identify biomarkers for early detection and targeted therapy. This study aimed to identify key regulatory genes and molecular pathways associated with bile duct cancer using an integrated bioinformatics approach.

Methods: Gene expression datasets (GSE131027 and GSE107754) were retrieved from the NCBI Gene Expression Omnibus (GEO) database, and samples annotated as cholangiocarcinoma were selected for analysis. Differentially expressed genes were identified using GEO2R. Protein–protein interaction networks were constructed using the STRING database and analyzed in Cytoscape with CytoHubba for hub gene identification.

Results: A total of over 1,000 differentially expressed genes (DEGs) were identified across the analyzed datasets. Network and functional enrichment analyses highlighted seven major hub genes—TP53, HIST1H3F, H2AFZ, FOS, POLR2B, CAV1, and SMAD3 that occupied central positions within the protein–protein interaction networks. These hub genes were associated with pathways related to transcriptional regulation, oxidative and nitrosative stress responses, RNA processing, and post-transcriptional gene regulation. Gene Ontology and KEGG pathway analyses further indicated enrichment of biological processes previously implicated in cancer-related molecular mechanisms, supporting their relevance in cholangiocarcinoma.

Conclusion: This integrative bioinformatics study identified critical hub genes and molecular pathways that may serve as potential biomarkers and therapeutic targets for bile duct cancer. Further validation and molecular docking studies could facilitate the development of targeted drugs and improve treatment outcomes.

Keywords: Cholangiocarcinoma, biomarkers, Hub genes, therapeutics, gene expression.

Introduction

Bile duct cancer (BDC), or cholangiocarcinoma, is a rare but highly aggressive malignancy arising from the epithelial lining of the bile ducts, which transport bile from the liver to the small intestine. Despite its low incidence, BDC presents significant clinical challenges due to late diagnosis, early local invasion, and poor overall prognosis. Anatomically, cholangiocarcinoma is classified into intrahepatic (iCCA) that originates within the liver parenchyma, perihilar (pCCA) arises at or near the hepatic duct bifurcation, and distal extrahepatic (dCCA) is where the tumor occurs closer to the pancreas and ampulla of vater. This classification is critical for determining surgical strategies, prognostic assessment, and adjuvant therapy planning [1–3]. The Japanese Society of Hepato-Biliary-Pancreatic Surgery (JSHBPS)

provides standardized definitions and classifications for hepatobiliary malignancies, including BDC, to guide clinical management and optimize outcomes [4–6]. The pathogenesis of BDC is multifactorial. Chronic inflammation of the biliary epithelium, such as from primary sclerosing cholangitis, bile duct strictures, and liver fluke infestations, predisposes to malignant transformation. Viral hepatitis B and C infections are significant contributors, particularly in East Asia, where hepatitis prevalence is high[7–9]. Environmental toxins and genetic mutations, including alterations in tumor suppressor genes and oncogenes also been involved [10,11]. BDC predominantly affects older adults, with a higher incidence in males and geographic regions such as Southeast Asia exhibiting elevated prevalence rates.

Clinical presentation is often nonspecific that leads to delayed diagnosis in BDC. Patients commonly present with obstructive jaundice, pruritus, weight loss, and cholangitis. Imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and endoscopic retrograde cholangiopancreatography (ERCP), assist in tumor localization and staging, yet precise assessment of tumor spread along the bile duct remains difficult [12,13]. Cytological evaluation of ERCP brushings is highly specific but has limited sensitivity (23–56%) due to low cellularity and reactive changes and the most of the diagnosis frequently requires surgical exploration [13,14]. Surgical resection remains the primary curative therapy for BDC. The tumor's location and degree of dissemination dictate the kind and scope of the surgery. iCCA is managed with hepatic resection aiming for negative margins. Perihilar tumors, particularly Bismuth type III and IV lesions, often require major hepatectomy with caudate lobectomy, extrahepatic bile duct excision, and regional lymphadenectomy. Distal cholangiocarcinoma is primarily treated with pancreaticoduodenectomy (PD), whereas selected middle-third common bile duct tumors may undergo segmental bile duct resection to avoid pancreatic transection while achieving oncological safe margins. The proximity of tumors to the hepatic artery, portal vein, and adjacent liver parenchyma complicates surgical planning, and extensive resections are associated with increased postoperative morbidity and mortality [15–17]. Long-term results are still limited despite advancements in surgery. In spite of the aggressive tumor and early recurrence with five-year survival rate after extrahepatic cholangiocarcinoma excision seldom exceeds 50%. Prognostic factors include depth of the tumor invasion, lymph node metastases, perineural invasion, histological differentiation, and resection margin status. Serum carbohydrate antigen 19-9 (CA19-9) serves as an adjunct biomarker, although it is insufficient for early detection [18,19]. Preoperative biliary drainage and stenting are frequently required in obstructive cases but are associated with higher postoperative infectious complications, emphasizing careful patient selection.

Recent advances highlight the significance of molecular profiling in understanding BDC pathogenesis and guiding therapeutic strategies. Analyses of gene expression networks reveal dysregulation of genes involved in cell cycle control, apoptosis, and signaling pathways. Bioinformatics tools such as Cytoscape and datasets from NCBI Gene Expression Omnibus (GEO) facilitate visualization of gene interactions, identification of key regulatory genes, and prediction of survival outcomes. Such analyses can inform targeted drug development and stratify patients based on prognostic risk. Identifying consistently dysregulated genes enables precision medicine approaches and offers potential biomarkers for early diagnosis and therapeutic intervention [20–23]. The concept of “field cancerization” is also relevant in BDC. Chronic bile duct injury and exposure to carcinogens may induce widespread epithelial alterations, predisposing patients to multifocal malignancies. This underscores the importance of thorough intraoperative assessment, histological examination of resection margins, and long-term surveillance for recurrence. Population-level studies utilizing administrative databases, such as the Korean National Health Insurance Service, have provided valuable insights into BDC risk factors, long-term prognosis, and treatment outcomes. These data complement molecular studies by correlating biomarker expression with clinical outcomes.

Advances in high-throughput transcriptomic technologies and computational biology have enabled systematic exploration of the molecular landscape of bile duct cancer. Publicly available gene expression datasets deposited in the NCBI Gene Expression Omnibus (GEO) provide valuable, large-scale data that allow unbiased identification of differentially expressed genes across independent patient cohorts, thereby enhancing the reliability and reproducibility of findings [24,25]. Integrative analysis of multiple GEO datasets helps overcome sample heterogeneity and platform-specific bias commonly encountered in individual studies. Furthermore, protein–protein interaction (PPI) network analysis offers a systems-level perspective by elucidating functional relationships among dysregulated genes and identifying coordinated molecular modules involved in tumorigenesis [26–28]. Within these networks, hub genes defined by their high topological connectivity often represent key regulatory elements that control critical biological processes such as cell cycle regulation, apoptosis, and stress responses [29,30]. Therefore, combining GEO-based transcriptomic analysis with PPI network construction and hub gene identification provides a powerful bioinformatics framework for uncovering potential biomarkers and therapeutic targets in cholangiocarcinoma, bridging clinical observations with molecular mechanisms.

Bile duct cancer is a rare, highly aggressive malignancy arising from complex interactions between chronic inflammation, viral infections, environmental factors, and genetic alterations. Anatomical classification into intrahepatic, perihilar, and distal types guides surgical management and prognosis. While curative resection remains the cornerstone of treatment, survival is limited by aggressive tumor biology and anatomical complexity. Integrating molecular profiling, gene network analysis, and targeted therapy strategies offers promising avenues for improving patient outcomes. Multidisciplinary management—including surgery, molecular diagnostics, supportive care, and palliative interventions—is essential for optimizing survival and quality of life in patients with this challenging malignancy.

Despite advances in imaging and surgical management, the molecular mechanisms underlying cholangiocarcinoma remain incompletely understood, and reliable molecular biomarkers for early diagnosis are limited. Publicly available transcriptomic datasets from the Gene Expression Omnibus (GEO) provide valuable resources for systematically investigating gene expression alterations associated with bile duct cancer. Integrative bioinformatics approaches that combine differential gene expression analysis with protein–protein interaction network construction and hub gene identification enable the discovery of key regulatory genes and biological pathways involved in tumor progression. In this study, we applied a systems biology–based workflow to identify hub genes and functionally enriched pathways associated with cholangiocarcinoma, aiming to provide insights into disease mechanisms and potential molecular targets.

Methods

Bibliographic Search

Gene expression datasets GSE131027 and GSE107754 were obtained from the NCBI Gene Expression Omnibus (GEO) database. GSE131027 is a pan-cancer microarray dataset generated using the Affymetrix Human Genome U133 Plus 2.0 Array, whereas GSE107754 represents an independent dataset containing transcriptomic profiles relevant to biliary tract malignancies. Samples annotated as cholangiocarcinoma were selected from both datasets based on GEO metadata for downstream analysis.

Differential Gene Expression Analysis

Differential gene expression analysis was performed using GEO2R, an interactive web-based tool provided by the NCBI Gene Expression Omnibus. GEO2R applies the limma statistical framework to identify differentially expressed genes between defined sample groups. Samples annotated as cholangiocarcinoma / bile duct cancer were selected based on GEO metadata and compared with corresponding reference samples. Genes with an absolute log2 fold change ($|\log2FC| > 1$ and an adjusted p-value < 0.05 (Benjamini–Hochberg correction, as implemented in GEO2R) were considered statistically significant and used for downstream protein–protein interaction and hub gene analyses

Formation of STRING network using the STRING network analysis and Cytoscape

Protein–protein interaction (PPI) networks were constructed using the STRING database to identify functional associations among differentially expressed genes. The interaction networks were visualized using Cytoscape through the STRINGApp plugin. Nodes represent proteins, and edges represent predicted or experimentally validated protein–protein interactions as curated by STRING. This network-based approach was used to explore molecular interaction patterns relevant to cholangiocarcinoma

Identification of Hub genes using Cytohubba

Hub genes were identified from the STRING-derived protein–protein interaction network using the CytoHubba plugin in Cytoscape. CytoHubba ranks nodes based on topological properties of the network to identify key regulatory genes. In this study, hub genes were ranked using the Degree, DMNC (Density of Maximum Neighborhood Component), and Bottleneck algorithms. Genes consistently ranked among the top candidates were considered hub genes and selected for subsequent functional enrichment analyses.

ClueGo and BINGO

ClueGo is a tool that generates the first binary gene term matrix with their selected associated terms and genes. By looking at this matrix, there is a similarity between the terms to determine the association between the pathways, using the kappa statistics to assess the strength of the terms. Finally, the created networks represent the network using the kappa statistical score level. Using the custom method, the kappa score will automatically adjust based on the positive scale from 0 to 1, which reflects the network connectivity. By gene clustering, ClueGO allows the visualization of the groups based on their network to cluster distribution over the terms. There are two main modes for selecting the set of genes that can be functionally profiled in the default mode, like choosing the nodes from the Cytoscape network or using the other plugins using the MCODE or BINGO in Cytoscape. They are compiled from different sources for a set of upregulated genes in microarray experiments displayed in STRING network analysis. BINGO shows the relevant GO annotations, propagates them through the GO hierarchy, and shows a similar set of parental categories.

Software and Reproducibility

Differential gene expression analysis was performed using GEO2R, a publicly accessible web-based tool provided by the NCBI Gene Expression Omnibus, which applies the limma statistical framework. Protein–protein interaction networks were generated using the STRING database (version 12.0). Network visualization and analysis were conducted using Cytoscape (version 3.10.4). Hub gene identification was carried out using the CytoHubba plugin, and functional enrichment analyses were performed using the ClueGO and BiNGO plugins with default parameters. All datasets used in this study are publicly available, and the analysis can be reproduced by applying the same sample grouping and threshold criteria described above.

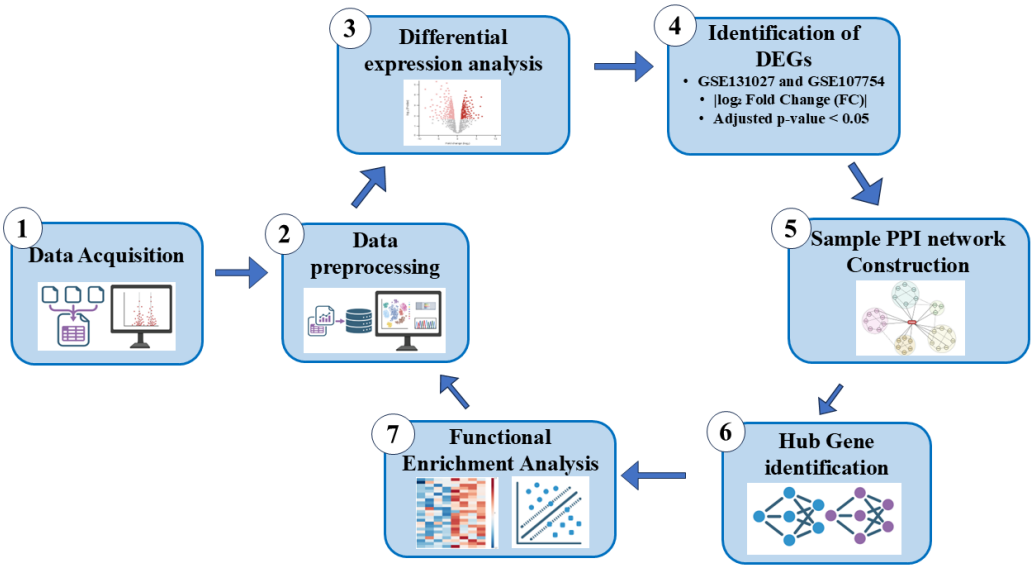


Figure 1 – Integrated Bioinformatics workflow for Cholangiocarcinoma analysis

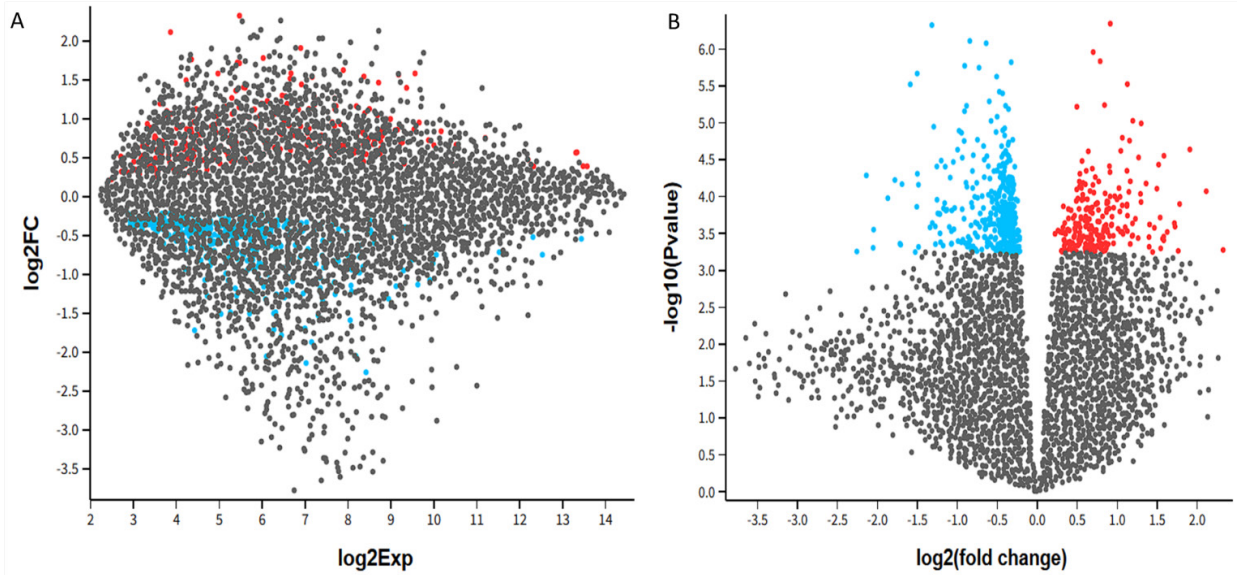


Figure 2 – (a) Mean-difference (MD) plot showing differentially expressed genes (DEGs) between bile duct cancer and control samples. Upregulated genes are indicated in red, while downregulated genes are shown in blue. (b) Volcano plot illustrating the distribution of DEGs based on $\log_2$ fold change and statistical significance. Red dots represent significantly upregulated genes, and blue dots represent significantly downregulated genes, as defined by $|\log_2 FC| > 1.0$ and adjusted p-value < 0.05.

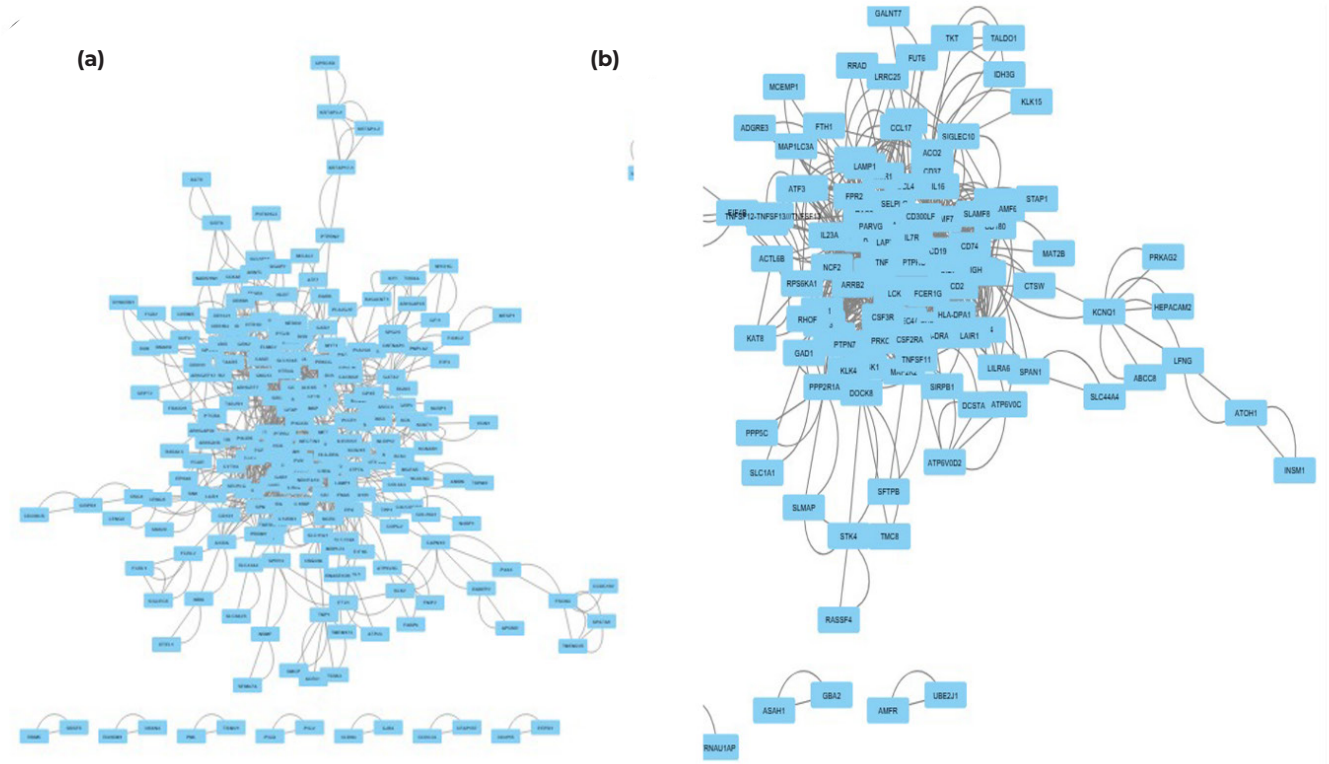


Figure 3 – (a) Protein–protein interaction (PPI) network of upregulated differentially expressed genes (DEGs) from Set 1 (GSE131027) constructed using the STRING database and visualized in Cytoscape. Nodes represent proteins encoded by upregulated genes, while edges indicate predicted or experimentally validated interactions based on STRING confidence scores. (b) Protein–protein interaction (PPI) network of downregulated DEGs from Set 1 (GSE131027) generated using the same STRING-based approach. The network highlights functional connectivity and interaction patterns among downregulated genes associated with bile duct cancer.

Schematic overview of the integrative bioinformatics workflow used to identify differentially expressed genes, hub genes, and enriched pathways in cholangiocarcinoma was shown in Figure 1.

Results

Selection of samples from the selected GEO datasets

Through the NCBI GEO database, they selected two different datasets from the bile duct cancer. After analyzing the STRING data and the samples, they collected the number of samples from sets 1 and 2 from the database.

Identification of DEG

Differential expression analysis revealed substantial transcriptomic alterations across the two analyzed GEO datasets. In both datasets, more than 1,000 genes were identified as significantly differentially expressed, including both upregulated and downregulated genes.In dataset set 1, the mean-difference plot (Figure 2a) illustrates a wide distribution of gene expression

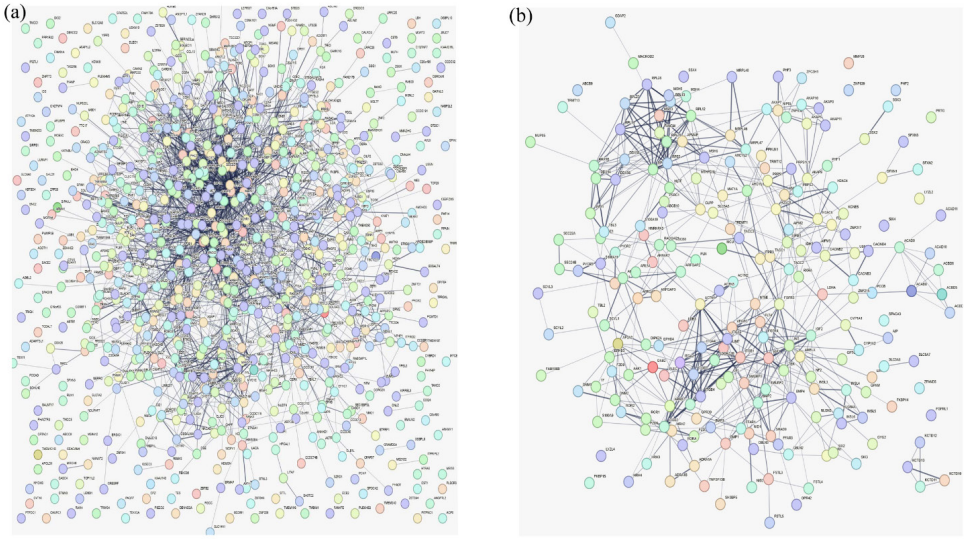


Figure 4 – (a) Protein–protein interaction (PPI) network of upregulated differentially expressed genes (DEGs) from Set 2 (GSE107754) constructed using the STRING database and visualized in Cytoscape. Nodes represent proteins encoded by upregulated genes, while edges denote known and predicted protein–protein interactions based on STRING confidence scores. (b) Protein–protein interaction (PPI) network of downregulated DEGs from Set 2 (GSE107754) generated using the same STRING-based approach. The network illustrates interaction density and functional organization among downregulated genes associated with bile duct cancer.

Table 1 Top 15 Upregulated and Downregulated Genes in Set 1 and Set 2 Based on DEG Analysis

Set 1 (Upregulated)				**Set 1 (Down-regulated)**			
S.NO	NAME	RANK	SCORE	S.NO	NAME	RANK	SCORE
1	TP53	1	138	1	PTPRC	1	88
2	HIST1H3F	2	84	2	TNF	2	76
3	H2AFZ	3	70	3	CD19	3	60
4	FOS	4	68	4	ACTB	4	58
5	POLR2B	5	52	5	FCER1G	4	58
6	CAV1	6	48	6	LCK	6	46
7	CS	7	46	7	IL7R	7	44
8	FEN1	8	44	8	CD2	7	44
9	NUP98	9	42	9	CD74	9	42
10	SMAD3	10	40	10	RAC2	10	38
11	RAN	10	40	11	CCL4	11	34
12	TPI1	12	38	12	HLA-DRA	11	34
13	RAC1	12	38	13	CCRL2	13	30
14	RAE1	14	36	14	CD53	13	30
15	EIF4A3	15	34	15	LAPTM5	15	26
Set 2 (Upregulated)				Set 2 (Down-regulated)			
S.NO	NAME	RANK	SCORE	S.NO	NAME	RANK	SCORE
1	INS	1	94	1	VCAM1	1	10735
2	PTEN	2	84	2	HGF	2	10443
3	CDKN2A	3	54	3	INS	3	9652
4	IGF1	3	54	4	CCR2	4	8478
5	VCAM1	5	50	5	IGF1	5	8307
6	CDK4	5	50	6	ITCB1	6	8119
7	CCR2	7	48	7	NT5E	7	7383
8	HGF	7	48	8	TEK	8	5906
9	RPS2	9	44	9	CD80	9	3425
10	RPL26	9	44	10	PTEN	10	2680
11	RARA	9	44	11	BDNF	11	1745
12	BDNF	9	44	12	FLT4	12	1563
13	NT5E	9	44	13	CCR5	13	1139
14	CD80	14	42	14	TNFSF13B	14	930
15	H2AC8	15	40	15	TNFRSF1B	15	928

The table lists the top 15 upregulated and downregulated genes identified in Set 1 and Set 2, ranked by their respective scores. The scores reflect the significance or impact of each gene based on expression analysis. Notably, genes such as TP53, INS, PTEN, and VCAM1 appear prominently across the datasets, indicating their potential biological relevance in the studied conditions.

changes, with a clear separation between significantly regulated genes and genes showing no significant variation. Dataset set 2 is represented using a volcano plot (Figure 2b), which highlights genes exhibiting both high statistical significance and large magnitude changes in expression.

Overall, the DEG profiles of the two datasets showed comparable patterns of differential regulation, indicating consistent transcriptomic dysregulation associated with cholangiocarcinoma across independent datasets.

Identification of genes from the STRING interactions

To identify key interaction patterns among differentially expressed genes, protein–protein interaction (PPI) networks were constructed using the STRING database for both upregulated and downregulated gene sets from the two datasets. The STRING-derived networks for Set 1 are shown in Figures 3a and3b, while the corresponding upregulated and downregulated networks for Set 2 are illustrated in (Figures 4a and 4b).

For Set 2, the upregulated PPI network exhibited an average node degree of 5.58 and an average local clustering coefficient of 0.3444, whereas the downregulated network showed an average node degree of 3.87 and an average local clustering coefficient of 0.339. In Set 1, the upregulated network had an average node degree of 4.32 and an average local clustering coefficient of 0.385, while the downregulated network demonstrated an average node degree of 3.12 and an average local clustering coefficient of 0.358.

In these networks, nodes represent proteins encoded by individual gene loci, with splice isoforms and post-translationally modified forms combined into a single node. Edges represent functional protein–protein associations curated by STRING, with confidence scores categorized as low (0.150), medium (0.400), high (0.700), and highest (0.900).

Identification of Top hub genes from the STRING networks

Table 1 summarizes the top 15 hub genes identified from the two gene expression datasets (Set 1 and Set 2) based on CytoHubba ranking scores derived from the STRING protein–protein interaction networks. These hub genes represent highly connected nodes within the networks, indicating their central positions in the interaction architecture.

In Set 1, genes such as TP53, HIST1H3F, H2AFZ, FOS, and POLR2B consistently ranked among the top hub genes in both upregulated and downregulated networks, reflecting their high network connectivity across different regulatory contexts. The ranking patterns suggest differential expression dynamics rather than binary on–off regulation.

In Set 2, the distribution of hub genes showed greater separation between upregulated and downregulated categories. Genes including INS, PTEN, CDKN2A, and IGF1 ranked highly in the upregulated network, while VCAM1, HGF, and CCR2 were among the top-ranked genes in the downregulated network.

The STRING-derived interaction networks for Set 1 and Set 2 are illustrated in Figures 4 and 5. The upregulated and downregulated hub gene networks for Set 1 consisted of 15 nodes with 50 and 74 edges, respectively (Figure 5a–b). For Set 2, the upregulated and downregulated networks comprised 15 nodes with 43 and 33 edges, respectively (Figure 6a–b). These hub genes were subsequently used for functional enrichment analysis using ClueGO.

Exploration of pathway and GO functional correlation using Cluego analysis

We have used the ranks of hub genes and placed them in the Cluego analysis to perform the pathways and GO functional correlation. Cytoscape has five databases for pathways analysis: Bio Carta, Elsevier pathway, KEGG, Reactome, and Wikipathway. Where set 1 upregulated and down-regulated clue consists of 34 nodes and 109 edges, as shown in Fig. 7(a). Set 2 upregulated clue consists of 34 nodes and 109 edges, shown in Fig. 7(b). Set 2 down-regulated clue contains 107 nodes and 236 edges, as shown in Fig. 8 (a, b). Fig. 7 (a, b) illustrates functional enrichment networks derived from Gene

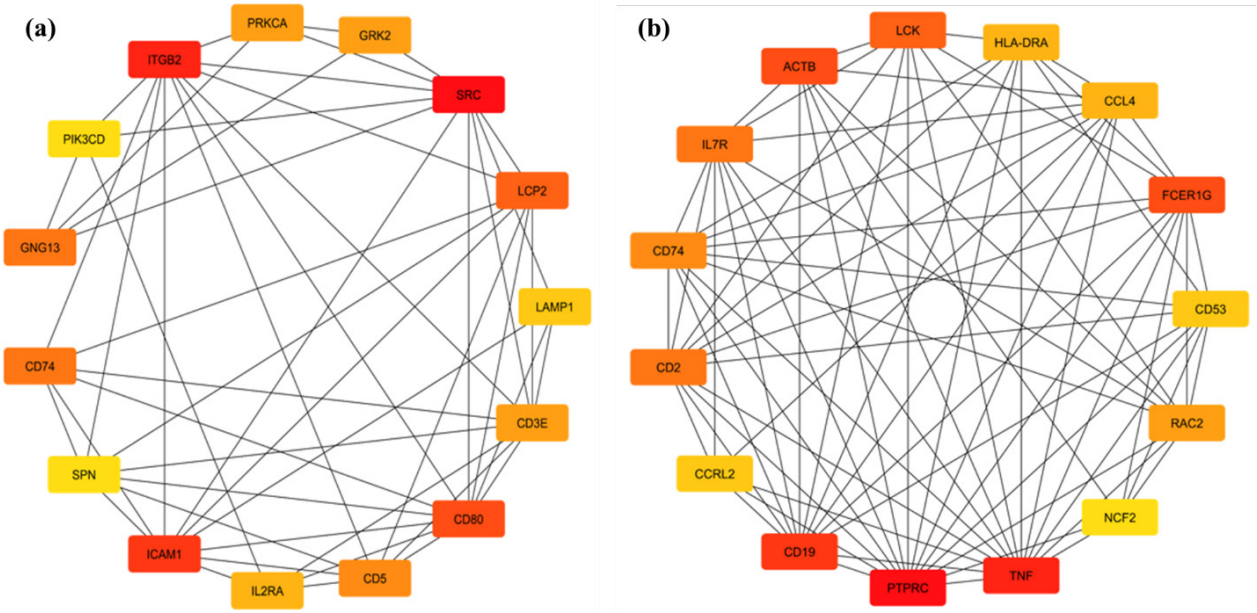


Figure 5 – (a) Hub gene interaction network of upregulated differentially expressed genes (DEGs) from Set 1 (GSE131027) identified using the CytoHubba plugin in Cytoscape. (b) Hub gene interaction network of downregulated DEGs from Set 1 (GSE131027) identified using the same CytoHubba criteria. Highly connected nodes indicate key regulatory genes potentially involved in immune signaling, inflammation, and tumor-related pathways in bile duct cancer.

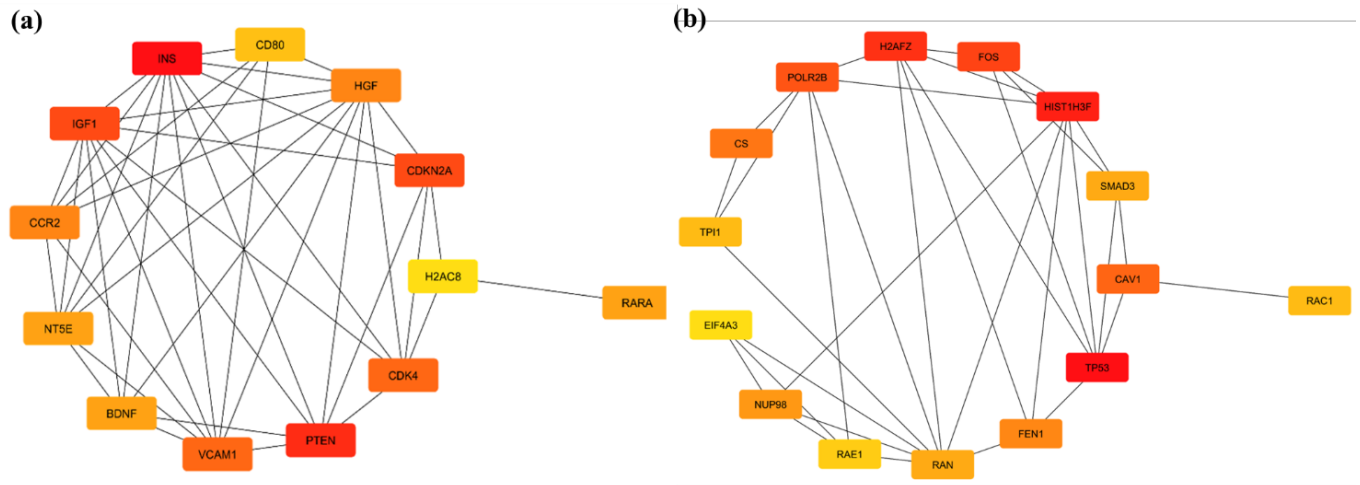


Figure 6 – (a) Hub gene interaction network of upregulated differentially expressed genes (DEGs) from Set 2 (GSE107754) identified using the CytoHubba plugin in Cytoscape. Hub genes were ranked based on network topological parameters, including Degree and BottleNeck algorithms. (b) Hub gene interaction network of downregulated DEGs from Set 2 (GSE107754) identified using the same CytoHubba criteria.

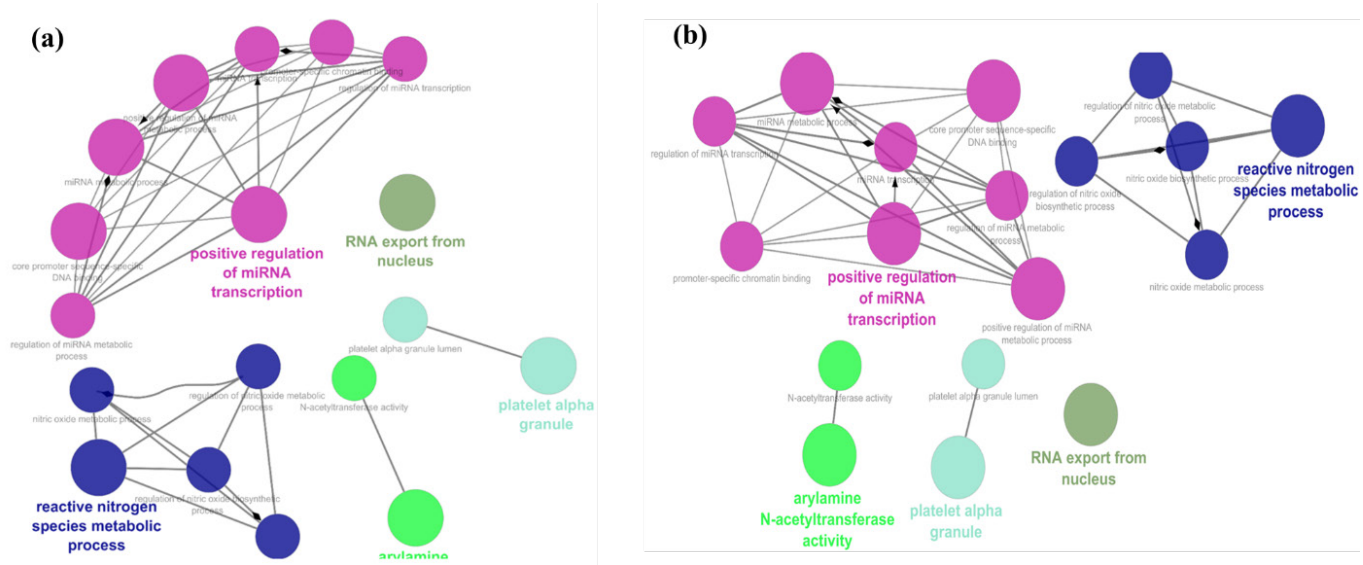


Figure 7 – (a) Set 1 top upregulated hub genes visualized as a functionally grouped network, highlighting significant Gene Ontology (GO) biological processes and molecular functions. (b) Set 1 top downregulated hub genes represented as a ClueGO network showing enriched GO terms predominantly associated metabolic process and cellular components.

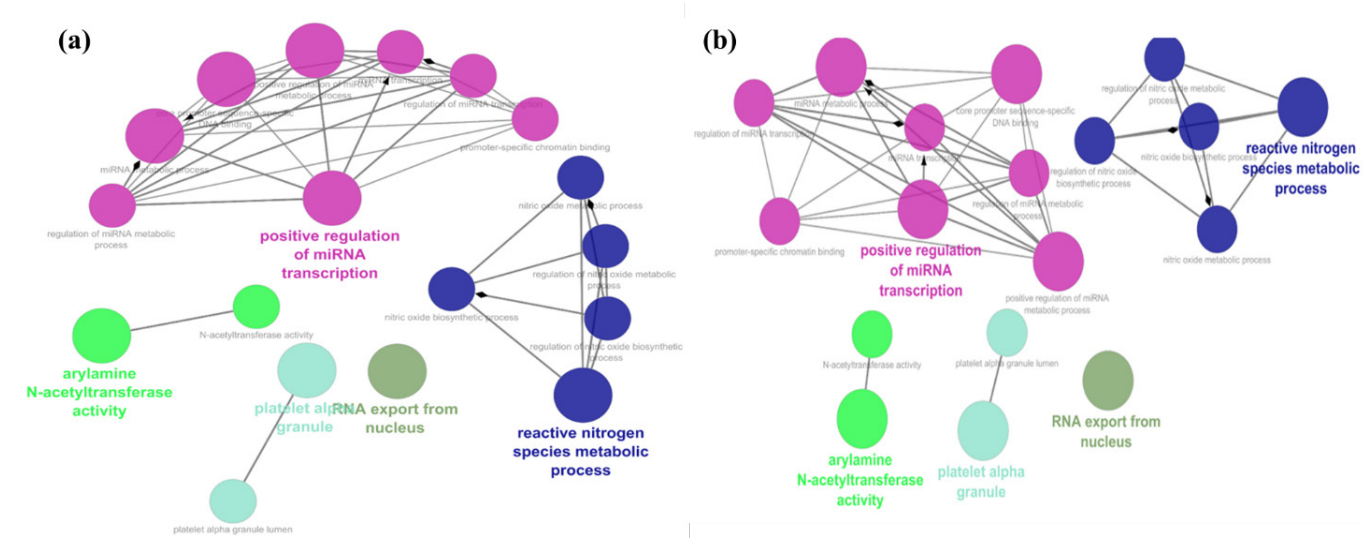


Figure 8 – (a) Set 2 top upregulated hub genes showing significantly enriched Gene Ontology (GO) terms, with dominant functional clusters related to metabolic process. (b) Set 2 top down-regulated hub genes illustrating enriched GO biological processes, molecular functions and cellular components.

Ontology analysis, comparing biological processes for Set 1 in Panel A and Set 2 in Panel B. Variations in gene cluster frequencies are represented by colored nodes categorized by functional features, including miRNA transcription (magenta),

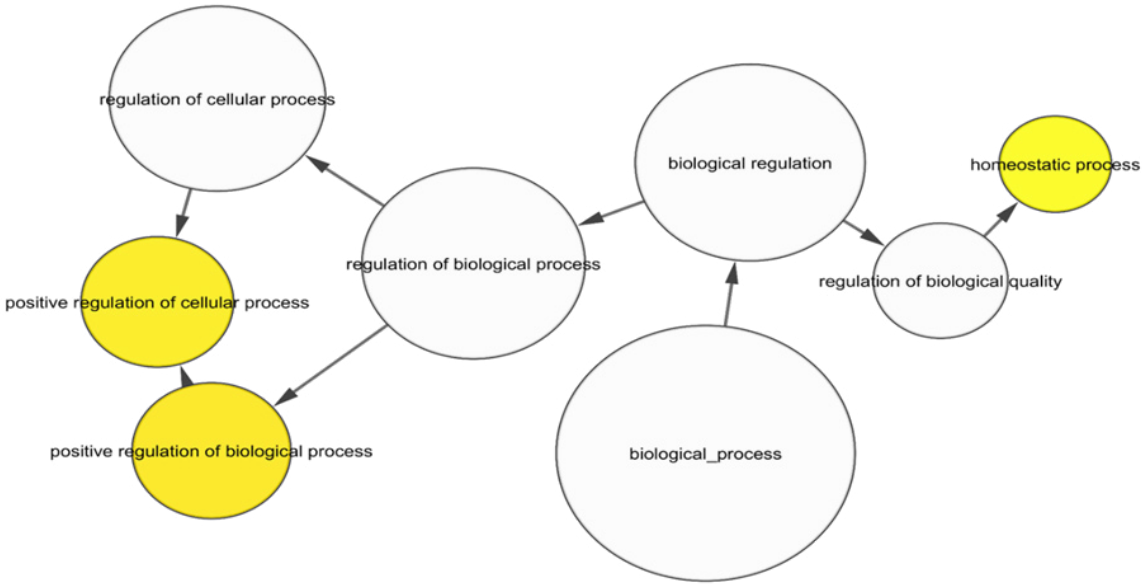


Figure 9 – The network depicts significantly overrepresented Gene Ontology (GO) biological process terms associated with Set 2 upregulated hub genes. Major enriched categories include biological process, biological regulation, regulation of biological process, regulation of cellular process, and homeostatic process.

reactive nitrogen species metabolism (dark blue), RNA export (gray-green), platelet granules (aqua), and N-acetyltransferase activity (green). The network lets users discern that nodes expand proportionately to word relevance or gene count while edges connect functional and hierarchical relationships among GO terms. Panel A's primary cluster (Set 1) centers on "positive regulation of miRNA transcription," illustrating a significant amplification of miRNA regulatory mechanisms. The procedures activate gene silencing post-transcription and regulate gene expression in response to cellular stimuli. A considerable quantity of "reactive nitrogen species" molecular constituents arises alongside "positive regulation of miRNA transcription" as primary clusters indicative of active oxidative/nitrosative stress responses. Despite the limited connections between RNA export from the nucleus and the actions of platelet alpha granules and arylamine N-acetyltransferase, they signify distinct biological processes regulated within this dataset.

The "positive regulation of miRNA transcription" hub is central in the results of Set 2, mirroring its position in Set 1, indicating the continued significance of miRNA-mediated gene regulation across datasets. The "reactive nitrogen species metabolic process" cluster in Set 2 functions as an integrated system with numerous interrelated processes, in contrast to Set 1, indicating its potential as a chronic or systemic oxidative stress response. The "arylamine N-acetyltransferase activity" node is more prominent in Set 2, suggesting its role in detoxification or drug metabolism under specific testing settings. The "platelet alpha granule" pathway persists in analysis, exhibiting varying degrees of significance relative to the initial configuration, while "RNA export from nucleus" functions similarly with negligible alterations.

Fig. 8 (a, b) illustrates GO term-based functional enrichment network maps for Set 1 (Panel A) and Set 2 (Panel B) gene groups, highlighting the biological processes and molecular activities associated with differentially expressed genes. Colored nodes depict distinct GO terms, while edges illustrate functional linkages and hierarchical connections between these terms. The illustration depicts various biological activities represented by colored pathways: magenta for miRNA transcription, dark blue for reactive nitrogen species metabolism, green for N-acetyltransferase activity, cyan for platelet alpha granule, and olive for RNA export from the nucleus.

The primary cluster identified in Panel A (Set 1) emphasizes "positive regulation of miRNA transcription," highlighting the importance of miRNA synthesis in post-transcriptional regulation for the examined biological situation. Set 1 genes significantly influence RNA-mediated gene expression modulation, which appears to be triggered by cellular stress signals and differentiation cues. The phrases "arylamine N-acetyltransferase activity," "platelet alpha granule," and "RNA export from nucleus" represent disparate categories, signifying varied physiological activities encompassing metabolic detoxification, vesicle formation, and mRNA transport.

This panel illustrates analogous functional groupings that exhibit changes in network architectural relationships relative to the initial set. A prominent cluster of "positive regulation of miRNA transcription" links both Set 1 and Set 2. However, the "reactive nitrogen species metabolic process" exhibits more significant divergence from the central network in this panel. The setup of Set 2 displays indicators for specific oxidative processes. The N-acetyltransferase activity and platelet-related granule pathways are present in Set 2, although their connection patterns vary, indicating distinct functional regulators that modify their activation patterns.

Determination of gene ontology using BINGO

The GO enrichment analysis via BiNGO reveals biological processes associated with the upregulated genes in set 2, as illustrated in Fig. 9. The GO network, utilizing STRING network data, illustrates functional associations among differentially expressed genes throughout biological pathways. Each network node displays a GO word that connects to several associated genes, with node size and yellow hue indicating statistically significant relationships. The hierarchical structure in the network is illustrated by arrows indicating that subordinate parts transmit data to overarching parent portions. Several significant GO keywords are prominent in the network, including "positive regulation of the biological process," "positive regulation of the

cellular process," and "homeostatic process." The GO keywords suggest that the genes in set 2 facilitate biological functions and sustain operational stability within cellular systems. The relationships between the nodes "regulation of biological process" and "biological regulation" culminate in the broader category "biological process," signifying their significant regulatory mechanisms inside cellular contexts that encompass the genes from set 2. The enrichment analysis failed to uncover any GO pathways associated with the downregulated genes in set 2. The deficiency of genetic annotations suggests that many down-regulated genes and their related proteins remain unrecognized, as no existing studies link them to biological processes. These genetic components suggest potential discoveries as they exist outside traditional databases despite their possible significance in context-specific biological responses. The Functional analysis pathway table was given10 in Supplementary tables 1 and 2.

Identification of Hub Genes from DEG and PPI Network Analysis

Differential expression analysis of the two GEO datasets (Set 1 and Set 2) identified more than 1,000 significantly differentially expressed genes (DEGs) in each dataset based on the applied statistical thresholds. To further prioritize genes with potential regulatory importance, protein–protein interaction (PPI) networks were constructed using the STRING database, and hub genes were identified using the CytoHubba plugin in Cytoscape. Table 1 summarizes the top 15 upregulated and downregulated hub genes identified from each dataset based on network topological scoring. In Set 1, TP53 emerged as the highest-ranked hub gene among upregulated genes, followed by HIST1H3F, H2AFZ, FOS, POLR2B, and CAV1, indicating their strong connectivity and central roles within the interaction network. These genes are primarily associated with transcriptional regulation, chromatin organization, stress response, and signal transduction. The downregulated hub genes in Set 1 included immune- and signaling-related genes such as PTPRC, TNF, CD19, and LCK, suggesting alterations in immune-associated pathways in bile duct cancer.

In Set 2, hub gene analysis revealed prominent enrichment of genes involved in metabolic regulation and growth factor signaling. INS, PTEN, CDKN2A, and IGF1 were identified among the top upregulated hub genes, while VCAM1, HGF, CCR2, and NT5E were among the most strongly downregulated genes. Notably, some genes such as INS, IGF1, PTEN, and VCAM1 appeared in both upregulated and downregulated lists, reflecting context-dependent regulation across different sample groups within the dataset.

Overall, the hub gene profiles differed between Set 1 and Set 2, likely reflecting differences in dataset composition and biological context. Nevertheless, several genes with known roles in tumor suppression, epigenetic regulation, and oncogenic signaling consistently ranked highly, supporting their potential relevance in cholangiocarcinoma biology and justifying their further functional interpretation.

Discussion

In the present study, an integrative bioinformatics framework was applied to identify key hub genes and functionally enriched biological pathways associated with cholangiocarcinoma using publicly available Gene Expression Omnibus (GEO) datasets. By integrating differential gene expression analysis with protein–protein interaction (PPI) network construction and hub gene prioritization, this study aimed to uncover genes with potential regulatory significance in bile duct cancer pathogenesis.

Among the identified hub genes, TP53 consistently ranked as the most central node across both datasets. TP53 is a well-established tumor suppressor gene involved in cell cycle regulation, apoptosis, and maintenance of genomic stability, and its dysregulation has been frequently reported

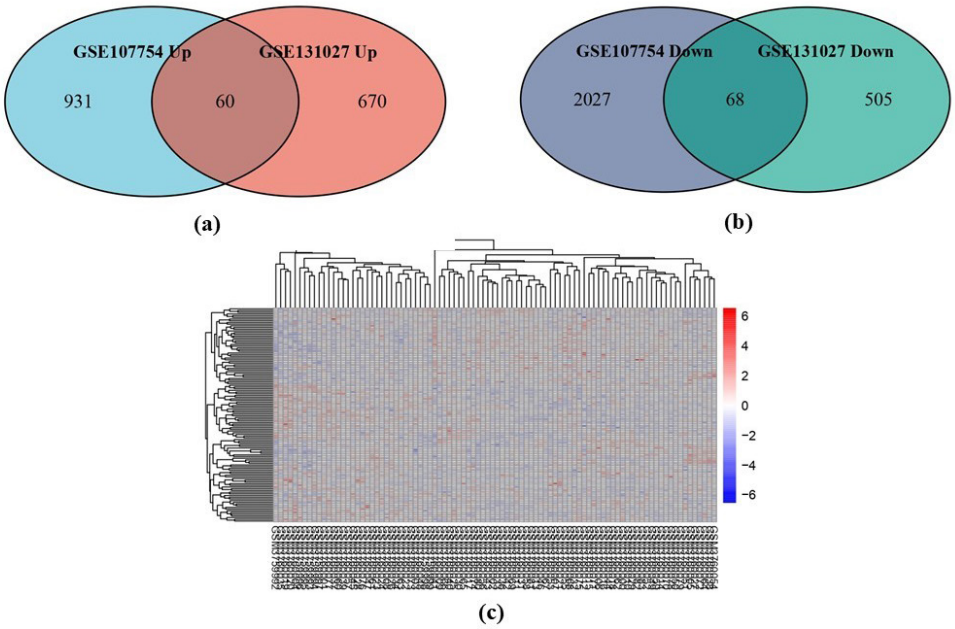


Figure 10 – (a) Venn diagram showing the overlap of upregulated DEGs between GSE107754 and GSE131027, with 60 genes commonly upregulated in both datasets. (b) Venn diagram illustrating the overlap of downregulated DEGs between GSE107754 and GSE131027, identifying 68 commonly downregulated genes. (c) Hierarchical clustering heatmap of the common DEGs, displaying their expression patterns across samples; red indicates higher expression levels and blue indicates lower expression levels, as shown by the color scale.

in cholangiocarcinoma. The prominent network centrality of TP53 observed in this study further supports its pivotal role in bile duct cancer pathogenesis. Previous studies have reported that alterations in TP53 are associated with aggressive tumor behavior, poor prognosis, and increased tumor mutation burden in intrahepatic cholangiocarcinoma, further highlighting its relevance in bile duct cancer biology [31]

In addition to TP53, SMAD3 emerged as a highly connected hub gene, underscoring the importance of the transforming growth factor-β (TGF-β) signaling pathway in cholangiocarcinoma. SMAD3 functions as a key intracellular mediator of TGF-β signaling and has been shown to promote epithelial–mesenchymal transition (EMT), tumor invasiveness, and metastatic dissemination in bile duct cancer. Experimental studies have demonstrated that suppression of SMAD3 phosphorylation attenuates EMT and inhibits tumor progression, highlighting the oncogenic role of aberrant TGF-β1/SMAD3 activation in cholangiocarcinoma [32]. The identification of SMAD3 as a hub gene in our network analysis further supports its biological relevance in disease progression.

We propose that the HSP module maintains proteostasis by improving the structural integrity of proteins necessary to oncogenic proteins. Many genes are upregulated and downregulated. In this study, HSP significantly contributes to oncogenesis via homologous recombination repair. In our study, several key genes implicated in thermoresistance were marked overexpressed in tumors of different tissues. Moreover, the expression patterns of the thermogenesis like TP53, HIST1H3F, H2AFZ, FOS, POLR2B, CAV1, CS, FEN1, NUP98, SMAD3, RAN, TPI1, RAC1, RAE1, EIF4A3 these are the top upregulated genes which shows the positively correlated in diverse types of the bile duct cancer [33–42]. Histone-related genes such as HIST1H3F and H2AFZ were also identified among the hub genes, indicating that alterations in chromatin organization and transcriptional regulation may contribute to the abnormal gene expression landscape observed in cholangiocarcinoma. These findings are consistent with previous reports highlighting the importance of epigenetic dysregulation in biliary tract cancers.

Caveolin-1 (CAV1) was another prominent hub gene identified in this study, highlighting the contribution of tumor–stromal interactions to cholangiocarcinoma progression. Previous investigations have demonstrated that elevated CAV1 expression in cancer-associated fibroblasts correlates with poor overall and recurrence-free survival in intrahepatic cholangiocarcinoma. Mechanistically, CAV1-positive fibroblasts promote an immunosuppressive tumor microenvironment by enhancing regulatory T-cell infiltration and impairing effective anti-tumor immune responses. These findings suggest that CAV1 contributes to disease aggressiveness through modulation of stromal signaling and immune regulation, supporting its potential value as both a prognostic biomarker and a therapeutic target.

The transcription factor FOS, a core component of the activator protein-1 (AP-1) complex, was also identified as a central hub gene. AP-1 regulates gene expression programs involved in cell proliferation, apoptosis, stress responses, and oncogenic transformation. Integrative epigenomic profiling studies have demonstrated that AP-1 binding sites are enriched in regions exhibiting altered histone modifications in intrahepatic cholangiocarcinoma, indicating a critical role for FOS-mediated transcriptional regulation. Dysregulation of AP-1 components, including FOS, has been associated with tumor progression and unfavorable clinical outcomes, supporting its biological and prognostic relevance in bile duct cancer [43].

In one dataset, genes involved in metabolic and growth factor signaling, including INS, IGF1, and PTEN, were identified as hub genes. However, as this dataset represents a broader pan-cancer cohort, the involvement of these genes may reflect tissue-contextual or systemic regulatory effects rather than cholangiocarcinoma-specific oncogenic drivers. Consequently, these findings should be interpreted with caution and require further experimental validation to establish their disease-specific relevance.

Functional enrichment analysis using ClueGO and BiNGO revealed that the identified hub genes were significantly associated with biological processes related to transcriptional regulation, oxidative and nitrosative stress responses, RNA processing, and intracellular transport. These processes have been previously implicated in cancer initiation and progression and support the biological plausibility of the identified gene networks. Overall, the integration of network-based hub gene identification with functional enrichment analysis provides a systems-level perspective on the molecular mechanisms underlying cholangiocarcinoma and highlights potential targets for future therapeutic exploration. Hierarchical clustering heatmap analysis of the common differentially expressed genes (DEGs) demonstrated distinct expression profiles between the sample groups. The consistent clustering pattern indicates robust differential expression and supports the reliability of the identified DEGs across datasets was shown in Fig. 10.

The hub genes identified in this study provide important biological insights into the molecular mechanisms underlying bile duct cancer progression. TP53 emerged as the most significant hub gene, reaffirming its critical role in regulating cell cycle control, DNA repair, and apoptosis, with its dysregulation contributing to genomic instability in cholangiocarcinoma [10,11]. Epigenetic regulators such as HIST1H3F and H2AFZ suggest that chromatin remodeling and transcriptional reprogramming are central to tumor development [44,45]. FOS links chronic inflammation to oncogenic signaling, while POLR2B reflects increased transcriptional activity in malignant cells. Additionally, CAV1 and SMAD3 implicate tumor–stroma interactions, TGF-β signaling, and epithelial–mesenchymal transition in disease progression [46–49].

Our study highlights key differentially expressed and hub genes potentially involved in bile duct cancer progression, providing insights into its molecular mechanisms. By integrating two independent GEO datasets, we enhanced the robustness of our findings. However, the sample size remains limited, which may reduce statistical power and limit representation of molecular heterogeneity across diverse patient populations and clinical subtypes. The analyses were entirely computational, and experimental validation of the identified genes was not performed. Therefore, the biological functions and clinical relevance of these hub genes require confirmation through in vitro and in vivo experiments, including quantitative PCR, protein expression studies, and functional assays.

Protein–protein interaction networks were constructed using the STRING database, which combines experimentally validated and predicted interactions. While this approach offers a systems-level perspective, it may introduce biases due to incomplete annotations, potential false positives, and lack of tissue- or disease-specific context, warranting cautious interpretation of network relationships. Furthermore, clinical

variables such as tumor stage, histological subtype, treatment status, and patient outcomes were unavailable, limiting correlations between gene expression and clinical features. Future studies incorporating larger sample sizes, detailed clinical data, and experimental validation are essential to confirm the diagnostic and therapeutic potential of the identified hub genes and facilitate their translation into clinical applications.

Limitations

This study is limited by the use of publicly available datasets with relatively small sample sizes and by the absence of experimental validation. Additionally, GSE131027 represents a pan-cancer cohort; although bile duct cancer samples were extracted for analysis, residual biological heterogeneity may influence the results. Furthermore, protein–protein interaction networks derived from STRING represent predicted and curated interactions that may not fully capture tissue-specific biological contexts. Future studies incorporating independent cohorts and experimental validation are required to confirm these findings.

Conclusion

In conclusion, this study applied an integrated bioinformatics framework to investigate the molecular landscape of cholangiocarcinoma using publicly available GEO transcriptomic datasets. By combining differential gene expression analysis, protein–protein interaction network construction, hub gene ranking, and functional enrichment analysis, we identified several highly connected genes and biologically relevant pathways associated with bile duct cancer. Network-based analysis consistently highlighted TP53 as a central hub gene, along with other key regulators such as SMAD3, CAV1, FOS, HIST1H3F, H2AFZ, and POLR2B, suggesting coordinated dysregulation of pathways related to transcriptional control, signaling, and chromatin organization. Functional enrichment further supported the involvement of processes linked to gene regulation, oxidative and nitrosative stress responses, and RNA processing in cholangiocarcinoma. Although the findings are derived from in silico analyses and require experimental validation, this study provides a systems-level perspective on the molecular mechanisms underlying bile duct cancer. The identified hub genes and pathways offer a valuable resource for future experimental studies and may contribute to improved understanding of cholangiocarcinoma biology.

Supplementary materials

The Supplementary information includes tables:

• Supplementary Table 1. KEGG pathway enrichment analysis of differentially expressed genes (DEGs);

• Supplementary Table 2. Gene Ontology (GO) Biological Process enrichment analysis of differentially expressed genes (DEGs).

This supplemental materials have been provided by the authors to give readers additional information about their work.

The file can be accessed using: https://www.editorialpark.com/download/article-supp/758/Supplementary-data.docx.

Author Contributions: Conceptualization, M. K.; methodology, R. N., P. H. V. and M. K; validation, M. K. and P. H. V; formal analysis, R. N.; investigation, P. H. V.; resources, P. H. V.; data curation, R. N., P. H. V.; writing – original draft preparation, P. H. V. and R. N.; writing – review and editing, R. N. and P. H. V.; visualization, R. N. and P. H. V.; supervision, M. K.; project administration, M. K. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: Authors are thankful to Saveetha School of Engineering, SIMATS, India for providing support to carry out the work.

Funding: None.

Data availability statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Chun K. Recent Classifications of the Common Bile Duct Injury. Korean J. Hepato-Biliary-Pancreatic Surg. 2014; 18: 69, https://doi.org/10.14701/kjhbps.2014.18.3.69.
2. Joo I, Lee JM. Imaging Bile Duct Tumors: Pathologic Concepts, Classification, and Early Tumor Detection. Abdom. Imaging. 2013; 38: 1334–1350. https://doi.org/10.1007/s00261-013-0027-3.
3. Lee SH, Song SY. Recent Advancement in Diagnosis of Biliary Tract Cancer through Pathological and Molecular Classifications. Cancers (Basel). 2024; 16: 1761. https://doi.org/10.3390/cancers16091761.
4. Oba A, Del Chiaro M, Satoi S, Kim S, Takahashi H, Yu J, Hioki M, Tanaka M, Kato Y, Ariake K, Wu YHA, Inoue Y, Takahashi Y, Hackert T, Wolfgang CL, Besselink MG, Schulick RD, Nagakawa Y, Isaji S, Tsuchida A, Endo I. New Criteria of Resectability for Pancreatic Cancer: A Position Paper by the Japanese Society of Hepato-Biliary-Pancreatic Surgery (JSHBPS). J. Hepatobiliary. Pancreat. Sci. 2022; 29: 725–731. https://doi.org/10.1002/jhbp.1049.
5. Otsubo T, Kobayashi S, Sano K, Misawa T, Ota T, Katagiri S, Yanaga K, Yamaue H, Kokudo N, Unno M, Fujimoto J, Miura F, Miyazaki M, Yamamoto M. Safety-Related Outcomes of the Japanese Society of Hepato-Biliary-Pancreatic Surgery Board Certification System for Expert Surgeons. J. Hepatobiliary. Pancreat. Sci. 2017; 24: 252–261. https://doi.org/10.1002/jhbp.444.
6. Kuroda S, Kobayashi T, Hatano E, Kubo S, Endo I, Ohdan H. Questionnaire on the Surgical Indications for Intrahepatic Cholangiocarcinoma Administered to Japanese Board-certified Expert Hepatobiliary and Pancreatic Surgeons and Instructors. J. Hepatobiliary. Pancreat. Sci.

2025; 32: 179–193. https://doi.org/10.1002/jhbp.12108.
7. Baison GN, Bonds MM, Helton WS, Kozarek RA. Choledochal Cysts: Similarities and Differences between Asian and Western Countries. World J. Gastroenterol. 2019; 25: 3334–3343. https://doi.org/10.3748/wjg.v25.i26.3334.
8. Oze I, Ito H, Koyanagi YN, Abe SK, Rahman MS, Islam MR, Saito E, Gupta PC, Sawada N, Tamakoshi A, et al. Obesity Is Associated with Biliary Tract Cancer Mortality and Incidence: A Pooled Analysis of 21 Cohort Studies in the Asia Cohort Consortium. Int. J. Cancer. 2024; 154: 1174–1190. https://doi.org/10.1002/ijc.34794.
9. Keane MG, Horsfall L, Rait G, Pereira SP. A Case–Control Study Comparing the Incidence of Early Symptoms in Pancreatic and Biliary Tract Cancer. BMJ Open. 2014; 4: e005720. https://doi.org/10.1136/bmjopen-2014-005720.
10. Guo L, Zhou F, Liu H, Kou X, Zhang H, Chen X, Qiu J. Genomic Mutation Characteristics and Prognosis of Biliary Tract Cancer. Am. J. Transl. Res. 2022; 14: 4990–5002.
11. Wardell CP, Fujita M, Yamada T, Simbolo M, Fassan M, Karlic R, Polak P, Kim J, Hatanaka Y, Maejima K et al. Genomic Characterization of Biliary Tract Cancers Identifies Driver Genes and Predisposing Mutations. J. Hepatol. 2018; 68: 959–969. https://doi.org/10.1016/j.jhep.2018.01.009.
12. Tompkins RK, Saunders K, Roslyn JJ, Longmire WP. Changing Patterns in Diagnosis and Management of Bile Duct Cancer. Ann. Surg. 1990; 211: 614–620, discussion 620–621.
13. Tanaka K, Kida M. Role of endoscopy in screening of early pancreatic cancer and bile duct cancer. Dig. Endosc. 2009; 21. https://doi.org/10.1111/j.1443-1661.2009.00856.x.
14. Ryan ME. Cytologic Brushings of Ductal Lesions during ERCP. Gastrointest. Endosc. 1991; 37: 139–142. https://doi.org/10.1016/S0016-5107(91)70671-8.
15. Seyama Y, Makuuchi M. Current Surgical Treatment for Bile Duct Cancer. World J. Gastroenterol. 2007; 13: 1505–1515. https://doi.org/10.3748/wjg.v13.i10.1505.
16. Cowzer D, Harding, JJ. Advanced Bile Duct Cancers: A Focused Review on Current and Emerging Systemic Treatments. Cancers (Basel). 2022; 14: 1800. https://doi.org/10.3390/cancers14071800.
17. Huguet JM, Lobo M, Labrador JM, Boix C, Albert C, Ferrer-Barceló L, Durá AB, Suárez P, Iranzo I, Gil-Raga M, Burgos CB, Sempere J. Diagnostic-Therapeutic Management of Bile Duct Cancer. World J. Clin. Cases. 2019; 7: 1732–1752. https://doi.org/10.12998/wjcc.v7.i14.1732.
18. Min L, Ziyu D, Xiaofei Z, Shunhe X, Bolin W. Analysis of Levels and Clinical Value of CA19-9, NLR and SIRI in Patients with Pancreatic Cancer with Different Clinical Features. Cell. Mol. Biol. 2022; 67: 302–308. https://doi.org/10.14715/cmb/2021.67.4.41.
19. Liu J, Gao J, Du Y, Li Z, Ren Y, Gu J, Wang X, Gong Y, Wang W, Kong X. Combination of Plasma MicroRNAs with Serum CA19-9 for Early Detection of Pancreatic Cancer. Int. J. Cancer. 2012; 131: 683–691. https://doi.org/10.1002/ijc.26422.
20. Liao B. Research on the Factors That Affecting the Occurrence of Gastric Cancer Based on NCBI Gene Expression Omnibus Database. AIP Conference Proceedings. 2020; 2208 (1): 020007. https://doi.org/10.1063/5.0000016.
21. Yu-jing T, Wen-jing T, Biao T. Integrated Analysis of Hub Genes and Pathways In Esophageal Carcinoma Based on NCBI's Gene Expression Omnibus (GEO) Database: A Bioinformatics Analysis. Med. Sci. Monit. 2020; 26. https://doi.org/10.12659/MSM.923934.
22. Kumar SU, Kumar DT, Siva R, Doss CGP, Zayed H. Integrative Bioinformatics Approaches to Map Potential Novel Genes and Pathways Involved in Ovarian Cancer. Front. Bioeng. Biotechnol. 2019; 7. https://doi.org/10.3389/fbioe.2019.00391.
23. Yang Y, Qi S, Shi C, Han X, Yu J, Zhang L, Qin S, Gao Y. Identification of Metastasis and Prognosis-Associated Genes for Serous Ovarian Cancer. Biosci. Rep. 2020; 40. https://doi.org/10.1042/BSR20194324.
24. Barrett T, Edgar R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO). Methods Mol Biol. 2006; 338: 175–190. https://doi.org/10.1385/1-59745-097-9:175.
25. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol. 2016; 1418: 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.
26. Vella D, Marini S, Vitali F, Di Silvestre D, Mauri G, Bellazzi R. MTGO: PPI Network Analysis Via Topological and Functional Module Identification. Sci. Rep. 2018; 8: 5499, https://doi.org/10.1038/s41598-018-23672-0.
27. Murakami Y, Tripathi LP, Prathipati P, Mizuguchi K. Network Analysis and in Silico Prediction of Protein–Protein Interactions with Applications in Drug Discovery. Curr. Opin. Struct. Biol. 2017; 44: 134–142. https://doi.org/10.1016/j.sbi.2017.02.005.
28. Tomkins JE, Manzoni C. Advances in Protein-Protein Interaction Network Analysis for Parkinson's Disease. Neurobiol. Dis. 2021; 155: 105395. https://doi.org/10.1016/j.nbd.2021.105395.
29. Li T, Gao X, Han L, Yu J, Li H. Identification of Hub Genes with Prognostic Values in Gastric Cancer by Bioinformatics Analysis. World J. Surg. Oncol. 2018; 16: 114. https://doi.org/10.1186/s12957-018-1409-3.
30. Lv J, Li L. Hub Genes and Key Pathway Identification in Colorectal Cancer Based on Bioinformatic Analysis. Biomed Res. Int. 2019; 2019: 1–13. https://doi.org/10.1155/2019/1545680.
31. Guo C, Liu Z, Yu Y, Chen Y, Liu H, Guo Y, Peng Z, Cai G, Hua Z, Han X, Li Z. TP53 /KRAS Co-Mutations Create Divergent Prognosis Signatures in Intrahepatic Cholangiocarcinoma. Front. Genet. 2022; 13. https://doi.org/10.3389/fgene.2022.844800.
32. Deng L, Bao W, Zhang B, Zhang S, Chen Z, Zhu X, He B, Wu L, Chen X, Deng T, Chen B, Yu Z, Wang Y, Chen G. AZGP1 Activation by Lenvatinib Suppresses Intrahepatic Cholangiocarcinoma Epithelial-Mesenchymal Transition through the TGF-B1/Smad3 Pathway. Cell Death Dis. 2023; 14: 590. https://doi.org/10.1038/s41419-023-06092-5.
33. Zou W, Zhang Q, Sun R, Li X, He S. Study on TFF1 and PALB2 Gene Variants Associated with Gastric Carcinoma Risk in the Chinese Han Population. Cancer Epidemiol. 2023; 83: 102333. https://doi.org/10.1016/j.canep.2023.102333.
34. Ghojazadeh M, Somi MH, Naseri A, Salehi-Pourmehr H, Hassannezhad S, Hajikamanaj Olia A, Kafshdouz L, Nikniaz Z. Systematic Review and Meta-Analysis of TP53, HER2/ERBB2, KRAS, APC, and PIK3CA Genes Expression Pattern in Gastric Cancer. Middle East J. Dig. Dis. 2022; 14: 335–345, https://doi.org/10.34172/mejdd.2022.292.
35. Wu Y, Zhao H. Circ_0074027 Binds to EIF4A3 and Promotes Gastric Cancer Progression. Oncol. Lett. 2021; 22: 704, https://doi.org/10.3892/ol.2021.12965.
36. Liang X, Chen W, Shi H, Gu X, Li Y, Qi Y, Xu K, Zhao A, Liu J. PTBP3 Contributes to the Metastasis of Gastric Cancer by Mediating CAV1 Alternative Splicing. Cell Death Dis. 2018; 9: 569, https://doi.org/10.1038/s41419-018-0608-8.

37. Peng J, Liang S, Li L. SFRP1 Exerts Effects on Gastric Cancer Cells through GSK3β/Rac1-mediated Restraint of TGFβ/Smad3 Signaling. Oncol. Rep. 2018. https://doi.org/10.3892/or.2018.6838.
38. Fenoglio-Preiser CM, Wang J, Stemmermann GN, Noffsinger A. TP53 and Gastric Carcinoma: A Review. Hum. Mutat. 2003; 21: 258–270, https://doi.org/10.1002/humu.10180.
39. Zhou Q, Zheng X, Chen L, Xu B, Yang X, Jiang J, Wu C. Smad2/3/4 Pathway Contributes to TGF-β-Induced MiRNA-181b Expression to Promote Gastric Cancer Metastasis by Targeting Timp3. Cell. Physiol. Biochem. 2016; 39: 453–466, https://doi.org/10.1159/000445638.
40. Liu L, Zhou C, Zhou L, Peng L, Li D, Zhang X, Zhou M, Kuang P, Yuan Q, Song X, Yang M. Functional FEN1 Genetic Variants Contribute to Risk of Hepatocellular Carcinoma, Esophageal Cancer, Gastric Cancer and Colorectal Cancer. Carcinogenesis. 2012; 33: 119–123, https://doi.org/10.1093/carcin/bgr250.
41. Zhu Z, Peng R, Cai H. The Value of Nucleoporin 188 in Diagnosis, Prognosis and Immunoregulation: From Pan-Cancer Analysis to Gastric Cancer Verification. Front. Immunol. 2025; 16. https://doi.org/10.3389/fimmu.2025.1586784.
42. Muste Sadurni, M, Saponaro, M. Deregulations of RNA Pol II Subunits in Cancer. Appl. Biosci. 2023; 2: 459–476, https://doi.org/10.3390/applbiosci2030029.
43. He K, Feng Y, An S, Liu F, Xiang G. Integrative Epigenomic Profiling Reveal AP-1 Is a Key Regulator in Intrahepatic Cholangiocarcinoma. Genomics. 2022; 114: 241–252, https://doi.org/10.1016/j.ygeno.2021.12.008.
44. Subramanian V, Fields PA, Boyer LA. H2A.Z: A Molecular Rheostat for Transcriptional Control. F1000Prime Rep. 2015; 7. https://doi.org/10.12703/P7-01.
45. Vardabasso C, Hasson D, Ratnakumar K, Chung C-Y, Duarte LF, Bernstein E. Histone Variants: Emerging Players in Cancer Biology. Cell. Mol. Life Sci. 2014; 71: 379–404, https://doi.org/10.1007/s00018-013-1343-z.
46. Jafri Z, Li Y, Zhang J, O'Meara CH, Khachigian LM. Jun, an Oncological Foe or Friend? Int. J. Mol. Sci. 2025; 26: 555, https://doi.org/10.3390/ijms26020555.
47. Bradner JE, Hnisz D, Young RA. Transcriptional Addiction in Cancer. Cell. 2017; 168: 629–643, https://doi.org/10.1016/j.cell.2016.12.013.
48. Goetz JG, Lajoie P, Wiseman SM, Nabi IR. Caveolin-1 in Tumor Progression: The Good, the Bad and the Ugly. Cancer Metastasis Rev. 2008; 27: 715–735, https://doi.org/10.1007/s10555-008-9160-9.
49. Xu J, Lamouille S, Derynck R. TGF-β-Induced Epithelial to Mesenchymal Transition. Cell Res. 2009; 19: 156–172, https://doi.org/10.1038/cr.2009.5.

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/10.23950/jcmk/18532

Prediction of Paroxysmal Atrial Fibrillation in Patients with Sinus Rhythm

Olga Germanova[1], Yulia Reshetnikova[2], Andrey Germanov[1], Giuseppe Galati[3], Inga Prokhorenko[1]

[1]*Department of _____, Medical University Reaviz, Samara, Russia*
[2]*Department of _____, Samara State Medical University, Samara, Russia*
[3]*Department of _____, I.R.C.C.S. Ospedale Multimedica – Cardiovascular Scientific Institute, Milan, Italy*

Received: 2025-12-09.
Accepted: 2026-05-06.

J Clin Med Kaz 2026; 23(3): 24-30

Corresponding author:
Olga Germanova.
E-mail: olga_germ@mail.ru.
ORCID: 0000-0003-4833-4563.

Abstract

Objective: To create a tool for the prediction of paroxysmal atrial fibrillation (PAF) in patients with sinus rhythm.

Methods: Single-center, case-control study. None of the patients had a prior diagnosis of AF or reported symptoms of heart arrhythmias. Among the cohort of 6,630 individuals, paroxysms of AF were incidentally detected during 24-hours Holter ECG monitoring in 97 patients (main group). The control group - 99 patients from the same primary cohort without PAF. We assessed supraventricular and ventricular ectopic activity, the presence of pauses and blocks, changes of ST segment, QT interval durations, and heart rate variability.

Results: We formulated a regression equation to estimate the probability of PAF in patients with sinus rhythm. The most significant risk predictors: early "P on T" premature ectopic complexes (p<0.0001); coupled premature ventricular ectopic complexes (p=0.021); ventricular allorhythmias (OR=0.997). Other analyzed factors, including the frequency of both atrial and atrioventricular premature ectopic complexes, as well as single ventricular ectopic complexes, did not exhibit a statistically significant effect on the risk of PAF, according to this model.

Conclusion: The final regression equation, based on the evaluation of data from 24-hours Holter ECG monitoring, incorporates the following criteria: gender, the number of atrial and atrioventricular supraventricular complexes, the count of single and paired ventricular complexes, variations in rhythms with ventricular complexes, as well as the presence or absence of early "P on T" complexes (AUC=0.996).

Keywords: atrial fibrillation, premature complexes, paroxysmal atrial fibrillation, predictive equation.

Introduction

Paroxysmal atrial fibrillation (PAF) is one of the most common heart arrhythmia all over the world, characterized by intermittent episodes of atrial fibrillation (AF) that typically resolve spontaneously. It poses significant clinical challenges due to its potential to progress to persistent forms of AF, leading to increased morbidity and mortality. Understanding the risk factors and physiological mechanisms underlying PAF is crucial for early identification and intervention in at-risk populations.

Recent studies have highlighted the importance of identifying patients with sinus rhythm who have higher risk of developing PAF. This is particularly relevant for asymptomatic patients who have PAF episodes and due to the limitations of current diagnostic methods. Advances in technology and analytical methodologies have opened new avenues for risk stratification, including the use of machine learning algorithms and wearable devices that can monitor cardiac rhythm continuously [1–4].

Risk factors, associated with PAF, include age, arterial hypertension, diabetes mellitus, obesity, and structural heart disease, but emerging evidence suggests that biomarkers and genetic predisposition may also play significant roles [5–8]. Additionally, lifestyle factors, such as alcohol consumption and physical inactivity, have been implicated in the development of PAF, indicating that preventive strategies could be implemented in appropriate patient populations [9–12].

The integration of these diverse risk factors into predictive models holds promise for stratifying patients with sinus rhythm, allowing for more personalized approaches to the management and prevention of PAF. Recent research efforts have focused on developing and validating such predictive models, which may incorporate clinical parameters, echocardiographic findings, and novel biomarkers [13, 14]. As the field of cardiology moves towards a more precision-based approach, understanding how to effectively predict PAF in patients with sinus rhythm may provide critical insights for clinicians and patients alike.

When improving the possibility of PAF prediction before its manifestation, prevention strategies take a leading role, including lifestyle modification, pharmacological therapy, or even catheter ablation in select populations, which could significantly improve patients outcomes while reducing healthcare costs related to PAF complications [15–18].

Objective: To create a tool for the prediction of PAF in patients with sinus rhythm.

Methods

A single-center, case-control study was conducted. We analyzed the data from 6,630 patients who underwent routine examinations of 24-hours Holter ECG monitoring during hospitalization in Samara State Medical University Clinics. None of the patients had a prior diagnosis of AF or reported symptoms, related to rhythm disturbances before the study commenced. Among this cohort of 6,630 individuals, paroxysms of AF were incidentally detected during 24-hours Holter ECG monitoring in 97 patients. They were included into the main group. The control group consisted of 99 patients from the same primary cohort, without paroxysms of AF. We selected the patients for the control group from the whole cohort the way that both groups were equal in terms of anthropometric parameters and comorbidities.

All patients underwent standard laboratory and instrumental examinations. In addition to the 24-hours Holter ECG monitoring, instrumental methods included transthoracic echocardiography and Doppler ultrasound of the brachiocephalic arteries. Stress echocardiography with physical exercise or pharmacological testing, as well as coronary angiography, were performed if indicated. During the data analysis of 24-hours Holter ECG monitoring, we studied the following key parameters: ECG registration time, heart rhythm pacemaker, supraventricular and ventricular ectopic activity, the presence of pauses and blocks, changes of the ST segment, QT interval durations, and heart rate variability. Special attention was given to the presence of early extrasystoles of the “P on T” and “R on T” types.

The research was conducted in accordance with the Helsinki Declaration, and the local ethical committee approved study protocol (№248 dated 27.04.2022, SamSMU University Ethical committee). All patients signed an informed agreement to participate in the study.

In statistical analysis, we followed the principles of evidence-based medicine. Initially, the normality of the distribution of the analyzed parameters was determined. For normally distributed data, parametric criteria were used (quantitative variables were characterized by calculating the mean and standard deviation; intergroup comparisons were performed using one-way ANOVA with F-statistic, degrees of freedom “df,” and statistical significance “p”). For non-normally distributed data, non-parametric criteria were applied (for quantitative indicators, medians and the first and third quartiles – Q1 and Q3 – were provided; intergroup comparisons were conducted using the Kruskal-Wallis method, indicating values for H and “p”). Differences between groups were considered statistically significant at $p \leq 0.05$. For all statistical tests, the criterion for statistical significance was set at $p \leq 0.05$. Statistical analysis was performed using MedCalc® Statistical Software version 20.118 (MedCalc Software Ltd, Ostend, Belgium; https://www.medcalc.org; 2022), GraphPad Prism for Windows, version 10.1.0 (GraphPad Software, San Diego, California, USA, www.graphpad.com), and the open-source R software environment (https://cran.rstudio.com/).

Results

The main and control groups had no statistically significant differences in anthropometric parameters or comorbidities. However, the analysis of data from 24-hours Holter ECG monitoring revealed statistically significant variations in both supraventricular and ventricular ectopic activity characteristics (see Table 1 on the next page).

In summary, the parameters observed in 24-hours Holter ECG monitoring, such as the presence of ectopic complexes and paroxysmal tachycardia, occurred significantly more frequently and with higher values in the main group with PAF. Specifically, the vast majority (97.9%) of patients in the main group with diagnosed PAF exhibited early atrial ectopic complexes of the “P on T” type, compared to only 4.0% in the control group (odds ratio [OR] = 8461.648 [382.1983; 187336]). The frequency of supraventricular ectopic complexes was also significantly higher in the main group, including single, coupled, and grouped ectopic complexes. Additionally, the interval durations were notably longer in the group with PAF. However, the rates of ventricular ectopic complexes and ST segment depression did not differ significantly between the groups. In other words, patients of the main group with PAF demonstrated distinctive rhythm and conduction changes when compared to the control group, despite both groups being equal in key anthropometric characteristics and comorbidities.

To develop a predictive tool for PAF in patients with sinus rhythm, we analyzed the relevance and significance of various parameters, including early "P on T" ectopic complexes, coupled ventricular ectopic complexes, and ventricular allorhythmias. The presence of early “P on T” premature ectopic complexes substantially increased the likelihood of developing PAF by a factor of 8461 compared to its absence ($p<0.0001$). Conversely, an increase in paired premature ventricular ectopic complexes

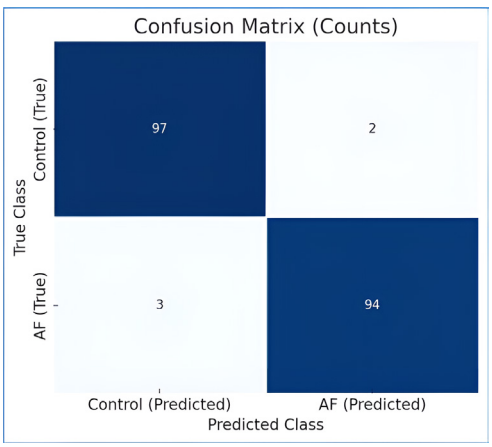


Figure 1 – Confusion plot. AF – main group (n=97), control – control group (n=99)

Table 1 24-hours Holter ECG monitoring data within the main and control groups

Parameters, Median (Q1, Q3)	Main group n=97	Control group n=99	Statistics
Gender, n (%)			
F	53 (27.04)	53 (27.04)	χ2=0.000, p=0.991
M	44 (22.45)	46 (23.47)	
Age, years old	72.0 (65.0, 78.0)	71.0 (64.0, 79.0)	H=0.007, p=0.933
ECG time registration, hours	22.7 (21.9, 23.2)	22.3 (21.4, 23.1)	H=1.376, p=0.120
Time of recorded AF, seconds	81.0 (14.0, 8925.0)	0.0 (0.0, 0.0)	H=167.876, p<0.001
Heart rate in AF	107.0 (94.0, 117.0)	Not applicable	H=167.142, p<0.001
Supraventricular paroxysmal tachycardia, seconds	5.0 (0.0, 34.0)	0.0 (0.0, 9.0)	H=9.840, p=0.002
Ventricular paroxysmal tachycardia, seconds	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	H=0.099, p=0.753
Supraventricular extrasystoles	633.0 (284.0, 2098.0)	79.0 (27.5, 349.5)	H=40.635, p<0.001
Single supraventricular extrasystoles	461.0 (238.0, 1767.0)	69.0 (22.5, 315.0)	H=37.952, p<0.001
Coupled and group supraventricular extrasystoles	26.0 (6.0, 93.0)	4.0 (1.0, 12.0)	H=38.272, p<0.001
Allorhythmia in supraventricular extrasystoles	24.0 (1.0, 401.0)	0.0 (0.0, 9.5)	H=31.647, p<0.001
Atrial extrasystoles	602.0 (262.0, 2028.0)	74.0 (27.5, 314.0)	H=36.843, p<0.001
Ventricular extrasystoles	24.0 (3.0, 149.0)	16.0 (3.0, 488.5)	H=0.176, p=0.675
Single ventricular extrasystoles	20.0 (2.0, 143.0)	14.0 (2.0, 328.5)	H=0.121, p=0.727
Coupled ventricular extrasystoles	0.0 (0.0, 6.0)	0.0 (0.0, 2.0)	H=2.100, p=0.147
Allorhythmia in ventricular extrasystoles	0.0 (0.0, 0.0)	0.0 (0.0, 3.0)	H=0.193, p=0.661
Extrasystoles type «R on T»	2.0 (1.0, 5.0)	0.0 (0.0, 0.0)	H=74.763, p<0.001
Extrasystoles type «P on T», n (%)	95 (97.9%)	4 (4.0%)	χ2= 172.81, p<0.001
QT interval max, seconds	0.5 (0.5, 0.6)	0.5 (0.5, 0.6)	H=2.495, p=0.114
RR interval max, seconds	1.7 (1.6, 1.9)	1.5 (1.4, 1.8)	H=14.172, p<0.001
Loss of QRS complex	0.0 (0.0, 2.0)	0.0 (0.0, 0.0)	H=11.732, p=0.001
ST depression, n (%)	13 (6.63%)	8 (4.08%)	χ2=2.377, p=0.305
Pacemaker, n (%)	2 (1.02%)	0 (0%)	χ2=0.526, p=0.468

AF – atrial fibrillation; ECG – electrocardiography; F – feminine; M – masculine.

reduced the probability of PAF, with this effect also being statistically significant (p=0.021). Moreover, a greater number of episodes of ventricular allorhythmia was associated with a lower risk of PAF (OR=0.997) (see Figure 1).

Based on these findings, we formulated a regression equation to estimate the probability of PAF in patients with sinus rhythm. We identified the following as the most significant risk predictors for developing PAF:

1) The presence of early "P on T" premature ectopic complexes, which had a highly significant impact (p<0.0001). When "P on T" ectopic complexes were present, the likelihood of developing PAF increased 8461 times compared to their absence, highlighting the strong predictive value of this factor.

2) Coupled premature ventricular ectopic complexes, where an increase in their number correlated with a reduced probability of PAF, also with statistical significance (p=0.021).

3) Ventricular allorhythmias. A higher incidence of these episodes during ventricular ectopic activity was linked to a diminished risk of PAF (OR=0.997).

Other analyzed factors, including the frequency of both atrial and atrioventricular premature ectopic complexes, as well as single ventricular ectopic complexes, did not exhibit a statistically significant effect on the risk of PAF, according to this model.

The proposed assessment model demonstrated a high accuracy in forecasting. We validated our proposed model on the same cohort of the analyzed patients due to the principles of evidence-based statistical analysis. The ROC curve analysis revealed that the area under the curve (AUC) was 0.996, with an optimal risk coefficient for predicting PAF set at 0.5, resulting in a prediction accuracy of 97.45%. We named the developed equation "Arfa" (see Figure 2).

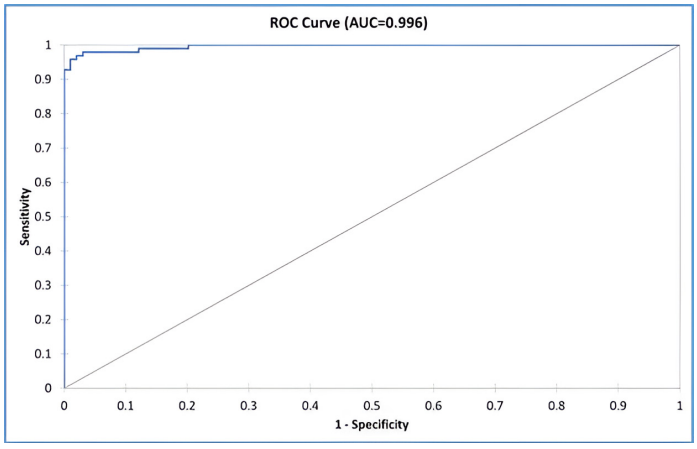


Figure 2 – ROC-curve of "Arfa" (AUC=0,996)

The final regression equation, based on the evaluation of data from 24-hours Holter ECG monitoring, incorporates the following criteria: gender, the number of atrial and atrioventricular supraventricular complexes, the count of single and paired ventricular complexes, variations in rhythms with ventricular complexes, as well as the presence or absence of early "P on T" extrasystoles. This predictive tool for the onset of PAF in patients with sinus rhythm, utilizing the "Arfa" equation, recommends the implementation of the following regression equation:

Table 2 Model parameters (groups of variables)

Parameter	Importance	Standard error	Wald Chi-Square	Pr>Chi²	Wald Lower bound (95%)	Wald Upper bound (95%)	Odds ratio	Odds ratio Lower bound (95%)	Odds ratio Upper bound (95%)
Atrial extrasystoles	0.001	0.000	3.966	0.046	0.000	0.001	1.000677172	1.000010727	1.001344061
Ventricular extrasystoles	-0.001	0.003	0.070	0.791	-0.007	0.005	0.9991973805	0.9932764343	1.005153622
Single ventricular extrasystoles	-0.002	0.002	0.652	0.419	-0.006	0.003	0.9981502839	0.9936747847	1.002645941
Coupled ventricular extrasystoles	0.017	0.015	1.211	0.271	-0.013	0.046	1.016670605	0.9871675159	1.047055442
Ventricular allorhythmias	-0.006	0.005	1.477	0.224	-0.016	0.004	0.9940329032	0.9844853485	1.00367305
Gender - M	-2.040	3.226	0.400	0.527	-8.362	4.283	0.1300607651	0.0002334862027	72.44883181
«P on T» extrasystoles	17.475	11.945	2.140	0.143	-5.936	40.887	38857742.95	0.002642938799	5.71305E+17

M – masculine.

Table 3 Standard coefficients (Variable group)

Parameter	Importance	Standard error	Wald Chi-Square	Pr>Chi²	Wald Lower bound (95%)	Wald Upper bound (95%)	PL Lower bound (95%)	PL Upper bound (95%)
Atrial extrasystoles	1.406	0.706	3.966	0.046	0.022	2.789	0	0
Ventricular extrasystoles	-0.648	2.446	0.070	0.791	-5.443	4.147	0	0
Single ventricular extrasystoles	-0.982	1.216	0.652	0.419	-3.366	1.402	0	0
Coupled ventricular extrasystoles	1.654	1.503	1.211	0.271	-1.292	4.601	0	0
Ventricular allorhythmias	-1.706	1.404	1.477	0.224	-4.458	1.045	0	0
Gender - M	-0.559	0.884	0.400	0.527	-2.291	1.173	0	0
«P on T» extrasystoles	4.817	3.292	2.140	0.143	-1.636	11.270	0	0

M – masculine.

Figure 3 – Standard coefficients (95% confidential interval)

P on T – extrasystoles type "P on T"; Paried VE – paired premature ventricular contractions; Single VE - single premature ventricular contractions; AV Ectopics – supraventricular extrasystoles; Atrial Ectopics – atrial extrasystoles.

Pr (Arfa) = 1 / (1 + EXP (- (- 3,65982 + 0,000144 * Atrial ES + 0,000354 * Atrioventricular ES + 0,000842 * Single ventricular ES - 0,01613 * Couple ventricular ES - 0,002863 * Ventricular allorhythmias - 1,596521 * Gender m + 9,043299 * "P on T" ES))).

Notes: ES – premature complexes, gender f = 0, m = 1; "P on T" ES = 1, absence of «P on T» ES = 0.

Hosmer-Lemeshow test:

Pr (Arfa) = 1 / (1 + EXP (- (- 4.428948 + 0.000677 * Atrial ES - 0.000803 * Atrioventricular ES - 0.001851 * Single ventricular ES + 0.016533 * Couple ventricular ES - 0.005985 * Ventricular allorhythmias - 2.039754 * Gender m + 17.475418 * "P on T" ES -1))).

Notes. ES – premature complexes.

The model parameters are detailed in Tables 2 and 3, as well as in Figure 3.

The risk of developing PAF in patients with sinus rhythm can be assessed using the "Arfa" regression equation as follows:

If Pr (Arfa) > 0.5, there is a high risk of PAF, in which case we recommend treating ectopic complexes and subsequently

monitoring with a 24-hour Holter ECG. Conversely, if Pr (Arfa) < 0.5, the risk of developing PAF is considered low.

Discussion

PAF remains a significant clinical challenge, given its association with increased morbidity and mortality. The development of predictive models for PAF in patients maintaining sinus rhythm plays crucial role for identifying the individuals at heightened risk who may benefit from preventive strategies. Earlier investigations demonstrated the utility of various clinical, demographic, and electrophysiological factors in predicting PAF recurrence [19, 20]. Moreover, the incorporation of machine learning algorithms appears promising, providing enhanced predictive power over traditional models [21].

Recent literature emphasizes the role of clinical risk factors such as arterial hypertension, heart failure, and age as key contributors to the development of PAF [5, 8]. In the publications, it was demonstrated that advancing age is a significant predictor, with a clear relationship between age and the likelihood of developing PAF observed. Older adults exhibited a higher burden of AF due to the structural and electrical remodeling in the heart [11, 12].

Furthermore, comorbidities, such as obesity and diabetes mellitus significantly enhance the risk of PAF, reinforcing insights from several recent studies. Obesity, in particular, affects atrial structure through increased stretch and pressure, thereby facilitating the onset of arrhythmias [22, 23].

Another noteworthy aspect of our predictive modeling involves the integration of sleep apnea as a potential risk factor for PAF [24]. Nocturnal hypoxia and sympathetic overactivity stemming from obstructive sleep apnea may lead to atrial remodeling and increased arrhythmogenicity. Although this study did not focus solely on obstructive sleep apnea, the patterns suggest that screening patients for sleep-disordered breathing could be a vital component of the risk stratification strategy in those with AF.

Lifestyle factors also surfaced as significant predictors of PAF. So, smoking and alcohol consumption have been implicated in the development of PAF. Recent studies suggest that modifiable lifestyle can significantly reduce the risk of AF developing [9, 11, 12]. The advent of wearable technology and digital health has revolutionized how we monitor patients with potential arrhythmias. Continuous heart rhythm monitoring facilitates plays an important role in early detection of PAF episodes and offers the potential benefit for personalized strategies [13, 18].

Despite above-mentioned advances, there remain limitations and contradictions. There is still no universal model for predicting the development of AF that could be used in the everyday clinical practice.

The routine method of 24-hours ECG monitoring belongs to the highly accurate detection method of heart arrhythmias, including PAF. However, before AF paroxysm of occurs, other ECG changes are observed that precede this arrhythmia. If they are identified, the risk of PAF in the patient is higher. Primarily, we are referring to supraventricular premature complexes, especially early ones, “P on T” type. To explain this phenomenon, we believe it is important to consider not only the electrophysiological mechanisms, but also the intra-heart hemodynamics, which also play a crucial role in the development PAF. In our previously published works, we studied the role of heart biomechanics and intra-arterial hemodynamics in arrhythmias, including AF [25–27].

When “P on T” type of premature complexes appear on the ECG, the P wave of the extrasystolic complex lands on the descending limb of the T wave. What mechanisms are taking part during this moment in the biomechanical cardiac cycle? At this moment, the atrioventricular valvular leaflets are closed. If an extrasystole occurs, it catches the atria in a state isolated from the ventricles. In response to the electrical stimulus, a kind of isometric contraction of the atria occurs. Blood, remaining incompressible, exerts additional mechanical pressure on the contracting atria, leading to their stretching. Frequent early atrial premature complexes type “P on T” result in repeated mechanical impacts on the walls of the atria, causing further stretching, altering their morphology and thereby leading to the formation of AF.

Importantly, using the advanced imaging techniques, such as and echocardiography or cardiac magnetic resonance (MRI), helps to improve the risk prediction models. These modalities allow for the assessment of left atrial volume and function – parameters that have been linked to PAF. Integrating these imaging-derived metrics into predictive algorithms could enhance risk stratification, ensuring that we prioritize management for high-risk individuals. In light of this, we believe it is promising to study the stiffness of the left atrium using speckle tracking echocardiography as an additional marker for PAF. In our research, we are planning to include these parameters in our further publications.

As the early “P on T” premature ectopic complexes ($p<0.0001$) are the most significant risk predictor, the patients who have these kind of arrhythmia are highly possible to have already AF paroxysms, not diagnosed. So, to this category of patients can be recommended to perform three days or seven days long ECG monitoring to reveal the arrhythmia.

We believe that the results of our study will contribute to the development of personalized treatment strategies for patients with “P on T” type extrasystoles or PAF initiated by them. Concept of PAF pathophysiological mechanisms will undergo a transformation, taking into account the biomechanical cardiac cycle. Further application of new approaches of risk stratification in clinical practice could provide new data for formulating new strategies for the AF prevention.

Limitations

1) The study is single-center, case-control design. This design introduces a significant risk of selection bias, as the recruited cases (patients with PAF) and controls may not be fully representative of the broader population. The specific practices, patient demographics, and referral patterns of our center limit the generalizability of our results.

2) Limited cohort size, particularly of patients with PAF. Despite the primary cohort included 6,630 patients, in the PAF group it were included 97 of them. A small sample size, particularly of the outcome group, reduces the statistical power of the study. This increases the likelihood of Type II errors (failing to identify true predictor variables) and limits the complexity of the predictive model that can be reliably developed.

3) Potential for model overfitting, as suggested by high AUC. An overfitted model performs well on the data it was trained on but is likely to perform poorly and, greatly diminishing its clinical utility.

4) Lack of external validation. The predictive model was only validated internally using techniques like cross-validation on the original dataset. Internal validation, while valuable, is insufficient to prove the model's robustness and general applicability. This is the most critical step required to transition a research model into a clinically usable tool, and its absence is a major limitation.

In summary, these limitations collectively urge caution in interpreting the model's performance as definitive. The findings should be considered hypothesis-generating. Future prospective, multi-center studies with larger, more diverse cohorts and rigorous external validation are essential to confirm the model's validity and assess its true potential for clinical deployment.

Conclusion

The final regression equation, based on the evaluation of data from 24-hours Holter ECG monitoring, incorporates the following criteria: gender, the number of atrial and atrioventricular supraventricular complexes, the count of single and paired ventricular complexes, variations in rhythms with ventricular complexes, as well as the presence or absence of early "P on T" complexes (AUC=0.996).

We believe that the proposed predictive model appears promising. However, it requires validation in larger cohorts and preferably in prospective or randomized studies before clinical implementation. In the future, it can be helpful to use of the proposed “Arfa” equation in the everyday clinical work of general practitioners, cardiologists and vascular surgeons as the tool that can be easily implemented as the desktop calculator.

Author Contributions: Conceptualization, A. G.; methodology, O. G. and A. G.; validation, O. G. and Y. R.; formal analysis, G. G.; investigation, O. G. and Y. R.; resources, Y. R.; data curation, A. G.; writing – original draft preparation, O. G.; writing – review and editing, A. G. and G. G.; visualization, Y. R.; supervision, I. P.; project administration, O. G. and I. P.; funding acquisition, Y. R. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: Authors express the gratitude to Professor Alexander Kolsanov for the possibility to do the research on the base of Samara State Medical University Clinics; and MD, PhD Timur Syunyakov for the help with statistical analysis.

Funding: None.

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

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Vyas R, Jain S, Thakre A, Thotamgari SR, Raina S, Brar V, Sengupta P, Agrawal P. Smart watch applications in atrial fibrillation detection: Current state and future directions. J Cardiovasc Electrophysiol. 2024;35(12):2474-2482. https://doi.org/10.1111/jce.16451.
2. Desteghe L, Heidbuchel H. Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting: authors' response. Europace. 2017;19(8):1408-1409. https://doi.org/10.1093/europace/euw237.
3. Chiuariu T, Anghel L, Popa DM, Bîrgoan GS, Fechet ȘD, Zanfirescu RL, Balasanian MO, Sascău RA, Stătescu C. Predictors for Device-Detected Subclinical Atrial Fibrillation: An Up-to-Date Narrative Review. J Clin Med. 2026;15(2):578. https://doi.org/10.3390/jcm15020578.
4. Sau A, Sieliwonczyk E, Barker J, Zeidaabadi B, Pastika L, Patlatzoglou K, Khattak GR, McGurk KA, Peters NS, Kramer DB, Waks JW, Ware JS, Ng FS. Prediction of incident atrial fibrillation: A comprehensive evaluation of conventional and artificial intelligence-enhanced approaches. Heart Rhythm. 2026;23(2):e183-e191. https://doi.org/10.1016/j.hrthm.2025.08.024.
5. Manemann SM, Alonso A, Noseworthy PA, Siontis KC, Gersh BJ, Roger VL, Ryu E, Killian JM, Weston SA, Vaughan LE, Chamberlain AM. Multimorbidity in Atrial Fibrillation: Impact on Outcomes. J Am Heart Assoc. 2026;15(5):e040612. https://doi.org/10.1161/JAHA.124.040612.
6. Taha A, Martinsson A, Nielsen SJ, Rezk M, Pivodic A, Gudbjartsson T, Herrmann FEM, Bergfeldt LB, Jeppsson A. New-onset atrial fibrillation after coronary surgery and stroke risk: a nationwide cohort study. Heart. 2024;111(1):18-26. https://doi.org/10.1136/heartjnl-2024-324573.
7. Wu LD, Li F, Chen JY, Zhang J, Qian LL, Wang RX. Analysis of potential genetic biomarkers using machine learning methods and immune infiltration regulatory mechanisms underlying atrial fibrillation. BMC Med Genomics. 2022;15(1):64. https://doi.org/10.1186/s12920-022-01212-0.
8. Lu Z, Ntlapo N, Tilly MJ, Geurts S, Aribas E, Ikram MK, de Groot NMS, Kavousi M. Burden of cardiometabolic disorders and lifetime risk of new-onset atrial fibrillation among men and women: the Rotterdam Study. Eur J Prev Cardiol. 2024;31(9):1141-1149. https://doi.org/10.1093/eurjpc/zwae045.
9. Chung MK, Eckhardt LL, Chen LY, Ahmed HM, Gopinathannair R, Joglar JA, Noseworthy PA, Pack QR, Sanders P, Trulock KM; American Heart Association Electrocardiography and Arrhythmias Committee and Exercise, Cardiac Rehabilitation, and Secondary Prevention Committee of the Council on Clinical Cardiology; Council on Arteriosclerosis, Thrombosis and Vascular Biology; Council on Cardiovascular and Stroke Nursing; and Council on Lifestyle and Cardiometabolic Health. Lifestyle and Risk Factor Modification for Reduction of Atrial Fibrillation: A Scientific Statement From the American Heart Association. Circulation. 2020;141(16):e750-e772. https://doi.org/10.1161/CIR.0000000000000748.
10. Piano MR, Marcus GM, Aycock DM, Buckman J, Hwang CL, Larsson SC, Mukamal KJ, Roerecke M; on behalf the American Heart Association Council on Lifestyle and Cardiometabolic Health; Council on Cardiovascular and Stroke Nursing; Council on Clinical

Cardiology; and Stroke Council. Alcohol Use and Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation. 2025;152(1):e7-e21. https://doi.org/10.1161/CIR.0000000000001341.
11. Lee KY, Lee SR, Choi EK, Choi J, Ahn HJ, Kwon S, Han KD, Oh S, Lip GYH. Cardiovascular benefits of early rhythm control and healthy lifestyle in young atrial fibrillation. Eur J Clin Invest. 2025;55(6):e70018. https://doi.org/10.1111/eci.70018.
12. Deng H, Mei Y, Wu C, Gong C, Lai Z, Huang J, Zheng M, Chen J, Xie Y, Fan H, Wu X, Cai X, Xue Y, Wu S, Liu X. Association of healthy lifestyle and the incidence of atrial fibrillation in senior adults: a prospective cohort study. BMC Geriatr. 2025;25(1):160. https://doi.org/10.1186/s12877-025-05825-9.
13. Atta-Fosu T, LaBarbera M, Ghose S, Schoenhagen P, Saliba W, Tchou PJ, Lindsay BD, Desai MY, Kwon D, Chung MK, Madabhushi A. A new machine learning approach for predicting likelihood of recurrence following ablation for atrial fibrillation from CT. BMC Med Imaging. 2021;21(1):45. https://doi.org/10.1186/s12880-021-00578-4.
14. Zacharia E, Papageorgiou N, Ioannou A, Siasos G, Papaioannou S, Vavuranakis M, Latsios G, Vlachopoulos C, Toutouzas K, Deftereos S, Providência R, Tousoulis D. Inflammatory Biomarkers in Atrial Fibrillation. Curr Med Chem. 2019;26(5):837-854. https://doi.org/10.2174/0929867324666170727103357.
15. Miyasaka Y. Atrial fibrillation progression: what is the impact and how can we intervene? European Heart Journal. 2021;42(29):2872-2881.
16. Gunawardene M, Schmidt B. Management von Risikofaktoren und Begleiterkrankungen bei Vorhofflimmern [Management of Risk Factors and Comobordities in Atrial Fibrillation] [in German]. Dtsch Med Wochenschr. 2025;150(16):945-953. https://doi.org/10.1055/a-2516-8410.
17. Lane DA, Andrade JG, Arbelo E, Boriani G, Hendriks JM, Lee SR, Lip GYH, Mant J, Middeldorp ME. Atrial fibrillation. Lancet. 2026;407(10532):1000-1013. https://doi.org/10.1016/S0140-6736(25)02166-X.
18. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, Boriani G, Castella M, Dan GA, Dilaveris PE, Fauchier L, Filippatos G, Kalman JM, La Meir M, Lane DA, Lebeau JP, Lettino M, Lip GYH, Pinto FJ, Thomas GN, Valgimigli M, Van Gelder IC, Van Putte BP, Watkins CL; ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J. 2021;42(5):373-498. https://doi.org/10.1093/eurheartj/ehaa612.
19. Liu X, Jiang J, Wei L, Xing W, Shang H, Liu G, Liu F. Prediction of all-cause mortality in coronary artery disease patients with atrial fibrillation based on machine learning models. BMC Cardiovasc Disord. 2021;21(1):499. https://doi.org/10.1186/s12872-021-02314-w.
20. Gupta N, Haft JI, Bajaj S, Samuel A, Parikh R, Pandya D, Shamoon F. Role of the combined CHADS2 score and echocardiographic abnormalities in predicting stroke in patients with paroxysmal atrial fibrillation. J Clin Neurosci. 2012;19(9):1242-5. https://doi.org/10.1016/j.jocn.2011.12.008.
21. Xie C, Wang Z, Yang C, Liu J, Liang H. Machine Learning for Detecting Atrial Fibrillation from ECGs: Systematic Review and Meta-Analysis. Rev Cardiovasc Med. 2024;25(1):8. https://doi.org/10.31083/j.rcm2501008.
22. Dziano JK, Ariyaratnam JP, Middeldorp ME, Sanders P, Elliott AD. Obesity and Atrial Fibrillation: From Mechanisms to Treatment. Heart Lung Circ. 2025;34(10):1021-1032. https://doi.org/10.1016/j.hlc.2025.08.003.
23. Sedighi J, Luedde M, Boettger P, Bengel P, Bauer P, Sossalla S, Rozo Sánchez SE, Kostev K. Overweight, obesity and incident atrial fibrillation: Real-world evidence from 400 000 patients in Germany. Diabetes Obes Metab. 2025;27(10):5822-5830. https://doi.org/10.1111/dom.16637.
24. Linz D, Nattel S, Kalman JM, Sanders P. Sleep Apnea and Atrial Fibrillation. Card Electrophysiol Clin. 2021;13(1):87-94. https://doi.org/10.1016/j.ccep.2020.10.003.
25. Kunts LD, Germanova OA, Reshetnikova YB, Galati G, Milevskaya IV, Biondi-Zoccai G. Extrasystolic arrhythmia as an atrial fibrillation predictor [in Russian]. Science and Innovations in Medicine. 2024;9(2):117-123. https://doi.org/10.35693/SIM624503.
26. Germanova OA, Galati G, Kunts LD, Usenova AA, Reshetnikova YB, Germanov AV, Stefanidis A. Predictors of paroxysmal atrial fibrillation: Analysis of 24-hour ECG Holter monitoring [in Russian]. Science and Innovations in Medicine. 2024;9(1):44-48. https://doi.org/10.35693/SIM626301.
27. Germanova O, Galati G, Germanov A, Stefanidis A. Atrial fibrillation as a new independent risk factor for thromboembolic events: hemodynamics and vascular consequence of long ventricular pauses. Minerva Cardiol Angiol. 2023;71(2):175-181. https://doi.org/10.23736/S2724-5683.22.06000-8.

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article
DOI: https://doi.org/10.23950/jcmk/18598

Impact of Microablative Fractional CO_2 Laser Applied in Menopausal Period on Vulvovaginal Atrophy and Dyspareunia: A Systematic Review and Meta-analysis Study

Ayşe Çuvadar[1], Handan Özcan[2], Yeter Çuvadar Baş[3]

[1]Department of Midwifery, Faculty of Health Sciences, Karabuk University, Karabuk, Turkey
[2]Department of Midwifery, Faculty of Health Sciences, University of Health Sciences, Istanbul, Turkey
[3]Department of Paramedic, Gedik Vocational School, T.C. Gedik University, Istanbul, Turkey

Received: 2026-02-04.
Accepted: 2026-05-22.

J Clin Med Kaz 2026; 23(3): 31-39

Corresponding author:
Ayşe Çuvadar
E-mail: aysecuvadar@karabuk.edu.tr.
ORCID: 0000-0002-7917-0576.

ABSTRACT

Aim: This study aimed to synthesize current evidence on the clinical effectiveness of microablative fractional CO_2 laser ($MFCO_2$) therapy for the management of menopausal vulvovaginal atrophy and dyspareunia.

Methods: A meta-analysis was conducted to identify relevant studies published within the last ten years. Electronic database searches were performed between March and June 2024 using PubMed, EBSCOhost, Web of Science, Google Scholar, and the YÖK National Thesis Center. Following the screening process and eligibility evaluation according to established inclusion criteria, eight studies were ultimately retained for analysis. The methodological rigor of the selected studies was evaluated using design-specific critical appraisal instruments developed by the Joanna Briggs Institute (JBI). Quantitative data synthesis was performed using CMA Version 2 software, and findings were interpreted using both statistical meta-analytic techniques and descriptive synthesis.

Results: The pooled analysis demonstrated that $MFCO_2$ laser therapy produced a statistically significant improvement in vulvovaginal atrophy among menopausal women (SMD: 1.437, 95% CI: 0.646–2.228; Z = 3.559, p < 0.001; I^2 = 94.93%). Similarly, treatment was associated with a significant reduction in dyspareunia severity (SMD: −1.820, 95% CI: −3.063 to −0.577; Z = −2.871, p = 0.004; I^2 = 96.17%). These findings indicate that $MFCO_2$ laser therapy may contribute to meaningful symptom improvement in menopausal genitourinary disorders. However, considerable heterogeneity among included studies suggests variability in intervention protocols, patient characteristics, and outcome assessment methods.

Conclusion: Microablative fractional CO_2 laser therapy may serve as a beneficial non-hormonal intervention for improving menopausal genitourinary symptoms, especially dyspareunia among sexually active women. Nevertheless, large-scale clinical trials with uniform treatment protocols and longer follow-up are needed to confirm long-term therapeutic efficacy and safety.

Keywords: Dyspareunia, microablative fractional CO_2 laser, vulvovaginal atrophy, menopause.

Introduction

Vulvovaginal atrophy (VVA) is currently classified within the clinical spectrum of genitourinary syndrome of menopause (GSM) and represents one of the most common consequences of estrogen depletion following menopause. The condition is characterized by progressive

structural and functional changes in vulvovaginal tissues. It is commonly associated with vaginal dryness, vulvovaginal irritation or discomfort, pruritus, pain during sexual activity, and urinary symptoms [1,2]. These symptoms may gradually intensify over time and can substantially interfere with sexual health, interpersonal relationships, and psychosocial well-being. Epidemiological data indicate that symptoms associated with GSM affect nearly half of postmenopausal women and are closely linked to impaired sexual functioning and decreased overall quality of life [3,4].

A variety of therapeutic strategies have been developed to manage GSM, including both hormonal and non-hormonal interventions. Non-hormonal therapies, particularly vaginal moisturizers and lubricants, are commonly recommended as supportive measures for symptom relief; however, these products typically provide short-term improvement and do not directly address underlying tissue degeneration. In contrast, local vaginal estrogen therapy has been shown to restore epithelial integrity and improve symptom severity over longer treatment durations and is widely considered an effective therapeutic option [5,6]. Nevertheless, hormonal therapies may not be appropriate for all patients due to contraindications, safety concerns, or personal treatment preferences. Consequently, increasing attention has been directed toward alternative therapeutic modalities that aim to provide effective symptom control while minimizing systemic exposure. Among emerging treatment approaches, energy-based technologies, particularly laser therapies, have gained significant clinical interest due to their capacity to stimulate tissue remodeling and support restoration of normal vaginal physiology [7].

Carbon dioxide (CO_2) laser technology, among the earliest gas-based laser systems developed for medical applications, has been widely used across various clinical fields, including gynecology [8,9]. Early clinical investigations into fractional vaginal CO_2 laser therapy were reported by Gaspar et al. in 2011, demonstrating improvements in both clinical symptomatology and histological characteristics of vaginal atrophy [10]. Subsequent observational studies have further supported these findings. For example, Salvatore and colleagues evaluated symptom outcomes following three monthly sessions of $MFCO_2$ laser therapy over 12-weeks period in a cohort of 49 women. Their findings demonstrated improvement in vaginal dryness (86.0%), vaginal burning (90.0%), vaginal itching (80.0%), and dyspareunia (74.0%) [11]. In addition, improvement in dyspareunia was reported among all sexually active women included in the study [12].

Despite increasing scientific and clinical interest in laser-based treatments, much of the existing literature consists of small-scale, short-term pilot studies, which limit the strength of the available evidence [13]. Currently, three main non-surgical energy-based therapeutic modalities are utilized in the management of menopausal vulvovaginal symptoms: fractional $MFCO_2$ laser therapy, erbium laser applications, and temperature-controlled radiofrequency treatments. Available research suggests that vaginal laser therapies, including both erbium and CO_2 laser techniques, may offer clinically meaningful improvements in symptom severity and vaginal tissue health [14].

Fractional CO_2 laser therapy is believed to promote regeneration of vaginal tissues through several biological mechanisms. These mechanisms include stimulation of collagen synthesis, increased glycogen production, thickening of the epithelial layer, and enhanced of local vascularization, all of which contribute to the restoration of vaginal mucosal integrity toward a premenopausal-like condition [15]. Although clinical improvements following laser therapy have been reported, the duration of therapeutic benefits remains uncertain, with current evidence suggesting symptom improvement may persist for approximately 12 months in many patients [14]. In addition to histological improvements, fractional CO_2 laser treatment has been associated with reductions in both patient-reported symptom severity and objective clinical findings, further supporting its therapeutic potential [16].

While recent investigations have indicated potential therapeutic benefits of laser interventions in GSM, the overall coherence and robustness of the available findings remain limited. Variations in study design, patient characteristics, intervention protocols, outcome measurement tools, and follow-up durations contribute to heterogeneity among published findings. Furthermore, while individual clinical studies have reported encouraging outcomes, comprehensive quantitative synthesis focusing specifically on the effectiveness of $MFCO_2$ laser therapy in the treatment of VVA and dyspareunia remains limited. Therefore, the present meta-analysis was conducted to evaluate the effectiveness of $MFCO_2$ laser therapy administered during the menopausal period in alleviating VVA and dyspareunia symptoms. In addition, this study aimed to synthesize available scientific evidence to support clinical decision-making processes and to identify areas requiring further high-quality research.

The hypotheses of the study

Primary Hypothesis: $MFCO_2$ laser therapy applied during the menopausal period significantly improves vulvovaginal atrophy in postmenopausal women.

Secondary Hypothesis: $MFCO_2$ laser therapy significantly reduces dyspareunia severity in menopausal women.

Additional Hypothesis: $MFCO_2$ laser therapy represents a safe and effective non-hormonal therapeutic alternative for the management of GSM.

Methods

This study was designed as a systematic review and meta-analysis and followed the PRISMA reporting standards (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [17]. To enhance methodological rigor and reduce the risk of bias, the processes of literature searching, study selection, and data extraction were carried out independently by four investigators (A.Ç., H.Ö., Y.Ç.B.). Each stage of the review was conducted separately by the assigned reviewers to ensure consistency and objectivity.

Following independent evaluations, the reviewers compared their decisions, and any disagreements were resolved through discussion until consensus was achieved. In addition, the number of screening steps performed for each record and the reviewers responsible for each phase were systematically documented to ensure transparency throughout the review process. The methodological quality of the studies included in the meta-analysis was also evaluated independently by the same research team, thereby strengthening the reliability and validity of the synthesized findings.

Study Selection and Eligibility Requirements

The studies were screened according to the PICOS framework:

P: Patient: Postmenopausal women with symptom/s

I: Intervention: Microablative fractional CO_2 laser ($MFCO_2$ laser) therapy

C: Comparison: No treatment, placebo, sham treatment, or other non-laser-based therapeutic approaches (e.g., hormonal or non-hormonal treatments), depending on the design of the included studies

O: Outcomes: Vulvovaginal atrophy, dyspareunia

S: Study design: Studies designed using experimental or quasi-experimental methodologies and published in either Turkish or English were considered eligible.

Letters to the editor, case reports, case presentations, and studies of a systematic or narrative review nature were excluded from the scope of this research.

Literature Identification Strategy

Between March and June 2024, a search was conducted using the keywords "carbon dioxide laser", "CO_2 Laser", "vulvovaginal atrophy", "vaginal atrophy", "genitourinary syndrome of menopause", and "GSM" in PubMed, Web of Science, EBSCOhost, YOK National Thesis Center, and Google Scholar. The studies were then transferred to Mendeley. In order to review the current literature, studies from the past 10 years were included in the search.

Study Screening and Selection

The initial search resulted in 5359 records. After duplicates and irrelevant studies were removed, 5249 records were screened for selection based on abstract and title. As a result of this screening, 64 studies were selected for full-text review. Subsequently, the 64 articles were reviewed according to inclusion and exclusion criteria for the application of $MFCO_2$ laser in menopausal VVA and dyspareunia, with 8 studies reporting the effects included in the analysis. The flow of study identification and inclusion is shown in Figure 1.

Data Retrieval and Extraction

A structured data extraction template developed by the research team was used to collect study data. This template facilitated the documentation of study characteristics such as author name, publication year, research design, study location, assessment tools employed, and sample size (Table 1). The extraction process was performed independently by two investigators using an identical standardized form. Any disagreements were resolved through discussion until agreement was reached. This systematic methodology enhanced both the transparency and reproducibility of the study.

Ethical Principles

This study employed a meta-analytic approach by synthesizing findings from previously published research.

Quality Evaluation of Included Studies

The methodological quality of the studies incorporated into this systematic review and meta-analysis was examined using study design–specific critical appraisal tools developed by the JBI [18]. The selection of appraisal instruments was based on the methodological structure of each study. Randomized controlled trials were evaluated using a 13-item checklist, whereas quasi-experimental studies were assessed using a 9-item checklist [19]. The checklists provide four response options: “Yes,” “No,” “Unclear,” and “Not Applicable.” Two independent researchers conducted the quality assessment process, and disagreements were resolved through discussion until consensus was obtained. The outcomes of the quality assessment are presented in Table 1 as Quality Scores.

Synthesis of the Collected Data

Statistical calculations were performed using the Comprehensive Meta-Analysis program (CMA, Version 2). Statistical heterogeneity across the included studies was

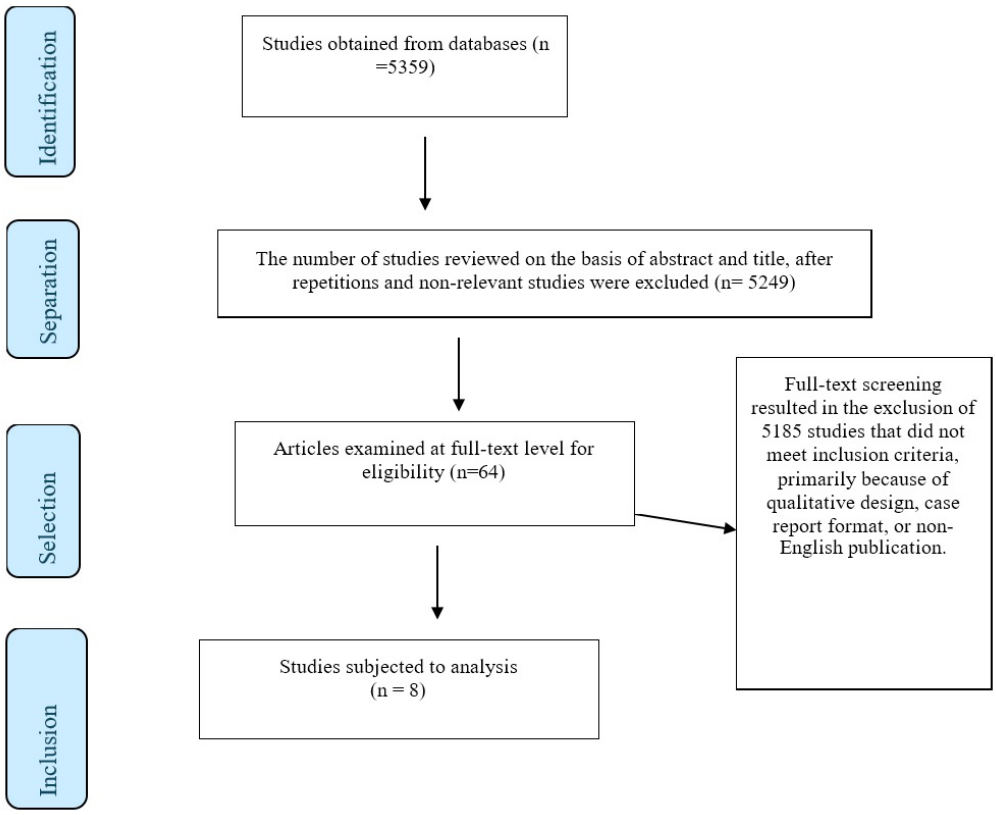


Figure 1 – Selection of studies according to PRISMA flow diagram

Table 1 Characteristics and results of the included studies

Author/Year	Study design	Sample size	Scale	Outcomes	Patient population	Quality Score
Eder SE. et al., 2018	Quasi-experimental	28 participants	Vaginal Health Index (VHI)	CO_2 laser therapy may represent an effective treatment alternative for postmenopausal women, as symptom improvement has been reported even after a single treatment session.	Postmenopausal women, dryness, itching, burning, dysuria or dyspareunia).	Yes:7/9 No:2/9 Uncertain:0/9 Not applicable:0/9
Pitsouni E. et al., 2016	Quasi-experimental	53 participants	Vaginal Health Index Score (VHI)	The findings of this study indicate that intravaginal CO_2 laser therapy may be a beneficial treatment option for postmenopausal women presenting with clinical manifestations of genitourinary syndrome of menopause (GSM), contributing to improvements in both vaginal tissue physiology and symptom severity. Furthermore, women in both comparison groups demonstrated enhanced sexual function and reduced levels of sexual distress.	Postmeno-pausal women with symptom/s (dyspareunia, genital dryness, burning, itching, dysuria, urinary frequency, urgency)	Yes:7/9 No:2/9 Uncertain:0/9 Not applicable:0/9
Ruanphoo P. et al., 2020	Randomized Control	Experimental group:44 Control group:44	Vaginal Health Index (VHI) Visual Analog Scale (VAS)	The findings of this study indicate that microablative fractional CO_2 laser therapy is associated with improvements in vaginal atrophy. This approach may represent a promising alternative treatment option for postmenopausal women experiencing this condition.	Postmeno-pausal women, dyspareunia, dryness, itching, burning, dysuria	Yes: 10/13 No:2/13 Uncertain:0/13 Not applicable: 1/13
Salvatore et al., 2015	Quasi-experimental	77 participants	Visual Analog Scale (VAS)	Fractional microablative CO_2 laser therapy has been associated with significant improvements in sexual function and overall sexual satisfaction among postmenopausal women experiencing symptoms of vulvovaginal atrophy (VVA).	Postmeno-pausal women, dyspareunia, dryness, itching, burning, dysuria	Yes:7/9 No:2/9 Uncertain:0/9 Not applicable:0/9
Salvatore S. et al., 2021	Randomized Control	Experimental group:28 Control group:30	Visual Analog Scale (VAS)	The findings of this study indicate that CO_2 laser therapy provides significant and sustained improvement in GSM-related symptoms. Moreover, compared with sham interventions, CO_2 laser treatment may serve as an effective alternative therapeutic option in the management of genitourinary syndrome of menopause.	Postmenopausal women diagnosed with GSM and bothersome dryness and dyspareunia	Yes: 10/13 No:2/13 Uncertain:0/13 Not applicable: 1/13
Siliquini GP. et al., 2017	Quasi-experimental	87 participants	Vaginal Health Index (VHI)	The results of this study indicate that CO_2 laser therapy is associated with significant and sustained symptom improvement.	Postmeno-pausal women, dyspareunia, dryness, itching, burning, dysuria	Yes:7/9 No:2/9 Uncertain:0/9 Not applicable:0/9
Sophie Page A. et al., 2022	Randomized Control	Experimental group grubu:28 Control group:29	Vaginal Health Index Visual Analog Scale (VAS)	Among women with GSM, treatment outcomes evaluated 12 weeks after laser therapy were found to be comparable to those observed following sham interventions.	Postmenopausal women GSM genitourinary syndrome of menopause	Yes: 10/13 No:2/13 Uncertain: 0/13 Not applicable: 1/13
Tenerowicz AR. et al., 2022	Quasi-experimental	205 participants	Vaginal Health Index (VHI)	Ablative CO_2 laser therapy may contribute to the reduction of vulvovaginal atrophy symptoms, including vaginal laxity, dryness, dyspareunia, and burning sensation, and may also lessen the severity of stress and urge urinary incontinence.	Perimenopausal dryness, dyspareunia, burning, vaginal laxity, urinary incontinence	Yes:7/9 No:2/9 Uncertain:0/9 Not applicable:0/9

examined using Cochran's Chi-square (Q) test and Higgins' I^2 statistic. The I^2 statistic was interpreted to quantify the proportion of total variation attributable to between-study heterogeneity rather than chance. An I^2 value above 50% was interpreted as indicating moderate to substantial heterogeneity, whereas lower values suggested acceptable consistency across studies. In addition, the statistical significance of heterogeneity was evaluated using the Q test, with a p-value threshold of 0.10, as recommended for heterogeneity testing due to the limited statistical power of this test in meta-analyses with small sample sizes.

Model selection was determined according to heterogeneity levels. A fixed-effects model was applied when heterogeneity was considered low ($I^2 \leq 50\%$ and $p > 0.10$), assuming that the included studies estimated a common underlying effect size. In contrast, when substantial heterogeneity was detected ($I^2 > 50\%$), a random-effects model was employed to account for potential variability in true effect sizes across studies [20]. Furthermore, Tau-square (τ^2) statistics were calculated to estimate the variance in true effect sizes among studies and to support the interpretation of between-study variability.

To allow comparison of outcomes measured using different assessment tools, standardized mean differences (SMD) were calculated along with their corresponding 95% confidence intervals (CI). The magnitude and direction of treatment effects were visually presented using forest plots, which illustrated individual study effect sizes and pooled estimates. The overall pooled effect size was obtained by calculating the weighted average of individual study effect sizes. The statistical significance of pooled results was determined by transforming the D statistic into a Z score and evaluating the associated p-value.

Publication bias was evaluated using both graphical and statistical approaches. Funnel plots were visually inspected to

assess potential asymmetry, which may suggest the presence of publication bias or small-study effects. Egger's regression test was additionally performed to provide a quantitative assessment of funnel plot asymmetry. All statistical analyses were performed using two-sided tests, and statistical significance was defined as a p-value less than 0.05 [21].

Results

The final analysis involved three randomized controlled experimental studies and five quasi-experimental studies. Overall, the studies included 100 participants in intervention groups, 103 participants in control groups, and 450 participants evaluated in single-group studies (Table 1).

Across all studies included in this systematic review and meta-analysis, more than half of the items in the quality appraisal checklists were satisfied (Table 1). This result indicates that the evidence synthesized in the present review is derived from studies demonstrating an acceptable level of methodological quality.

Meta-Analysis of $MFCO_2$ laser Effects on Vulvovaginal Atrophy

Two techniques were employed to examine potential publication bias, including graphical evaluation through funnel plot analysis and statistical testing using Egger's regression method [22].

Publication bias within the included studies was evaluated using Egger's regression test. The analysis yielded an intercept (B0) value of 4.21023 with a 95% confidence interval ranging from −6.60736 to 15.02783 (t = 1.08060, df = 4). The two-tailed p-value was calculated as 0.34068, indicating no statistically significant evidence of publication bias among the included studies.

Figure 3 presents the pooled findings of eight studies examining the effects of $MFCO_2$ laser therapy applied during the menopausal period on VVA. In these studies, the Vaginal Health Index Score (VHIS) was used as the primary outcome measure to evaluate vaginal tissue condition. The meta-analytic results demonstrated a statistically significant improvement in vaginal health among women receiving $MFCO_2$ laser therapy (SMD = 1.437; 95% CI: 0.646–2.228; Z = 3.559; $p < 0.001$).

However, considerable variability was observed across the included studies, as indicated by the heterogeneity analysis (I^2 = 94.933%; $p < 0.001$). This high heterogeneity suggests potential differences in study design, patient characteristics, intervention protocols, or outcome assessment methods among the included trials.

Meta-Analysis of $MFCO_2$ laser Effects on Dyspareunia

Potential publication bias among the included studies was assessed using Egger's regression analysis. The calculated

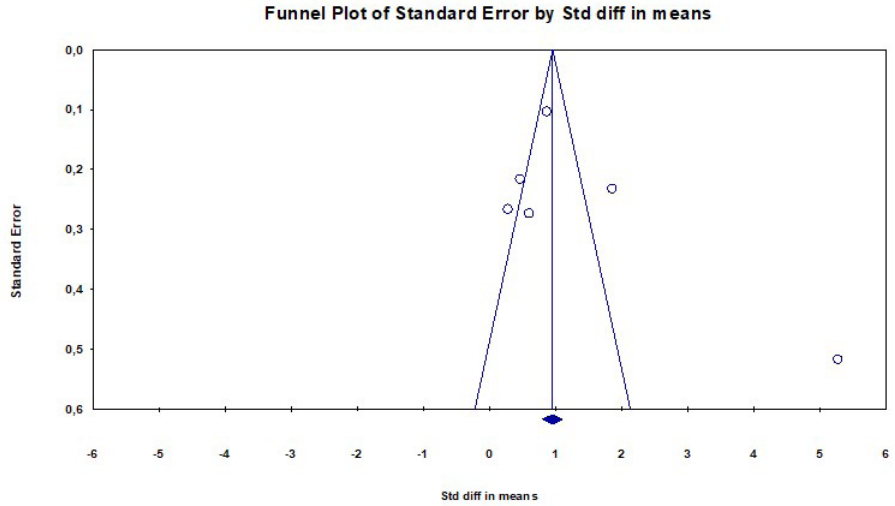


Figure 2 – Funnel plot reporting the results of studies on the effects of $MFCO_2$ laser on VVA

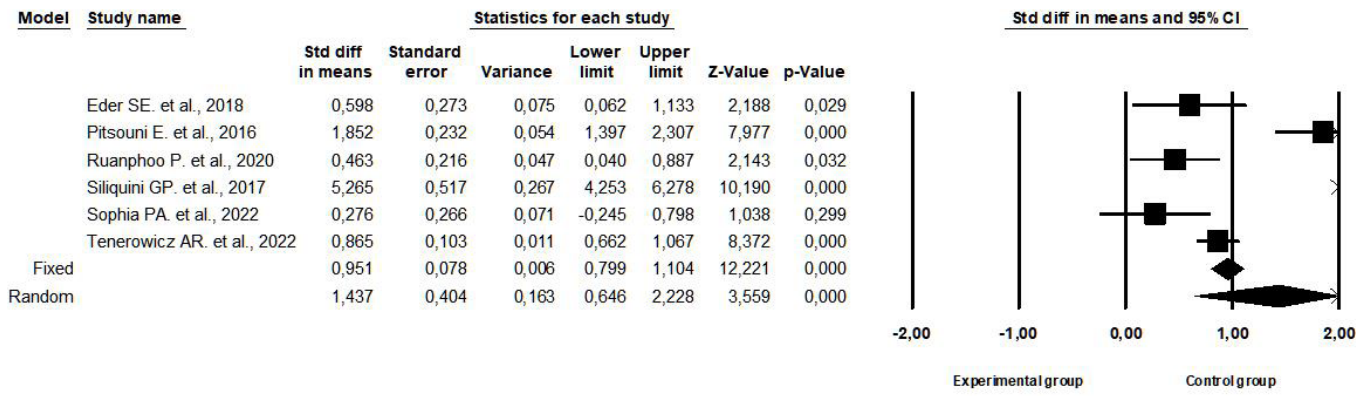


Figure 3 – Forest plot reporting the results of studies on the impact of $MFCO_2$ laser on VVA

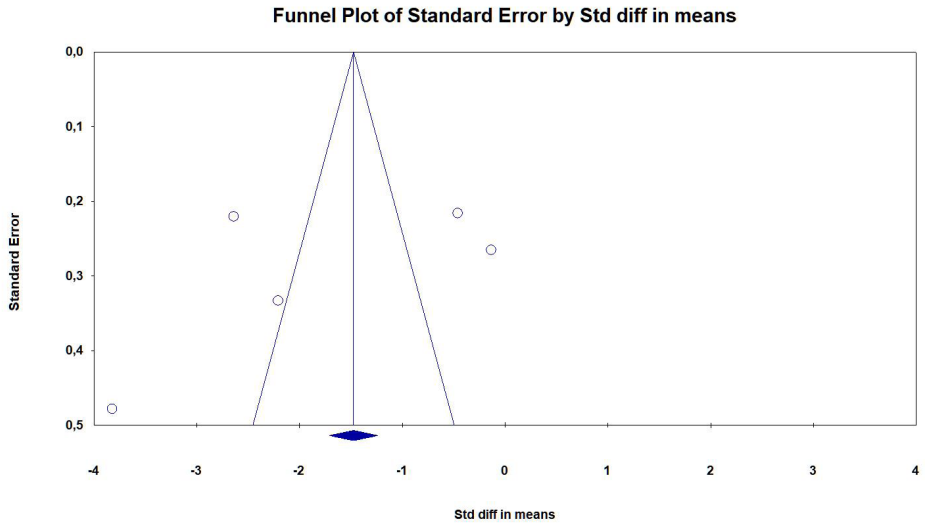


Figure 4 – Funnel plot reporting the results of studies on the effects of $MFCO_2$ laser on dyspareunia

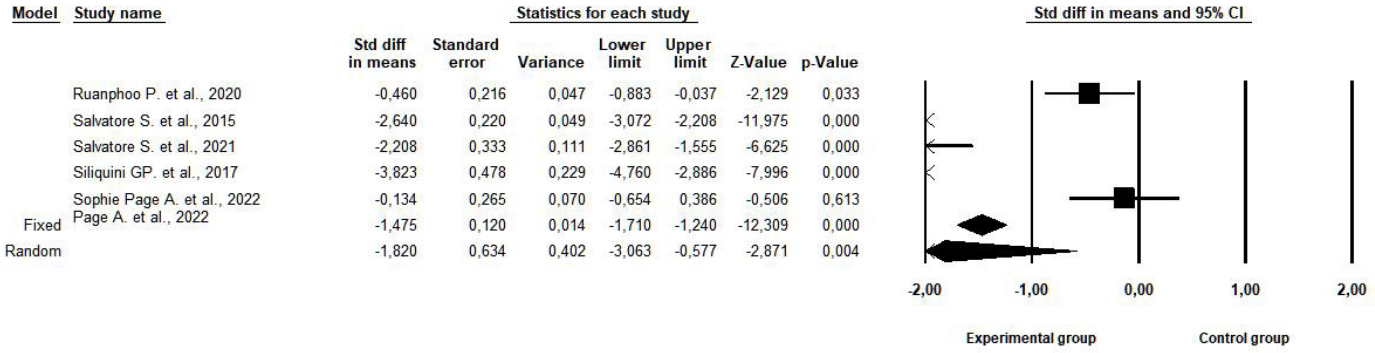


Figure 5 – Forest plot reporting the results of studies on the effect of $MFCO_2$ laser on dyspareunia

intercept (B0) was −7.96890 with a 95% confidence interval ranging from −37.18128 to 21.24349 (t = 0.86815, df = 3). The two-sided p-value was 0.44919, suggesting that there was no statistically significant evidence of publication bias within the analyzed dataset.

Figure 5 illustrates the pooled results of five studies evaluating the effects of $MFCO_2$ laser therapy administered during menopause on dyspareunia. Pain intensity in these studies was assessed using the Visual Analog Scale (VAS). Meta-analytic synthesis of the available data demonstrated a statistically significant reduction in dyspareunia following $MFCO_2$ laser treatment (SMD = −1.820; 95% CI: −3.063 to −0.577; Z = −2.871; p = 0.004).

Despite the observed therapeutic effect, heterogeneity analysis revealed considerable variability among the included studies (I^2 = 96.172%; p = 0.004). This elevated heterogeneity may reflect differences in treatment protocols, participant characteristics, follow-up duration, or methodological approaches used across the studies.

Discussion

GSM is a condition affecting the genital and lower urinary tract after menopause, primarily due to estrogen deficiency, which leads to VVA and related symptoms [31].

In the literature, there are some concerns regarding the safety of CO2 lasers, especially after the FDA's warning against their inappropriate use in the treatment of symptoms related to "vaginal rejuvenation" and sexual function, several important findings regarding the use of energy-based devices have been reported [32]. However, although the findings of the present study suggest that vaginal fractional CO_2 laser treatment is generally well tolerated and no serious complications were reported in the included studies, it should be noted that this meta-analysis primarily focused on symptom improvement and did not include a quantitative synthesis of safety outcomes. Therefore, conclusions regarding safety should be interpreted with caution.

In a meta-analysis conducted by Sarmento et al., CO2 laser has been identified as the most commonly used and scientifically proven effective treatment among physical methods, with the ability to improve all GSM symptoms up to 12 months after treatment [33]. In a meta-analysis study by Prodromidou et al. (2021), laser therapy was found to be effective on subjective and objective symptoms [34]. Liu et al. (2022) demonstrated in their meta-analysis that CO_2 laser therapy contributed to reductions in VVA symptoms and improvements in sexual function among postmenopausal women [2]. The consistency of these findings with our results strengthens the evidence supporting the therapeutic role of CO_2 laser treatment in GSM management. Nevertheless, variations in laser parameters, treatment sessions,

and follow-up durations among studies highlight the need for standardized treatment protocols.

Vulvovaginal atrophic changes may negatively affect women's quality of life and sexual function, highlighting the importance of effective management strategies for GSM symptoms [35,36]. These pathophysiological alterations emphasize the clinical importance of early identification and effective management of GSM symptoms, as untreated conditions may negatively influence not only physical health but also psychosocial well-being and intimate relationships.

In this meta-analysis, it was determined that $MFCO_2$ laser treatment applied during menopausal period has a generally significant effect on dyspareunia. The studies included in this meta-analysis used VAS to assess dyspareunia in most cases. In the systematic review and meta-analysis conducted by Filippini and colleagues, improvement in GSM symptoms before and after laser treatment was found [37]. The improvement in dyspareunia observed in our analysis may be explained by laser-induced stimulation of collagen remodeling, increased vascularization, and restoration of vaginal epithelial integrity. These biological mechanisms may contribute to improved tissue elasticity and lubrication, which are essential factors for reducing pain during sexual intercourse. In contrast to our study findings, Ni and Lian (2024) meta-analysis revealed that CO2 laser treatment did not make a significant difference in terms of dyspareunia, dryness, burning, itching, and dysuria scores on criteria such as GSM, FSFI, VHIS, and VAS [4]. This discrepancy may be related to methodological differences such as sample characteristics, intervention protocols, and outcome assessment methods. Additionally, differences in baseline symptom severity and patient selection criteria across studies may have influenced treatment responsiveness. Salvatore and colleagues' meta-analysis also found that Fractional CO2 laser improved sexual function and reduced pain in menopausal women affected by GSM [38]. In Mension and colleagues' (2022) meta-analysis, it was found that vaginal laser improved VAS and FSFI scores [13]. Taken together, the available evidence suggests that fractional CO_2 laser therapy may positively influence both pain-related and sexual function outcomes; however, the high heterogeneity observed across studies indicates that treatment outcomes may vary depending on clinical and methodological factors. The substantial heterogeneity ($I^2 > 90\%$) observed in this meta-analysis may be attributed to several factors, including differences in study design, variation in $MFCO_2$ laser treatment protocols (e.g., energy settings, number of sessions, and application techniques), heterogeneity in follow-up durations, and differences in outcome assessment methods across studies. These factors may have contributed to variability in the magnitude of the reported effects. Furthermore, although short- and medium-term results appear promising, there remains limited evidence regarding long-term effectiveness and safety. Additionally, the inclusion of both randomized controlled trials and quasi-experimental studies may have contributed to variability in the pooled results, as differences in study design, methodological rigor, and risk of bias can influence the magnitude and consistency of observed effects. Therefore, this methodological diversity should be considered when interpreting the findings.

Conclusion

In conclusion, fractional CO_2 laser therapy appears to be a promising non-hormonal intervention for the management of genitourinary syndrome of menopause, particularly in reducing dyspareunia among postmenopausal women. The findings of this meta-analysis indicate significant improvements in symptoms; however, the presence of substantial heterogeneity and the limited number of included studies warrant cautious interpretation. From a clinical perspective, this approach may be especially relevant for women who cannot or prefer not to use hormonal therapies. Nevertheless, further well-designed randomized controlled trials with larger sample sizes and longer follow-up periods are needed to confirm long-term efficacy, establish standardized treatment protocols, and support its integration into evidence-based clinical practice.

Limitations

Interpretation of the present meta-analytic results requires consideration of certain methodological constraints. In particular, the inclusion of studies with relatively small sample populations and quasi-experimental study designs may reduce the overall reliability and generalizability of the synthesized evidence. Additionally, although the inclusion of studies published within the last decade was intended to reflect contemporary clinical practice, this restriction may have limited the number of eligible studies and potentially excluded relevant earlier research. Another important limitation is the high level of heterogeneity observed among the included studies. Variations in study design, participant characteristics, treatment protocols, laser parameters, follow-up duration, and outcome measurement tools may have contributed to the observed heterogeneity. These differences may affect the comparability of study results and should be considered when interpreting the pooled effect sizes.

Furthermore, the diversity of assessment instruments used to evaluate outcomes such as vaginal health, dyspareunia, and sexual function may have influenced the consistency of the reported findings. Differences in subjective outcome measures may introduce variability related to patient perception and reporting bias. Although publication bias was not statistically significant, the relatively limited number of included studies may reduce the sensitivity of bias- detection methods. Moreover, for certain outcomes, such as dyspareunia, the number of included studies was particularly small, further limiting the robustness of the pooled estimates. In such cases, statistical tests for publication bias (e.g., Egger's test) should be interpreted with caution, as their reliability is reduced when the number of studies is limited. Therefore, the possibility of undetected publication bias cannot be completely excluded. Finally, the lack of long-term follow-up data in several studies restricts the ability to fully evaluate the sustainability of treatment effects and the long-term safety profile of laser therapy. Further high-quality clinical investigations with larger sample sizes, consistent intervention protocols, and longer observation periods are necessary to strengthen the current evidence. These limitations were carefully considered when interpreting the findings. Although they may influence the precision of the estimated effects, the overall direction and consistency of the results support the main conclusions of the study.

Theoretical Contributions

By integrating available research findings, this study offers theoretical insight into the role of $MFCO_2$ laser therapy as a non-hormonal therapeutic option for GSM. The findings support existing theoretical models suggesting that $MFCO_2$ laser treatment promotes vaginal tissue regeneration through collagen

remodeling, epithelial restoration, and enhanced vascularization, which are associated with improvements in dyspareunia and VVA. Additionally, by integrating heterogeneous research findings, this study strengthens the evidence base and supports the development of standardized, patient-centered therapeutic models for menopausal genitourinary symptom management.

Author Contributions: Conceptualization, A.Ç, H.Ö; methodology, A.Ç. H.Ö, Y.Ç.B; validation, A.Ç, H.Ö, Y.Ç.B; formal analysis, A.Ç. H.Ö; investigation, A.Ç. H.Ö, Y.Ç.B; resources, A.Ç. H.Ö, Y.Ç.B; data curation, A.Ç. H.Ö, Y.Ç.B; writing – original draft preparation, A.Ç. H.Ö, Y.Ç.B; writing – review and editing, A.Ç. H.Ö, Y.Ç.B; visualization, A.Ç. H.Ö, Y.Ç.B; supervision, A.Ç. H.Ö, Y.Ç.B; project administration, A.Ç. H.Ö, Y.Ç.B; funding acquisition, A.Ç. H.Ö, Y.Ç.B. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: None.

Data availability statement: The corresponding author can provide the data supporting the study's conclusions upon request. Due to ethical and privacy constraints, the data are not publicly accessible.

Artificial Intelligence (AI) Disclosure Statement: The authors confirm that generative artificial intelligence (AI) or AI-assisted technologies were not used in the design, data analysis, interpretation, writing, or editing of this manuscript. The entire manuscript was developed and finalized by the authors.

References

1. Gandhi J, Chen A, Dagur G, Suh Y, Smith NL, Cali B, Khan SA. Genitourinary syndrome of menopause: An overview of clinical manifestations, pathophysiology, etiology, evaluation, and management. American Journal of Obstetrics & Gynecology. 2016;215(6):704–711. https://doi.org/10.1016/j.ajog.2016.07.045
2. Liu M, Li F, Zhou Y, Cao Y, Li S, Li Q. Efficacy of CO_2 laser treatment in postmenopausal women with vulvovaginal atrophy: A meta-analysis. International Journal of Gynecology & Obstetrics. 2022;158(2):241–251. https://doi.org/10.1002/ijgo.13973
3. Parish SJ, Nappi RE, Krychman ML, Kellogg-Spadt S, Simon JA, Goldstein JA, Kingsberg SA. Impact of vulvovaginal health on postmenopausal women: A review of surveys on symptoms of vulvovaginal atrophy. International Journal of Women's Health. 2013;5:437–447. https://doi.org/10.2147/IJWH.S44579
4. Ni Y, Lian J. Carbon dioxide laser therapy for the management of genitourinary syndrome of menopause: A meta-analysis of randomized controlled trials. Experimental and Therapeutic Medicine. 2024;27:10. https://doi.org/10.3892/etm.2023.12297
5. Palacios S, Mejía A, Neyro JL. Treatment of the genitourinary syndrome of menopause. Climacteric. 2015;18(Suppl 1):23–29. https://doi.org/10.3109/13697137.2015.1079100
6. ACOG Committee Opinion No. 659. The use of vaginal estrogen in women with a history of estrogen-dependent breast cancer. Obstetrics & Gynecology. 2016;127(3):e93–e96. https://doi.org/10.1097/AOG.0000000000001331
7. Gambacciani M, Palacios S. Laser therapy for the restoration of vaginal function. Maturitas. 2017;99:10–15. https://doi.org/10.1016/j.maturitas.2017.01.012
8. Patel CKN. Continuous-wave laser action on vibrational-rotational transitions of CO_2. Physical Review. 1964;136(5A):A1187–A1193. https://doi.org/10.1103/PhysRev.136.A1187
9. Omi T, Numano K. The role of the CO_2 laser and fractional CO_2 laser in dermatology. Laser Therapy. 2014;23(1):49–60. https://doi.org/10.5978/islsm.14-RE-01
10. Gaspar A, Addamo G, Brandi H. Vaginal fractional CO_2 laser: A minimally invasive option for vaginal rejuvenation. American Journal of Cosmetic Surgery. 2011;28(3):156–162. Available at: https://www.monalisatouch.ch/wp-content/uploads/V2LR_Gaspar_et_al_AJCS_2011.eng_.pdf
11. Salvatore S, Nappi RE, Zerbinati N, Calligaro A, Ferrero S, Origoni M, Candiani M. A 12-week treatment with fractional CO_2 laser for vulvovaginal atrophy: A pilot study. Climacteric. 2014;17(4):363–369. https://doi.org/10.3109/13697137.2014.899347
12. Salvatore S, Maggiore ULR, Zerbinati N, Calligaro A, Ferrero S, Origoni M, Candiani M. Microablative fractional CO_2 laser improves dyspareunia related to vulvovaginal atrophy: A pilot study. Journal of Endometriosis. 2014;6(3):150–156. https://doi.org/10.5301/je.5000184
13. Mension E, Alonso I, Tortajada M, Matas I, Gómez S, Ribera L, Anglès S, Castelo-Branco C. Vaginal laser therapy for genitourinary syndrome of menopause: A systematic review. Maturitas. 2022;156:37–59. https://doi.org/10.1016/j.maturitas.2021.06.005
14. Zerbinati N, Serati M, Origoni M, Candiani M, Iannitti T, Salvatore S, Marotta F, Calligaro A. Microscopic and ultrastructural modifications of postmenopausal atrophic vaginal mu-cosa after fractional carbon dioxide laser treatment. Lasers in Medical Science. 2015;30:429–436. https://doi.org/10.1007/s10103-014-1677-2
15. Di Donato V, D'Oria O, Scudo M, Prata G, Fischetti M, Lecce F, Schiavi MC, Panici PB. Safety evaluation of fractional CO_2 laser treatment in post-menopausal women with vaginal atrophy: A prospective observational study. Maturitas. 2020;135:34–39. https://doi.org/10.1016/j.maturitas.2020.02.009
16. Perino A, Calligaro A, Forlani F, Trezza V, Falcone V, Cucinella G. Vulvo-vaginal atrophy: A new treatment modality using thermo-ablative fractional CO_2 laser. Maturitas. 2015;80(3):296–301. https://doi.org/10.1016/j.maturitas.2014.12.006
17. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic re-views and meta-analyses: The PRISMA statement. Physical Therapy. 2009;89(9):873–880. https://doi.org/10.1371/journal.pmed.1000097

* The articles used in the meta-analysis are indicated with an asterisk symbol (*).

18. The Joanna Briggs Institute. Critical appraisal tools for use in JBI systematic reviews. Available at: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Systematic_Reviews2017_0.pdf
19. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Systematic reviews of effective-ness. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis. Adelaide: JBI; 2024. https://doi.org/10.46658/JBIMES-24-03
20. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–560. https://doi.org/10.1136/bmj.327.7414.557
21. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. Chichester: John Wiley & Sons; 2021. https://doi.org/10.1002/9780470743386
22. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–634. https://doi.org/10. 1136/bmj.315.7109.629
23. *Eder SE, Heidinger C, Hanzal E. Early effect of fractional CO_2 laser treatment in post-menopausal women with vaginal atrophy. Laser Therapy. 2018;27(1):41–47. https://doi.org/10.5978/islsm.18-OR-04
24. *Pitsouni E, Grigoriadis T, Tsiveleka A, Zacharakis D, Salvatore S, Athanasiou S. Micro-ablative fractional CO_2 laser therapy and the genitourinary syndrome of menopause: An ob-servational study. Maturitas. 2016;94:131–136. https://doi.org/10.1016/j.maturitas.2016.09.012
25. *Ruanphoo P, Bunyavejchevin S. Treatment for vaginal atrophy using microablative frac-tional CO_2 laser: A randomized double-blinded sham-controlled trial. Menopause. 2020;27(8):858–863. https://doi.org/10.1097/GME.0000000000001542
26. *Salvatore S, Nappi RE, Parma M, Chionna R, Lagona F, Zerbinati N, Ferrero S, Origoni M, Candiani M, Maggiore LR. Sexual function after fractional microablative CO_2 laser in women with vulvovaginal atrophy. Climacteric. 2015;18(2):219–225. https://doi.org/10.3109/13697137.2014.975197
27. *Salvatore S, Pitsouni E, Grigoriadis T, Zacharakis D, Pantaleo G, Candiani M, Athanasiou S. CO_2 laser and the genitourinary syndrome of menopause: A randomized sham-controlled trial. Climacteric. 2021;24(2):187–193. https://doi.org/10.1080/13697137.2020.1829584
28. *Siliquini GP, Tuninetti V, Bounous VE, Bert F, Biglia N. Fractional CO_2 laser therapy: A new challenge for vulvovaginal atrophy in postmenopausal women. Climacteric. 2017;20(4):379–384. https://doi.org/10.1080/13697137.2017.1319815
29. *Page AS, Verbakel JY, Verhaeghe J, Latul YP, Housmans S. Laser versus sham for genito-urinary syndrome of menopause: A randomized controlled trial. BJOG. 2022;130:312. https://doi.org/10.1111/1471-0528.17335
30. *Tenerowicz AR, Zimmer-Stelmach A, Zimmer M. CO_2 ablative laser treatment in peri-menopausal patients with vulvovaginal atrophy. Ginekologia Polska. 2022;93(5):374–380. https://doi.org/10.5603/gp.a2021.0140
31. Portman DJ, Gass MLS. Genitourinary syndrome of menopause: New terminology for vul-vovaginal atrophy. Journal of Sexual Medicine. 2014;11(12):2865–2872. https://doi.org/10.1097/gme.0000000000000329
32. U.S. Food and Drug Administration. FDA warns against using lasers for vaginal rejuvena-tion. Available at: https://www.docwirenews.com/post/fda-issues-warning-against-vaginal-rejuvenation-procedures
33. Sarmento ACA, Lírio JF, Medeiros KS. Physical methods for the treatment of genitourinary syndrome of menopause: A systematic review. International Journal of Gynecology & Obstetrics. 2020;153(2):200–219. https://doi.org/10.1002/ijgo.13561
34. Prodromidou A, Zacharakis D, Athanasiou S, Kathopoulis N, Varthaliti A, Douligeris A, Michala L, Athanasiou V, Salvatore S, Grigoriadis T. CO_2 laser versus sham control for the management of genitourinary syndrome of menopause: A systematic review and meta-analysis of randomized controlled trials. Journal of Personalized Medicine. 2023;13(12):1694. https://doi.org/10.3390/jpm13121694
35. Purzand B, Rokhgireh S, Zanjani MS, Eshraghi N, Mohamadianamiri M, Esmailzadeh A, Alkatout I, Gitas G, Allahqoli L. Comparison of soybean and fish oil supplementation on menopausal symptoms in postmenopausal women: A randomized double-blind placebo-controlled trial. Complementary Therapies in Clinical Practice. 2020;41:101239. https://doi.org/10.1016/j.ctcp.2020.101239
36. Royal College of Nursing. Menopause guidance for nurses, midwives and health visitors. 2020. Available at: https://www.rcn.org.uk/library/-/media/Royal-College-Of-Nursing/Documents/Publications/2025/August/012-073.pdf
37. Filippini M, Porcari I, Ruffolo AF, Casiraghi A, Farinelli M, Uccella S, Franchi M, Candi-ani M, Salvatore S. CO_2-laser therapy and genitourinary syndrome of menopause: A sys-tematic review and meta-analysis. Journal of Sexual Medicine. 2022;19(3):452–470. https://doi.org/10.1016/j.jsxm.2021.12.010
38. Salvatore S, Pitsouni E, Del Deo F, Athanasiou S, Grigoriadis T, Candiani M. Sexual func-tion in women suffering from genitourinary syndrome of menopause treated with fraction-ated CO_2 laser. Sexual Medicine Reviews. 2017;5(4):486–492. https://doi.org/10.1016/j.sxmr.2017.07.003

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Review Article DOI: https://doi.org/10.23950/jcmk/18568

Cross-Modal Self-Attention Fusion for Breast Cancer Subtype Classification Using Multi-Omics Data

Kurmash Zhumagozhayev[1], Tomiris Zhaksylyk[1], Beibit Abdikenov[1], Temirlan Karibekov[2], Liliya Skvortsova[3], Adil Faizullin[1]

[1]Science and Innovation Center "Artificial Intelligence", Astana IT University, Astana, Kazakhstan
[2]Science and Innovation Center "MedTech", Astana IT University, Astana, Kazakhstan
[3]Laboratory of Molecular Genetics, Institute of Genetics and Physiology, Committee of Science of the Ministry of Science and Higher Education, Almaty, Kazakhstan

Received: 2025-10-24.
Accepted: 2026-05-16.

J Clin Med Kaz 2026; 23(3): 40-51

Corresponding authors:

Tomiris Zhaksylyk
E-mail: zhaksylyk.tomiris@astanait.edu.kz.
ORCID: 0009-0002-8749-1967;

Beibit Abdikenov.
E-mail: beibit.abdikenov@astanait.edu.kz.
ORCID: 0000-0002-0284-0949.

ABSTRACT

Background: Accurate classification of breast cancer subtypes is essential for personalized therapy and prognosis. Traditional subtype classification basically relies on gene expression profiling, usually overlooking other genomic signals like copy-number alterations (CNA) and mutations. At the same time most of the multi-omics models often rely on early or late fusion strategies, which do not capture complex inter-modality interactions.

Methods: This study proposes a cross-modal transformer-based approach that integrates gene expression, copy number alterations, and mutation data for robust breast cancer subtype classification. Each omics modality is encoded as a separate sequence and projected into a shared embedding space. Gene expression is treated as the primary modality and enriched through cross-modal self-attention mechanisms with CNA and mutation features. The final enriched embeddings are flattened and passed through a residual-connected MLP classifier. We evaluate performance on the METABRIC dataset using ElasticNet-selected top-K features (K = 300, 500, 1000, 1500) and mostly focus on macro F1-score, weighted F1-score, and ROC AUC due to class imbalance.

Results: Integrating copy-number and mutation data with expression features improved subtype classification across most feature set sizes. The tri-omic model (EXP+CNA+MUT) achieved the best performance for smaller feature sets (K = 300–500), whereas for larger feature sets (K = 1000) the highest scores were obtained by the bi-omic model (EXP+CNA) with macro-F1 = 0.859, weighted F1 = 0.868, accuracy = 0.866 and ROC AUC = 0.969. Paired statistical tests across five folds showed that differences between modality configurations did not reach significance at any K (all $p > 0.09$), whereas feature-set size did.

Within the EXP+CNA configuration alone, macro-F1 increased significantly from K = 300 to K = 500 (paired t-test, $p = 0.012$) and from K = 300 to K = 1000 ($p = 0.036$); and in the higher-powered pooled analysis across all three modality configurations (n = 15 paired folds), K = 1000 also outperformed K = 300 ($p = 0.030$).

Conclusion: This pipeline demonstrates an application of cross-modal attention for omics integration in subtype classification task, offering a scalable and biologically grounded alternative to traditional fusion approaches.

Keywords: breast cancer; copy number alteration; cross-modal attention; gene expression; METABRIC; molecular subtype classification; multi-omics integration; mutation; transformer.

Introduction

Cancer remains a leading cause of mortality worldwide and in Kazakhstan, where socioeconomic development, urbanization, aging, and lifestyle transitions have driven a rising non-communicable disease burden. Nationwide data from 2014–2022 show stable overall cancer mortality but a temporary increase in the mortality-to-incidence ratio during the COVID-19 pandemic, as reported in recent national analyses [1, 2]. Breast cancer significantly contributes to this burden, with increasing prevalence (from 30.4 to 50.6 per 10,000 population between 2014 and 2019) and incidence rates (peaking at 7.3 per 10,000 in 2016), alongside high mortality in older age groups and associations with comorbidities like diabetes [3]. Against this backdrop, the present study develops a cross-modal Transformer that directly fuses gene-expression, copy-number and mutation signals for breast-cancer molecular-subtype classification, evaluated on the METABRIC cohort as a methodological testbed for future application to national cohorts.

Effective prognosis and diagnosis of breast cancer are crucial for improving outcomes, as early detection and risk stratification can address aggressive tumors, treatment related complications, and long-term effects [4-7]. Advanced imaging modalities, including mammography, ultrasound, MRI, digital breast tomosynthesis, and breast CT, enhance visualization and early detection, particularly in dense breasts (BI-RADS C/D, linked to larger tumors and poorer outcomes), while supplemental imaging raises overall sensitivity to 90-97% [4, 5].

Bridging imaging-based diagnosis with molecular insights, liquid biopsies provide non-invasive real-time monitoring of tumor evolution through biomarkers like circulating tumor DNA (ctDNA) and cells, enabling detection of resistance mutations and minimal residual disease to guide precision oncology [8]. This integration underscores the need for multi-omics approaches to capture breast cancer's genetic heterogeneity.

Beyond this phenotypic-surveillance layer, effective stratification ultimately requires access to the tumor’s underlying molecular program. Accurate molecular subtyping of breast cancer is essential for guiding personalized therapy and prognosis. Gene expression profiling studies have revealed that breast cancer is not a single disease but a collection of subtypes with distinct molecular signatures and clinical outcomes [9]. Pioneering work by Sorlie et al. identified intrinsic subtypes Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like based on unsupervised clustering of tumor mRNA expression profiles, showing their relevance to treatment outcomes and survival trajectories [10]. These findings laid the foundation for the development of clinical assay such as the PAM50 classifier, a 50-gene expression panel designed to assign breast tumors into the intrinsic subtypes in a supervised manner [11]. In addition, researchers proposed a sixth distinct group, the Claudin-low subtype, characterized by low expression of cell-cell adhesion genes (e.g., claudin 3/4/7, E-cadherin) and high expression of mesenchymal and immune response markers, marking a more stem-like, aggressive phenotype [12].

Despite the clinical utility of expression-based classifiers, they might not fully capture key genomic alterations that drive breast cancer heterogeneity. For instance, CNAs gains and losses of chromosomal segments contribute significantly to oncogenic pathways and are highly prevalent in breast tumors. Somatic mutations in genes like TP53, PIK3CA, and BRCA1/2 further influence tumor behavior and therapeutic sensitivity. The METABRIC dataset comprising matched gene expression, CNA, and clinical profiles from around 2,000 breast tumors has been pivotal in uncovering these relationships [13]. At present, no locally curated breast cancer cohort in Kazakhstan combines matched gene expression, CNA and somatic mutation profiles at the scale required for supervised deep learning. We therefore utilize METABRIC dataset as a methodological basis, with architecture deliberately designed to transfer to matched national cohorts as they mature.

Given this complexity, multi-omics integration has become a key strategy for enhancing molecular classification [14-16]. However, most current multi-omics models employ either early fusion (feature concatenation) or late fusion (decision-level merging), both of which might have certain limitations in capturing inter-modality interactions [17]. Early fusion may dilute important signals across modalities, while late fusion ignores dependencies between, for example, a gene expression level and its mutational or copy number status [18]. Emerging deep learning methods, especially those based on self-attention mechanisms such as Transformers, offer a way to dynamically model cross-modal relationships through learned attention patterns [9, 18, 19].

Dedicated cross-modal attention where gene expression tokens directly attend to copy number and mutation tokens, as opposed to attention applied separately within each modality remains mostly unexplored for breast cancer subtype classification. To address this, we propose a cross-modal Transformer-based approach for breast cancer subtype classification using gene expression (EXP), copy number alterations (CNA), and mutation (MUT) data from METABRIC. Each omics modality is independently represented as a sequence of features and encoded into a shared latent embedding space. Gene expression is designated as the primary modality and is subsequently enriched through cross-modal attention, leveraging CNA and mutation embeddings as contextual sources to incorporate complementary biological signals. After the enriched representation is flattened and passed through a residual-connected multilayer perceptron (MLP) classifier.

Related work

learning has emerged as a powerful tool in bioinformatics, but it faces challenges with high-dimensional omics data and limited samples [20]. High feature counts (tens of thousands of genes) and class imbalance complicate model training and interpretation. A review by Nasser and Yusof highlights that while deep learning has achieved remarkable results in breast cancer imaging diagnostics and some genomic tasks, issues like interpretability and data heterogeneity remain obstacles [14]. Similarly, Abdikenov et al. survey the landscape of machine learning in breast cancer diagnosis and emphasize emerging trends in multimodal data integration combining different data sources (imaging, genomic, clinical) is seen as a key to improving predictive performance [21]. Our work aligns with these trends by focusing on multi-omics fusion via a novel transformer-based approach to better capture the complexity of breast tumors.

Feature Selection and Dimensionality Reduction

Because of the "large p, small n" problem in genomics (more features than samples), feature selection is often a crucial step for subtype classification tasks. One classical approach is to identify a small gene panel that discriminates against subtypes, as done with PAM50 [11]. Beyond expert-curated genes, data-driven methods are widely used. Differential expression analysis can filter genes; for instance, Choi and Chae (moBRCA-net) selected the top 1000 differentially expressed genes and associated features from other omics in their multi-omics model

[22]. Other works have used mutual information or entropy-based filtering. Liu et al. applied a mutual information feature selection to identify genes distinguishing triple-negative breast cancer subgroups [23]. An emerging strategy is to incorporate feature selection into the model training itself. Similarly, Guo et al. introduced a neural-network-based gene selection using knockoff filters to identify predictive genes with statistical control of false discoveries [24]. Considering high number of predictors versus number of observations, the ElasticNet regularization, which linearly combines L1 and L2 penalties, is another choice to enforce sparse gene selection while training a classifier [25]. Evolutionary and heuristic algorithms have also been explored for gene selection: Molaei et al. used particle swarm optimization to pick informative microRNAs before classification, and Andelic and Segota evolved symbolic expressions (combinations of a small number of genes) that achieved high accuracy in subtype classification (reporting ~99% accuracy on a small microarray dataset of 6 subtypes after balancing) [26, 27]. Unsupervised dimensionality reduction can likewise alleviate the curse of dimensionality. Bruno and Calimeri demonstrated that applying dimensionality reduction techniques such as PCA on gene expression data, alongside the integration of clinical features, could improve the visualization and classification of breast tumors [28]. In addition, autoencoder-based compression has been used to denoise and reduce genomic data: Arafa et al. developed a reduced-noise autoencoder that mitigated class imbalance effects and improved cancer genomic classification [29]. These methods show that carefully reducing feature space either through statistical selection or learned lower-dimensional representations can significantly benefit subtype classifiers by focusing on models on the most relevant biological signals.

Single-Omics Deep Learning Models

Early applications of machine learning for subtype classification often used single-omics data, primarily gene expression. Traditional classifiers (SVMs, Random Forests, etc.) achieved moderate success, but deep learning models have started to outperform them by capturing nonlinear patterns. Mostavi et al. proposed CNN-based models operating on gene expression profiles to classify tumor types across cancers; for breast cancer subtypes, such CNNs can automatically learn groups of co-expressed genes relevant to each subtype [30]. Mohamed et al. more recently designed a "bio-inspired" deep CNN for breast cancer detection using gene expression data, demonstrating that network architectures tailored to genomic data (e.g., using layers that mimic gene-gene interaction patterns) can improve accuracy [31]. Beykikhoshk et al. introduced an attention mechanism in a model called DeepTRIAGE to compute personalized biomarker scores from gene expression; they specifically classified Luminal A vs Luminal B subtypes and used attention to highlight genes contributing to the distinction for each patient (improving interpretability). These single-omics deep models typically report high performance on their focused tasks (e.g., DeepTRIAGE achieved over 84% accuracy distinguishing Luminal A vs B), but they inherently ignore other molecular information [9]. Our approach builds on the successes of deep learning in capturing transcriptomic patterns and extends it by integrating additional omics modalities to capture a more holistic view of the tumor.

Multi-Omics Integration Strategies. Integrating multiple data types (gene expression, CNAs, mutations, methylation, etc.) is challenging due to differing data scales and the potential lack of one-to-one feature correspondence. Nonetheless, many works have shown that multi-omics models can outperform single-omic ones for subtyping [16, 22]. Fusion strategies are generally categorized as early fusion (concatenate raw features or learned representations from each modality and then classify), late fusion (build separate models on each modality and then combine their predictions), or hybrid approaches in between. Early fusion is exemplified by Lin et al.'s DeepMO model, which trained parallel neural network subnetworks for mRNA, DNA methylation, and CNA data and then concatenated the learned features for final subtype prediction. DeepMO showed that a simple fully connected integration of multi-omic features already improved accuracy over single-omic models on TCGA data [16]. However, early fusion can struggle when one omics type dominates the signal or when there are many irrelevant features. Late fusion approaches, such as ensemble methods, build an expert model per modality and then aggregate decisions. For instance, Arya and Saha proposed a stacked ensemble where separate deep models for different modalities (expression CNA and clinical variables) were combined to predict breast cancer prognosis while their task was prognosis rather than subtype, the principle is similar – each data source is first mined independently for predictive insight [32]. Late fusion can be robust if one data type fails, but it may miss cross-modality feature interactions. More recent methods therefore explore intermediate or joint fusion methods that preserve modality-specific modeling while enabling inter-modality interaction. Graph-based models interpret multi-omics data as a network: Tanvir et al. introduced MOGAT, which builds a graph where nodes represent samples (with multi-omic feature vectors) and uses graph attention networks to learn sample embeddings that consider multi-omic similarity – yielding improved cancer subtype predictions by leveraging patient–patient relationships across omics [33]. Another graph-based approach, MoGCN by Li et al. compresses each omics layer with an autoencoder and fuses the resulting sample representations into a patient similarity network via Similarity Network Fusion; a graph convolutional network is the applied over this patient graph to classify cancer subtypes [18]. These graph-based models explicitly leverage cross-sample relationships induced by multi-omic similarity and have demonstrated strong performance in pan-cancer subtype classification.

Cascianelli et al. leveraged a pan-cancer multi-omic dataset, first training deep (especially semi-supervised) models across various cancer types including breast cancer and then tuning specifically for breast cancer subtype classification [15]. Their results indicate that both data quantity and heterogeneity, even beyond breast-specific samples, can improve model generalization and subtype discrimination. Attention mechanisms have also been applied at the feature level for multi-omics integration. One notable example is the proposed moBRCA-net framework by Choi and Chae [22]. In moBRCA-net, each omics modality gene expression, DNA methylation, and microRNA were processed independently through a self-attention module that assigns an importance weight to each feature within its respective modality. The resulting weighted representations are then concatenated and used for subtype classification. This design allows the model to highlight which genes, CpG sites, or miRNAs are most relevant, thus addressing the interpretability challenge commonly associated with deep models.

MoBRCA-net achieved an average accuracy of 89,1% and a F1-score of 0.887 on breast cancer subtypes from the TCGA dataset, outperforming ML-based models without attention mechanisms. However, it is important to note that moBRCA-net applies attention separately within each modality and does

not explicitly model cross-modal interactions. In contrast, our proposed method takes a step further by employing cross-modal attention, enabling features in one modality to directly attend to and integrate information from another. In this Transformer mechanism, cross-attention enables elements in one sequence (e.g., decoder tokens) to "attend" to elements in another sequence (e.g., encoder outputs), allowing rich contextualization. Similarly, in our model, a gene expression level can "look at" CNA or mutation features either from the same gene or from others when forming its representation [19]. By doing so, we allow direct cross-modal interactions, enhancing integration and interpretation of multi-omics data. To our knowledge, few works in cancer genomics have used transformer-style cross-modal attention for data fusion.

In summary, the state-of-the-art in breast cancer molecular subtype classification includes a spectrum of approaches: from conventional biomarkers and simple model ensembles to advanced deep learning architectures with attention, graph networks, and multi-omics integration.

Our approach is a cross-modal transformer that enriches gene expression features with CNA and mutation context, and which captures complex inter-modality relationships that early fusion or per-modality attention methods might miss. We hypothesize that this leads to more discriminative and robust representations for subtype classification, especially in scenarios of class imbalance or when the signal in any single modality is weak.

Methods

Data and Preprocessing

We evaluated our approach on the METABRIC dataset, a large breast cancer cohort with clinical, genomic, and transcriptomic data, that was acquired via cBioPortal [13, 34]. METABRIC provides gene expression profiles (originally from Illumina HT-12 v3 microarrays), somatic copy-number alterations (CNA), and somatic mutation data for each tumor, along with detailed clinical annotations.

We used the molecular subtype labels provided by METABRIC, which originally had six subtypes (5 PAM50 intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, Basal-like, Normal-like) and the Claudin-low subtype (the latter was assigned based on gene expression clustering) [12]. Due to the low number of samples of Normal-like and Claudin-low we decided to omit them. This resulted in a 4-class classification task.

The dataset comprises 1523 breast tumor samples with complete data across all three omics modalities. Samples with missing data in any modality were excluded. Gene expression data (20603 genes) were used as log2-transformed and z-score normalized values relative to the expression distribution across all samples. CNA values (22544 genes) were represented as discrete copy-number calls: "- 2" homozygous deletion, "-1" hemizygous deletion, "0" neutral, "1" gain, and "2" high-level amplification, following contamination correction and thresholding.

Mutation data for 173 genes were encoded as multi-hot vectors indicating the presence or absence of specific Gene + Variant Classification combinations, resulting in 885 binary features. In the training set, 13 distinct Variant Classification categories were observed, including Missense_Mutation, Nonsense_Mutation, Frame_Shift_Del, Frame_Shift_Ins, In_Frame_Del, In_Frame_Ins, Nonstop_Mutation, Translation_Start_Site, Splice_Site, Splice_Region, as well as synonymous and non-coding categories such as Silent, Intron, and 5'UTR. We used this encoding to allow the model to capture mutation-specific patterns relevant to subtype classification, while mitigating sparsity from rare individual mutations. For example, if a tumor sample had a missense mutation in the gene PIK3CA, the corresponding binary feature "PIK3CA_Missense_Mutation" was set to 1, while all other PIK3CA-related mutation features remained 0. Similarly, if a frameshift deletion was present in TP53, the feature "TP53_Frame_Shift_Del" was set to 1. If multiple mutations of different types occur in the same gene, multiple Gene + Variant Classification indicators can be simultaneously active for that sample.

We stratified the dataset into training and test subsets using an 80/20 split, preserving the subtype class distribution (which is imbalanced: Luminal A and Luminal B were the most common, while HER2-enriched and Basal-like were least represented) (see Table 1). The protocol governing use of this hold-out test set is described further under Evaluation protocol.

Table 1 Sample distributions by class

Subtype	Count
Luminal A	662
Luminal B	449
HER2-enriched	214
Basal-like	198
Total	**1523**

Feature selection

To reduce dimensionality and focus the model on the most informative genomic features, we applied ElasticNet logistic regression exclusively on the training set to avoid information leakage [25]. ElasticNet, combining L1 and L2 penalties, was fitted separately for gene expression (EXP) and copy-number alteration (CNA) data to predict the four molecular subtypes. This procedure produced two independent ranked lists of features based on the absolute values of their coefficients, yielding 1691 non-zero features for EXP and approximately 4000 for CNA.

Rather than fixing a single number of features, we explored four feature-set sizes (K = 300, 500, 1000, 1500), ranging from more stringent to more inclusive selections (limited by the EXP modality). For each value of K (K = 300, 500, 1000, 1500), we selected the top-K features from the EXP ranking and, in parallel, the top-K features from the CNA ranking, using the same K for both modalities. Thus, each experiment used K expression features and K CNA features, while the mutation (MUT) modality was kept fixed.

This design ensured consistent and controlled dimensionality across EXP and CNA inputs while preserving modality-specific feature selection. Moreover, it avoids any forced feature alignment between modalities, allowing the cross-modal attention mechanism to learn interactions without assuming shared feature indices.

Model training was performed on the training subset using stratified 5-fold cross-validation. It was used for model optimization and early stopping. In each fold, a portion of the training data was held out as a validation subset to monitor convergence and select the best model checkpoint. The overall steps of the proposed work are presented in Figure 1.

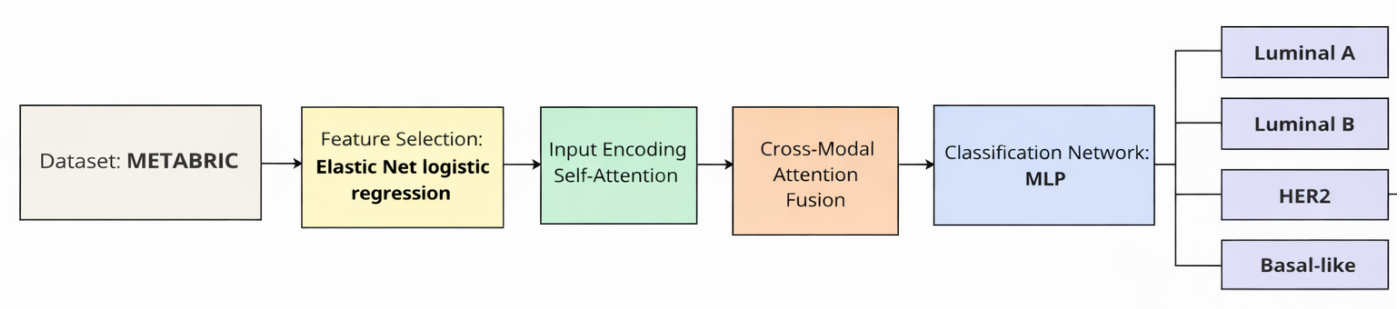


Figure 1 – Overall pipeline of the proposed method

Evaluation protocol

To prevent any information leakage, the external 20 % hold-out test set was fixed once at the start and was never used during feature ranking, pretraining, supervised training, hyperparameter tuning, or model-checkpoint selection. Within the 80 % training set we applied stratified 5-fold cross-validation: in each fold, four fifths of the training set were used for parameter updates and the remaining fifth served as an internal validation subset for early stopping and checkpoint selection. Five independent models (one per fold) were trained end-to-end, and each was then evaluated on the same, disjoint hold-out test set. The numbers reported in Table 4 are the mean (and, for macro-F1, the standard deviation) across these five independent test-set evaluations. No union of predictions, no re-fitting on the full training set, and no post-hoc thresholding on the test set were performed. Because the 5-fold split is defined strictly on the training 80 %, no internal validation fold can coincide with the external test set by construction. The best model checkpoint within each fold was selected based on minimum validation loss; no test-set information ever fed back into checkpoint selection, and hyperparameters were fixed a priori and kept identical across all Top-K experiments.

Our architecture consists of three key stages: (1) modality-specific self-attention encoding, (2) cross-modal transformer fusion that enriches EXP features using CNA and mutation signals, and (3) a residual MLP classifier that predicts the breast cancer subtype. The model supports ablation by selectively excluding modalities from the fusion step.

Input Encoding and Self-Attention

Each omics modality is treated as a sequence of scalar values one per gene or feature. Given a batch of N samples, the model processes the following inputs:

– Gene expression: $EXPraw \in R^{(N \times {}^{G}exp \times 1)}$

– Copy number alteration: $CNAraw \in R^{(N \times {}^{G}cna \times 1)}$

– Mutation features: $MUTraw \in R^{(N \times 885 \times 1)}$

Each tensor is first projected to a shared latent space using a linear layer ($1 \rightarrow 128$), then passed through a modality-specific Transformer encoder with self-attention:

– $EXPembed = SelfAtt(Linear_{1 \rightarrow 128}(EXPraw)) \in R^{(N \times {}^{G}_{exp} \times 128)}$

– $CNAembed = SelfAtt(Linear_{1 \rightarrow 128}(CNAraw)) \in R^{(N \times {}^{G}_{na} \times 128)}$

– $MUTembed = SelfAtt(Linear_{1 \rightarrow 128}(MUTraw)) \in R^{(N \times 885 \times 128)}$

The self-attended embeddings capture within-modality dependencies and are forwarded to the cross-modal attention blocks.

Self-attention block

Each modality uses a single standard Transformer encoder block with pre-norm residual structure. The block has four attention heads sharing a model width of d_model = 128 (per-head dimension 32), followed by a position-wise feed-forward network with hidden size 4•d_model = 512 and a GLU-style gated activation. Dropout of 0.1 is applied to attention weights, to the feed-forward output, and within the gating, together with a small stochastic-depth probability (DropPath, 0.05) on each residual branch. Residual branches are scaled by learnable LayerScale parameters initialised to 10-4. No positional encoding is applied. The feature order is fixed once per modality by the ElasticNet importance ranking (Feature selection) and is identical across all samples, so a gene's identity is carried by its position in the input tensor; adding sinusoidal or learned positional encodings is therefore neither necessary nor desirable and would introduce a spurious ordinal signal. The cross-modal attention block shares the same block structure, replacing self-attention with multi-head cross-attention in which the expression sequence is the query, and CNA or mutation sequences serve as key and value.

Cross-Modal Attention Fusion

To enrich the expression modality with auxiliary signals, we apply cross-attention using multi-head attention blocks. Gene expression embeddings serve as the query, and either CNA or mutation features act as the key and value:

EXP(CNA enriched) = CrossAttn(query = EXPembed, context = CNAembed)

EXP(MUT enriched) = CrossAttn(query = EXPembed, context = MUTembed)

We then compute the enriched expression tensor via residual addition:

EXPenriched_final = EXPembed + α EXP(CNA enriched) + β EXP(MUT enriched)

Where α and β are learnable scalars obtained via sigmoid(alpha) and sigmoid(beta) respectively. This mechanism allows the model to modulate the contribution of each auxiliary modality and refine the expression embeddings using cross-modal cues without overwriting original signals.

These gates are not driven to zero by the Stage 1 reconstruction term, for two complementary reasons. First, the reconstruction objective is optimized jointly with the supervised classification loss; trivially setting α = β = 0 would yield zero reconstruction but leave the classifier no better than an EXP-only baseline, and the joint objective therefore produces non-zero α and β whenever CNA or MUT information improves discrimination. Second, per-sample modality dropout applied during training randomly zeroes the CNA or MUT context with positive probability, so a solution that relies solely on EXP cannot attain a low joint loss across the training distribution. The

reconstruction term therefore acts as a stabilizing anchor against major drift of the enriched representation, not as an instruction to ignore auxiliary modalities.

Classifier Network

For classification, the enriched expression tensor of shape (N, G, 128) is flattened to a fixed-size vector of length G•128 (e.g., 128 000 features for K = 1000) and passed through a compact residual MLP. The MLP has a single hidden width H = 128, uses batch normalization after each linear layer, ReLU activations, and a skip connection that adds the output of the first block to the output of a bottleneck-and-expand sub-block:

– Linear: G 128 → 128 → BatchNorm → ReLU → Dropout (0.2) – produces the residual branch

– Linear: 128 → 64 → BatchNorm → ReLU → Dropout (0.1)

– Linear: 64 → 128→ Add residual branch → ReLU

– Output layer: 128 → 4 (corresponding to four molecular subtypes)

Ablation Variants

To investigate the contribution of each modality, we performed ablation studies using the following variants:

1. EXP – self-attention, no fusion.
2. EXP enriched by CNA only.
3. EXP enriched by both CNA and MUT.

The diagrams in Figures 2-4 illustrate the core model variants in our study.

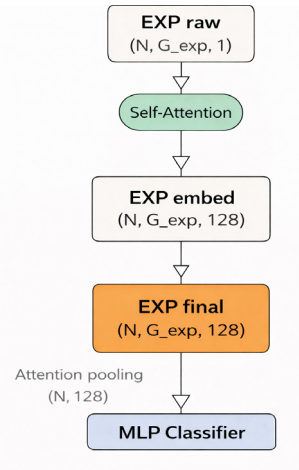


Figure 2 – EXP pipeline

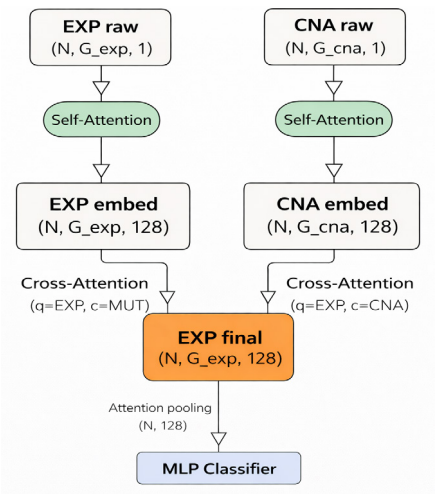


Figure 3 – Cross-modal attention from CNA

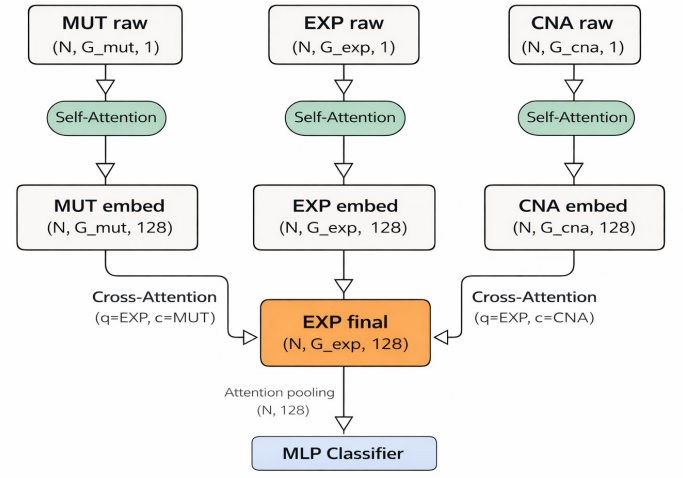


Figure 4 – Cross-modal attention from CNA and MUT

Training Procedure

Our training pipeline consisted of three sequential steps designed to learn within- and cross-modality representations before supervised classification.

Stage 1: Self and Cross-Modal Attention

Before classification, each input modality (EXP, CNA, MUT) was independently encoded using a self-attention Transformer. The resulting self-attended embeddings were then used to enrich the expression representation through cross-modal attention: the expression embedding served as the query, while CNA and mutation embeddings acted as context.

Let E denote the self-attended expression embedding, and AC, AM be the outputs of the CNA→EXP and MUT→EXP cross-attention blocks, respectively. The enriched expression tensor was computed as:

$$E' = E + \alpha \bullet AC + \beta \bullet AM$$

where α and β are learnable gating scalars obtained via sigmoid activation.

To guide this fusion, the model minimized a reconstruction loss between E′ and E:

$$L_{recon} = \| E' - E \|^2$$

Stage 1 uses a masked-feature reconstruction objective combined with an auxiliary classification head, the two losses balanced by a schedule that gradually shifts emphasis from reconstruction toward classification. This forces the encoder to recover masked gene positions from within-modality context while producing embeddings that are already discriminative for the four subtypes. Full numerical hyperparameters are given in Supplementary Table S1.

Stage 2: Cross-modal enrichment training

The cross-modal block is trained on the outputs of Stage 1 under the same masked-reconstruction plus auxiliary-classification recipe, with λ_rec fixed at 0.10 and the classification weight warmed up over the first eight epochs. Two scheduled regularisers are central to the behavior of the learned gates σ(α), σ(β): per-sample modality dropout zeroes the CNA or MUT context with positive probability during training, so a solution that relies solely on EXP cannot attain a low joint loss and the gates cannot collapse to zero; and a scheduled mask ratio on the expression query (5% → 12%) keeps the reconstruction anchor active throughout training. Full numerical hyperparameters are given in Supplementary Table S1.

Stage 3: Supervised Classification

Using the enriched embeddings E′, the MLP classifier was trained to predict molecular subtypes.

The flattened enriched tensor is standardized per fold using mean and standard deviation computed on the training indices of that fold only, so that no validation or test statistic ever leaks into the fit. The classifier input is regularized with input dropout and MixUp, and optimization uses AdamW with a ReduceLROnPlateau schedule and early stopping on validation loss. Class imbalance is addressed with a focal loss weighted by inverse class frequency:

$$L_{focal} = -(1/N) \Sigma \alpha y_i (1 - py_i)^{\gamma} \log(py_i),$$

where py_i is the predicted softmax probability for the true class y_i, αy_i is the inverse-frequency class weight, and γ is a focusing parameter whose value is given with all other Stage 3 hyperparameters in Supplementary Table S1. The combination of focal loss, class weighting, and MixUp penalises confident misclassifications while preserving gradient signal on minority subtypes.

All experiments were implemented in PyTorch using stratified 5-fold cross-validation on the training subset. Within each fold, 20% of the training data served as a validation subset for early stopping and model selection. The full procedure was repeated across different expression-feature sizes (K = 300, 500, 1000, 1500) and modality configurations (EXP-only, EXP + CNA, EXP + CNA + MUT).

The best model checkpoint was selected based on minimum validation loss and then evaluated on the fixed external hold-out test set.

Evaluation Metrics

We evaluated model performance using metrics that address class imbalance, which is critical in subtype classification. The macro-F1 score served as the primary criterion, emphasizing balanced performance across all subtypes, including minority classes. To complement this, weighted F1 reflected the overall performance adjusted for class frequencies, while accuracy was reported only as a secondary reference due to its bias toward dominant subtypes. We also computed ROC AUC as a threshold-independent measure of separability, and tracked precision, recall, and confusion matrices to analyze subtype-specific behavior. Macro-F1 and ROC AUC are highlighted as key indicators.

Statistical analysis

To quantify the uncertainty of the reported metrics and to evaluate whether observed differences between configurations are meaningful, we applied paired two-sided statistical tests across the five fold-specific models. All configurations — different modality setups at a given K, or different K values within a modality — share the same training and test samples fold-for-fold, so paired tests on the five per-fold macro-F1 values are the appropriate design. Two complementary tests were used for each comparison: the paired Student's t-test, and the non-parametric Wilcoxon signed-rank test (with n = 5 paired observations the Wilcoxon test has a minimum attainable two-sided p-value of 0.0625, which we note explicitly at borderline results). For the cross-K analysis we additionally ran a Friedman omnibus test across the four K-levels within each modality, and a pooled analysis across modalities (n = 15 paired observations per comparison, obtained by stacking the five folds of the three modality configurations). The Holm–Bonferroni correction was applied to the family of three pooled comparisons of K = 1000 against the other K-levels. Raw two-sided p-values are reported, with Holm-adjusted p-values where applicable; $p < 0.05$ is considered statistically significant and $0.05 \leq p < 0.10$ as suggestive. Effect sizes are reported as mean paired differences Δ in macro-F1 together with per-fold standard deviations. All tests were implemented with SciPy in Python 3.

Results

Tables 2 and 3 summarize the top-20 expression (EXP) and top-20 copy-number (CNA) features ranked by the absolute

Table 2 Top-20 ElasticNet-ranked features for gene expression (EXP) modality. Features are ordered by the absolute value of their coefficients

Rank	Gene	Coefficient
1	KRT17	0,04600026
2	KRT14	0,039189994
3	SLC39A6	0,038983226
4	COL17A1	0,03479651
5	SFRP1	0,03312954
6	KRT6B	0,03296845
7	FGFR4	0,03000498
8	ELP2	0,02835113
9	TYMSOS	0,025144786
10	CLDN11	0,023658585
11	KRT5	0,022679463
12	PBK	0,02123858
13	LDLRAD4	0,020831855
14	ESR1	0,020572215
15	NKX2-1	0,01995619
16	CALML3	0,019574217
17	FADS2	0,019254878
18	TUBB1	0,019099653
19	AXIN2	0,019011276
20	ERBB2	0,018967364

Table 3 Top-20 ElasticNet-ranked features for copy number alterations (CNA) modality. Features are ordered by the absolute value of their coefficients

Rank	Gene	Coefficient
1	PPP1R1B	0,012399
2	TCAP	0,012062
3	PNMT	0,012062
4	GRB7	0,011504
5	STARD3	0,011289
6	PGAP3	0,011123
7	ERBB2	0,011042
8	MIEN1	0,011042
9	MIR4728	0,011042
10	CTNNA2	0,010845
11	NEUROD2	0,00982
12	SH3RF3	0,009491
13	SHANK2-AS1	0,009424
14	MED1	0,009093
15	CDK12	0,008316
16	DEFB108B	0,008201
17	FBXL20	0,007518
18	ZNF92	0,007361
19	MUC16	0,007319
20	STRN3	0,007299

Table 4 Classification performance on METABRIC breast-cancer subtypes for different feature-set sizes and modality combinations. Bold values indicate the best results within each Top-K group.

Features (Top-K)	Modality	Acc	Pre	Rec	F1 (weighted)	Macro Pre	Macro Rec	Macro F1	ROC AUC
300	EXP	0.8472	0.8582	0.8472	0.8501	0.8336	0.8534	0.8402	0.9663
300	EXP+CNA	0.8400	0.8501	0.8400	0.8428	0.8233	0.8439	0.8306	0.9642
300	EXP+CNA+MUT	0.8623	0.8705	0.8623	0.8645	0.8488	0.8669	0.8551	0.9684
500	EXP	0.8518	0.8596	0.8518	0.8538	0.8360	0.8595	0.8451	0.9633
500	EXP+CNA	0.8518	0.8599	0.8518	0.8537	0.8370	0.8622	0.8469	0.9664
500	EXP+CNA+MUT	0.8557	0.8637	0.8557	0.8578	0.8461	0.8628	0.8519	0.9674
1000	EXP	0.8682	0.8727	0.8682	0.8694	0.8535	0.8654	0.8580	0.9680
1000	EXP+CNA	0.8656	0.8727	0.8656	0.8677	0.8548	0.8671	0.8589	0.9691
1000	EXP+CNA+MUT	0.8577	0.8619	0.8577	0.8590	0.8477	0.8551	0.8502	0.9669
1500	EXP	0.8511	0.8550	0.8511	0.8524	0.8443	0.8485	0.8453	0.9649
1500	EXP+CNA	0.8584	0.8631	0.8584	0.8598	0.8516	0.8610	0.8549	0.9682
1500	EXP+CNA+MUT	0.8544	0.8586	0.8544	0.8558	0.8495	0.8533	0.8503	0.9665

value of their ElasticNet coefficients, together with their ranks and coefficient values. These genes correspond to the highest-ranked entries in the modality-specific ElasticNet rankings used to construct all Top-K feature sets.

Further, we examined how the number of selected features (K) and the integration of omics modalities affected classification performance. For each configuration, the selected model checkpoint was applied to the independent hold-out test set (20% of the data), and all metrics reported in Table 4 were computed on this external test set. Experiments are grouped by feature-set size (Top-K) and evaluated under three modality setups:

- Expression (EXP) only;
- Expression + Copy Number Alterations (CNA);
- Expression + CNA + Mutation (MUT).

The experimental results indicate that both feature-set size (K) and the integration of multiple omics modalities significantly influence classification performance. As shown in Table 4, performance improved as the number of selected genes increased up to K = 1000, with macro-F1 scores rising from 0.84 to 0.86. Beyond this point (K = 1500), gains plateaued or slightly declined, with metrics stabilizing at around 0.85. Multi-omics integration consistently outperformed the single-modality (EXP-only) approach. Specifically, the tri-modal model (EXP+CNA+MUT) achieved the highest macro-F1 for smaller feature sets (K = 300–500). However, for larger feature sets ($K \geq 1000$), the bi-modal (EXP+CNA) configuration demonstrated the best overall performance. Across all configurations, ROC AUC remained high (~0.96–0.97), indicating robust class separability.

We tested whether the numerical differences observed in Table 4 are statistically meaningful. Across the three modality configurations at any K, paired differences in per-fold macro-F1 were consistent in direction but did not reach the $\alpha = 0.05$ threshold (smallest $p = 0.091$ for EXP+CNA vs. EXP+CNA+MUT at K = 1000).

In contrast, the effect of feature-set size was statistically demonstrable. Within EXP+CNA, macro-F1 increased significantly from K = 300 to K = 500 (paired t-test $p = 0.012$) and from K = 300 to K = 1000 ($p = 0.036$); the corresponding Wilcoxon signed-rank p-values (0.063 in both cases) hit the minimum attainable two-sided value for $n = 5$ paired observations and should therefore be read as suggestive rather than as failing to confirm the t-test result.

The Friedman omnibus test over the four K-levels within EXP+CNA returned $\chi^2 = 7.08$, $p = 0.069$, and the EXP+CNA+MUT configuration was flat across K ($\chi^2 = 0.12$, $p = 0.99$). Going from K = 500 to K = 1000 produced a further numerical gain that did not reach significance in any configuration (pooled $\Delta = +0.008$, $p = 0.20$), and increasing K from 1000 to 1500 yielded no significant gain either, with all three configurations showing a flat or slightly negative trend (pooled $\Delta = -0.006$, $p = 0.25$); this absence of further gain is consistent with the plateau evident in Figure 5 and supports K = 1000 as the operating point at which the significant gains end and the plateau begins.

Pooling across modalities ($n = 15$ paired observations per comparison) confirmed the same picture at the cohort level: K = 1000 significantly outperformed K = 300 (paired t-test $p = 0.030$; Wilcoxon $p = 0.031$), with the Friedman test across the four K-levels giving $\chi^2 = 6.84$, $p = 0.077$. After Holm–Bonferroni correction for the three headline pooled comparisons of K = 1000 against K = 300, 500 and 1500, the adjusted p-values are 0.089, 0.41 and 0.25. The most salient pairwise comparisons are reported in Table 5. Figure 5 shows macro-F1 as a function of K for the three modality configurations. The tri-modal configuration (EXP + CNA + MUT) saturates at K = 300 and remains statistically flat across K, while the bi-modal EXP + CNA configuration benefits from additional features and peaks at K = 1000.

A detailed analysis of the subtype-level performance for the top-performing model (K=1000, EXP+CNA) is presented in Figure 6 and Table 6. The use of Macro Average metrics ensures that the model's performance is evaluated equally across all subtypes, regardless of their prevalence in the dataset. The model demonstrated the highest discriminative power for Basal-like tumors, achieving an F1-score of 0.93. Similarly, Luminal A samples were classified with high precision (0.93) and recall (0.87). In contrast, the HER2-enriched subtype proved to be the most challenging, with a precision of 0.70. As shown in the confusion matrix, a significant portion of HER2-enriched misclassifications involved Luminal B and Luminal A, with 22 and 10 cases respectively. Conversely, Luminal B was most frequently confused with Luminal A (36 cases) and HER2-enriched (32 cases), highlighting the molecular overlap between these categories.

Table 5 Key paired statistical tests on per-fold macro-F1 (two-sided). Δ is the mean paired difference. n = 5 folds per comparison within a single modality; n = 15 for the pooled rows

Family	Comparison	Δ	p (t-test)	p (Wilcoxon)
Modality (K=300)	EXP+CNA+MUT vs EXP+CNA	+0.0246	0.096	0.125
Modality (K=1000)	EXP+CNA+MUT vs EXP+CNA	−0.0087	0.091	0.313
Cross-K (EXP+CNA)	K=500 vs K=300	+0.0163	0.012	0.063
Cross-K (EXP+CNA)	K=1000 vs K=300	+0.0284	0.036	0.063
Cross-K (EXP+CNA+MUT)	K=1000 vs K=300	−0.0050	0.473	0.813
Pooled (n=15)	K=1000 vs K=300	+0.0137	0.030	0.031
Pooled (n=15)	K=1000 vs K=500	+0.0077	0.204	0.244

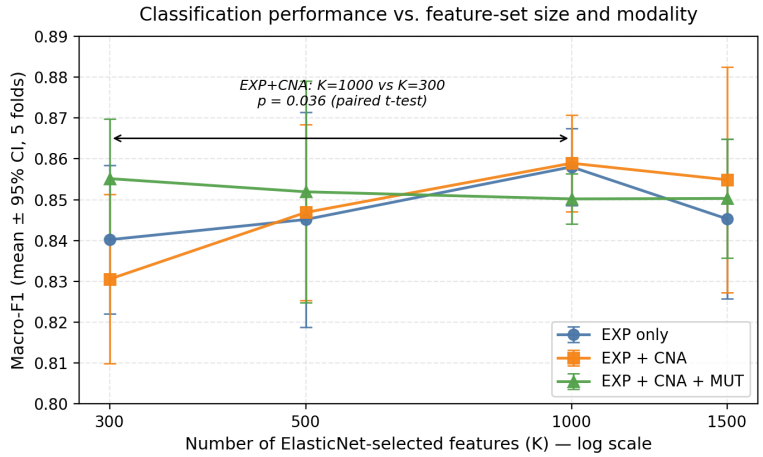


Figure 5 – Macro-F1 vs the number of features (K) for the three modality configurations

Each point is the mean across five fold-specific models; error bars show the 95% confidence interval of the mean. The bracket highlights the only within-modality comparison that reached significance at α = 0.05 (EXP+CNA, K = 300 vs K = 1000; paired t-test, p = 0.036).

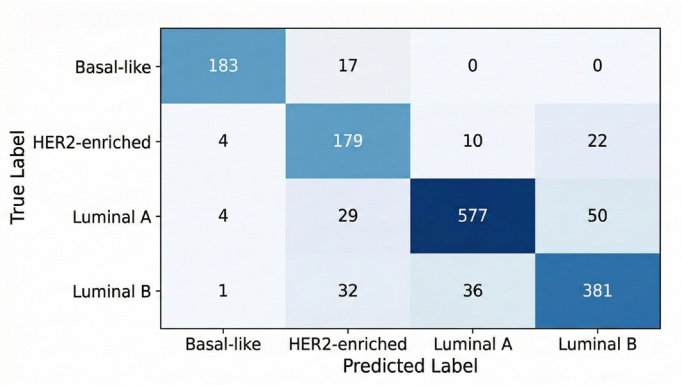


Figure 6 – Cross-modal attention from CNA

Table 6 Cumulative subtype-specific classification metrics (N=1525) Combined results for the K=1000 EXP_CNA model across five checkpoints (5 x 305 test predictions)

Subtype	Pre	Rec	F1	Support
Basal-like	0.9536	0.9150	0.9338	200
HER2-enriched	0.6978	0.8326	0.7587	215
Luminal A	0.9262	0.8742	0.8994	660
Luminal B	0.8417	0.8467	0.8438	450
Macro-Average	**0.8548**	**0.8671**	**0.8589**	**1525**

Discussion

ElasticNet-selected features demonstrate strong biological coherence with established molecular programs. The prioritization of KRT17, KRT14, and KRT5 underscores the unique epithelial differentiation of the Basal-like subtype, reflecting aggressive phenotypes [12]. Similarly, the ranking of ERBB2, GRB7, and STARD3 captures the focal 17q12 amplification characteristic of HER2-enriched tumors [12, 35]. The inclusion of ESR1 and FOXA1 further validates the model's ability to identify the regulatory backbone of luminal differentiation [35]. This alignment confirms that the cross-modal transformer is not merely identifying statistical noise but is operating on a biologically grounded hierarchy of breast cancer drivers.

Our findings highlight how feature-set size and multi-omics integration jointly influence the classification of molecular subtypes. Performance improved as the number of selected genes increased to approximately K = 1000, beyond which gains plateaued or slightly declined. This suggests that the top 1000 genes capture most of the biologically informative variance, while larger sets introduce redundancy or low-rank noise that reduces model efficiency.

The statistical analysis refines this qualitative picture. Pairwise differences between modality configurations at any K did not reach significance on five folds, so claims of strict superiority of one modality configuration over another should be interpreted as directional rather than established at this cohort size. In contrast, the effect of feature-set size is statistically supported: within EXP+CNA, K = 500 and K = 1000 both outperform K = 300 significantly (p = 0.012 and

0.036) and pooled across modalities K = 1000 outperform K = 300 (p = 0.030). The biologically meaningful lever at this scale is therefore the amount of information available to the cross-modal module, not which auxiliary modality is added.

The behavior of auxiliary modalities across different values of K provides deeper insight into the genomic architecture of breast cancer. At small feature sets (K = 300-500), mutation data provided the most significant performance boost, acting as a critical genetic anchor. When transcriptional coverage is limited, discrete somatic events in driver genes such as the high prevalence of TP53 mutations in Basal-like tumors or PIK3CA mutations in Luminal subtypes provide stable diagnostic cues that compensate for a sparse transcriptomic landscape.

Conversely, the dominance of the bi-modal EXP+CNA setup at K≥1000 highlights breast cancer as a primarily copy-number driven disease. In larger feature sets, the model effectively captures the downstream transcriptional effects of large-scale genomic events, such as the 17q12 (ERBB2) amplification. Because copy-number alterations affect entire gene clusters rather than single points, they provide a more robust and stable reinforcement of subtype identity compared to sparse mutational data, which may include non-functional passenger variants that introduce variability at higher dimensions. These results support a biologically coherent interpretation where gene expression captures the primary phenotypic state, while CNA provides a stable genomic backbone that reinforces subtype identity.

The granular analysis of the confusion matrix reveals specific areas where molecular biology challenges discrete classification. The lower precision for the HER2-enriched subtype (~0.70) highlights a significant genetic gray zone involving Luminal B tumors. Clinically, these triple-positive cases often harbor the HER2 amplification but are simultaneously driven by a dominant Estrogen Receptor program. The fact that many HER2-enriched samples were predicted as Luminal B suggests that the luminal transcriptional program can occasionally mask the HER2-driven signal [12]. Furthermore, the confusion between Luminal A and Luminal B reflects a biological spectrum rather than a binary divide. The transition between these subtypes is largely defined by a gradient of proliferation markers, and our results indicate that while cross-modal attention improves separation, tumors with intermediate proliferation remain difficult to categorize using standard thresholds.

Overall, our cross-modal transformer model achieved robust and competitive performance on the METABRIC dataset for PAM50-like breast cancer subtype classification, surpassing several previously reported multi-omics approaches. The comparative results are given in Table 7.

Limitations

Despite the promising performance of the cross-modal Transformer, several limitations constrain the interpretation of this work. Primarily, our evaluation was restricted to the METABRIC cohort; while this dataset provides a robust benchmark, external validation on independent cohorts such as TCGA-BRCA is essential to confirm the framework's generalizability across diverse populations. Furthermore, while the study is adequately powered (K=1000 vs. K=300) to detect significant architectural improvements, it remains underpowered for resolving sub-percent differences between specific modality configurations. Future iterations should utilize repeated cross-validation or bootstrap resampling to clarify these borderline comparisons. Additionally, the exclusion of Normal-like and Claudin-low subtypes due to class imbalance—reducing the task to a four-class problem—limits the current clinical scope. Technically, the reliance on Illumina HT-12 v3 microarrays necessitates a domain-adaptation step before the model can be applied to modern RNA-seq data. Finally, while our modality-agnostic design allows for the future integration of methylation and proteomic layers, the current lack of a formalized per-sample biomarker scoring system limits the immediate clinical interpretability of the produced attention maps.

Conclusion

This study demonstrates that integrating multi-omics data through a cross-modal transformer framework enables an accurate and biologically interpretable classification of breast cancer subtypes. By employing ElasticNet-based gene selection, we identified expression features aligned with canonical Basal, Luminal, and HER2 molecular programs, ensuring that the model's inputs are grounded in established oncological drivers. Our analysis revealed that a feature set of approximately 1000 genes maximizes predictive performance, beyond which the inclusion of lower-ranked features introduces redundancy that plateaus model accuracy.

A key finding of this work is the complementary role of different omics modalities. We propose that mutation data serves as a critical "genetic anchor" for classification when transcriptomic signals are limited, while copy-number alterations provide a stable genomic backbone that reinforces subtype identity in higher-dimensional spaces. The cross-modal attention mechanism effectively mimics the biological flow of information, weighting genomic alterations in the context of their transcriptional consequences.

The model's performance, characterized by a macro F1-score of 0.8589, proves robust across imbalanced classes. However, the persistent "gray zones" observed in the

Table 7 Comparison of Selected Methods for Breast Cancer Subtype Classification. Bold values indicate our results (Top-1000 EXP+CNA model)

Method	Modalities	Dataset/classes	Macro F1	Weighted F1/ Acc.
DeepMO [16]	EXP+CNA+Meth	TCGA (5 cls)	N/A	78.2% acc.
Islam et al. [17]	EXP+CNA	M-BRIC (5 cls)	N/A	79.2% acc.
moBRCA-net [22]	EXP+Meth+miRNA	TCGA (5 cls)	N/A	0.887 / 89.1%
MOGAT [33]	Multi-omics (8)	TCGA, M-BRIC (5 cls)	0.804	N/A
MoGCN [18]	EXP+CNV+RPPA+Clin	TCGA (4 cls)	~ 0.90	~0.90 / ~89%
Proposed	EXP+CNA	M-BRIC (4 cls)	0.8589	0.8677 / 0.8656

N/A = not reported.

misclassification of HER2-enriched and Luminal B tumors underscore the biological continuum of breast cancer, where luminal-driven transcriptional programs can partially mask HER2-driven signals. Overall, this pipeline offers a scalable and biologically grounded alternative to traditional fusion approaches, providing a foundation for more precise, data-driven diagnostics in precision oncology.

Supplementary materials

The Supplementary information includes tables:

- Supplementary Table 1. _____;
- Supplementary Table 2. _____.

This supplemental materials have been provided by the authors to give readers additional information about their work.

The file can be accessed using: https://www.editorialpark.com/download/article-supp/_____/Supplementary-data.docx.

Author Contributions: Conceptualization, K. Zh., T. Zh.; methodology / planning and organization, K. Zh.; validation, K. Zh., T. Zh., A. F.; formal analysis, T. Zh., A. F.; investigation, K. Zh., T. Zh.; clinical interpretation of the results, medical relevance of the study design, and validation the findings from a translational oncology perspective, T. K.; guidance on biological interpretation of omics features, validation of the relevance of selected molecular markers to breast cancer subtypes, L. S.; resources, B. A.; software, K. Zh.; data curation, K. Zh.; writing – original draft preparation, K. Zh.; writing – review and editing, K. Zh., T. Zh.; visualization, K. Zh.; supervision, B. A.; project administration, A. F., B. A.; funding acquisition, A. F., B. A. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number BR24993145.

Data availability statement: The corresponding author can provide the data supporting the study's conclusions upon request. Due to ethical and privacy constraints, the data are not publicly accessible.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Akhmedullin R, Aimyshev T, Zhakhina G, Yerdessov S, Beyembetova A, Ablayeva A, Biniyazova A, Seyil T, Abdukhakimova D, Segizbayeva A, Semenova Y, Gaipov A. In-depth analysis and trends of cancer mortality in Kazakhstan: a joinpoint analysis of nationwide healthcare data 2014–2022. BMC Cancer. 2024;24:1340. https://doi.org/10.1186/s12885-024-13128-2
2. Beyembetova A, Ablayeva A, Akhmedullin R, Abdukhakimova D, Biniyazova A, Gaipov A. National Electronic Oncology Registry in Kazakhstan: Patient's Journey. Epidemiol Health Data Insights. 2025;1(1):ehdi004. https://doi.org/10.63946/ehdi/16385
3. Midlenko A, Mussina K, Zhakhina G, Sakko Y, Rashidova G, Saktashev B, Adilbay D, Shatkovskaya O, Gaipov A. Prevalence, incidence, and mortality rates of breast cancer in Kazakhstan: data from the Unified National Electronic Health System, 2014–2019. Front Public Health. 2023;11:1132742. https://doi.org/10.3389/fpubh.2023.1132742
4. Chuvakova E, Zaripova L, Segizbayeva A, Baigenzhin A, Yegembay A, Idrissova D. Visualization of Breast Cancer and Safety: Review. J Clin Med Kaz. 2025;22(2):4–11. https://doi.org/10.23950/jcmk/16273
5. Iztleuov Y, Mutigulina G, Almagambetova A, Iztleuova G. Prognostic Role of Breast Architecture in Imaging, Histopathology, and Breast Cancer Outcome. J Clin Med Kaz. 2025; 22(5):73–79. https://doi.org/10.23950/jcmk/16879
6. Tombak Y, Umay EK, Unkazan FN, Karaahmet OZ, Sezer MK, Akyuz EU, Gurcay E. The Effect of Breast Cancer History on Bone Mineral Density in the Treatment of Postmenopausal Osteoporosis: One-Year Follow-Up Results. J Clin Med Kaz. 2024;21(6):85–90. https://doi.org/10.23950/jcmk/15703
7. Tlegenova Z, Balmagambetova S, Zholdin B, Kurmanalina G, Talipova I, Koyshybaev A, Nurmanova D, Sultanbekova G, Baspayeva M, Madinova S, Kubenova K, Urazova A. Stratifying breast cancer patients by baseline risk of cardiotoxic complications linked to chemotherapy. J Clin Med Kaz. 2023;20(3):75-81. https://doi.org/10.23950/jcmk/13325
8. Oladosu TA, Okafor CP, Nwosu PC, Ibukunoluwa AE, Monica UI, Aderanti TA. The Role of Liquid Biopsies in Tracking Tumor Evolution and Overcoming Therapeutic Resistance in Cancer. Oncol Nucl Med Transplantol. 2025;1(1):onmt006. https://doi.org/10.63946/onmt/17244
9. Beykikhoshk A, Quinn TP, Lee SC, Tran T, Venkatesh S. DeepTRIAGE: interpretable and individualised biomarker scores using attention mechanism for the classification of breast cancer sub-types. BMC Med Genomics. 2020;13(Suppl 3):20. https://doi.org/10.1186/s12920-020-0658-5
10. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Borresen-Dale A-L, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci. 2003;100(14):8418-8423. https://doi.org/10.1073/pnas.0932692100
11. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo JP, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160-1167. https://doi.org/10.1200/JCO.2008.18.1370
12. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68. https://doi.org/10.1186/bcr2635
13. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and

transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-352. https://doi.org/10.1038/nature10983
14. Nasser M, Yusof UK. Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction. Diagnostics. 2023;13(1):161. https://doi.org/10.3390/diagnostics13010161
15. Cristovao F, Cascianelli S, Canakoglu A, Carman M, Nanni L, Pinoli P. Investigating deep learning-based breast cancer subtyping using pan-cancer and multi-omic data. IEEE/ACM Trans Comput Biol Bioinform. 2022;19:121-134. https://doi.org/10.1109/TCBB.2020.3042309
16. Lin Y, Zhang W, Cao H, Li G, Du W. Classifying Breast Cancer Subtypes Using Deep Neural Networks Based on Multi-Omics Data. Genes. 2020;11(8):888. https://doi.org/10.3390/genes11080888
17. Islam MM, Huang S, Ajwad R, Chi C, Wang Y, Hu P. An integrative deep learning framework for classifying molecular subtypes of breast cancer. Comput Struct Biotechnol J. 2020;18:2185-2199. https://doi.org/10.1016/j.csbj.2020.08.005
18. Li X, Ma J, Leng L, Han M, Li M, He F, Zhu Y. MoGCN: A Multi-Omics Integration Method Based on Graph Convolutional Network for Cancer Subtype Analysis. Front Genet. 2022;13:806842. https://doi.org/10.3389/fgene.2022.806842
19. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is All You Need. Adv Neural Inf Process Syst. 2017;30. Available at: https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
20. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM, Xie W, Rosen GL, Lengerich BJ, Israeli J, Lanchantin J, Woloszynek S, Carpenter AE, Shrikumar A, Xu J, Cofer EM, Lavender CA, Turaga SC, Alexandari AM, Lu Z, Harris DJ, DeCaprio D, Qi Y, Kundaje A, Peng Y, Wiley LK, Segler MHS, Boca SM, Swamidass SJ, Huang A, Gitter A, Greene CS. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15(141):20170387. https://doi.org/10.1098/rsif.2017.0387
21. Abdikenov B, Zhaksylyk T, Shortanbaiuly O, Orazayev Y, Makhanov N, Karibekov T, Suvorov V, Imasheva A, Zhumagozhayev K, Seitova A. Future of Breast Cancer Diagnosis: A Review of DL and ML Applications and Emerging Trends for Multimodal Data. IEEE Access. 2025;13:136101–136143. https://doi.org/10.1109/ACCESS.2025.3585377
22. Choi JM, Chae H. moBRCA-net: a breast cancer subtype classification framework based on multi-omics attention neural networks. BMC Bioinformatics. 2023;24(1):169. https://doi.org/10.1186/s12859-023-05273-5
23. Liu J, Su R, Zhang J, Wei L. Classification and gene selection of triple-negative cancer subtype using ensemble learning and mutual information-based selection. Brief Bioinform. 2021;22(5):1-12. https://doi.org/10.1093/bib/bbaa395
24. Guo J, Jin M, Chen Y, Liu J. An embedded gene selection method using knockoffs optimizing neural network. BMC Bioinformatics. 2020;21:414. https://doi.org/10.1186/s12859-020-03717-w
25. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. 2005;67(2):301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x
26. Molaei S, Cirillo S, Solimando G. Cancer Detection Using a New Hybrid Method Based on Pattern Recognition in MicroRNAs Combining Particle Swarm Optimization Algorithm and Artificial Neural Network. Big Data Cogn Comput. 2024;8(3):33. https://doi.org/10.3390/bdcc8030033
27. Anđelić N, Šegota SB. Development of symbolic expressions ensemble for breast cancer type classification using genetic programming symbolic classifier and decision tree classifier. Cancers. 2023;15(1):3411. https://doi.org/10.3390/cancers15133411
28. Bruno P, Calimeri F, Kitanidis AS, Momi E. Data reduction and data visualization for automatic diagnosis using gene expression and clinical data. Artif Intell Med. 2020;107:101884. https://doi.org/10.1016/j.artmed.2020.101884
29. Arafa A, El-Fishawy N, Badawy M, Radad M. RN-Autoencoder: Reduced Noise Autoencoder for classifying imbalanced cancer genomic data. J Biol Eng. 2023;17(1):7. https://doi.org/10.1186/s13036-022-00319-3
30. Mostavi M, Chiu Y-C, Huang Y, Chen Y. Convolutional neural network models for cancer type prediction based on gene expression. BMC Med Genomics. 2020;13(Suppl 5):44. https://doi.org/10.1186/s12920-020-0677-2
31. Mohamed T, Ezugwu A, Fonou-Dombeu JV, Ikotun AM, Mohammed M. A bio-inspired convolution neural network architecture for automatic breast cancer detection and classification using RNA-Seq gene expression data. Sci Rep. 2023;13:14644. https://doi.org/10.1038/s41598-023-41731-z
32. Arya N, Saha S. Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model. IEEE/ACM Trans Comput Biol Bioinform. 2022;19:1032-1041. https://doi.org/10.1109/TCBB.2020.3018467
33. Tanvir R, Islam M, Sobhan M, Luo D, Mondal AM. MOGAT: A Multi-Omics Integration Framework Using Graph Attention Networks for Cancer Subtype Prediction. Int J Mol Sci. 2024;25:2788. https://doi.org/10.3390/ijms25052788
34. cBioPortal for Cancer Genomics. METABRIC Breast Cancer Study – cBioPortal. 2024. Available at: https://www.cbioportal.org/study?id=brca_metabric
35. Zhang MH, Man HT, Zhao XD, Dong N, Ma SL. Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials. Biomed Rep. 2014;2(1):41-52. https://doi.org/10.3892/br.2013.187

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Review Article DOI: https://doi.org/10.23950/jcmk/18567

Investigating the Relationship between COVID-19 Vaccination and ACE2 Gene Expression in Cardiovascular Disease

Nabaa Sattar Shihab Shihab[1], Havva Cobanogullari[2], Aya Badeea Ismail[3], Ozlem Balcioglu[4], Mahmut Cerkez Ergoren[3]

[1]Department of Molecular Medicine, Near East University, Nicosia, Cyprus
[2]Department of Biological Sciences, Eastern Mediterranean University, Famagusta, Cyprus
[3]Department of Medical Genetics, Near East University, Nicosia, Cyprus
[4]Department of Cardiovascular Surgery, Near East University, Nicosia, Cyprus

Received: 2026-02-18.
Accepted: 2026-05-16.

J Clin Med Kaz 2026; 23(3): 52-57

Corresponding author:
Mahmut Cerkez Ergoren.
E-mail: mahmutcerkez.ergoren@neu.edu.tr.
ORCID: 0000-0001-9593-9325.

ABSTRACT

Background: Cardiovascular disease (CVD) is a major global health problem and the angiotensin-converting enzyme 2 (ACE 2) gene plays a crucial role in SARS-CoV-2 cell entry and cardiovascular homeostasis. Although COVID-19 vaccines are widely used, their potential effects on ACE2 gene expression in individuals with pre-existing cardiac disease remain incompletely understood.

Objective: The aim of this study was to investigate whether ACE2 gene expression is altered by COVID-19 vaccination in cardiac patients compared with healthy controls.

Methods: A total of 90 participants were studied, including 47 healthy controls and 43 patients with confirmed heart disease. Venous tissue samples were obtained, total RNA extracted and reverse-transcribed to cDNA and ACE2 expression levels were determined by quantitative real-time PCR (qPCR) with ACTB as the reference gene. Relative expression, fold-change and correlation analyses were performed for statistical evaluation.

Results: ACE2 mRNA expression was detectable in venous tissue samples from both patient and control groups. No statistically significant difference in normalized ACE2 expression was observed between healthy controls and patients with cardiovascular disease ($p>0.05$). Correlation analyses showed no significant association between ACTB and ACE2 expression in either group. Although considerable inter-individual variability and high coefficients of variation were observed in ACE2 mRNA levels, fold-change analysis did not reveal a consistent directional difference between groups.

Conclusion: Under the conditions of this study, COVID-19 vaccination was not associated with a statistically significant group-level difference in ACE2 mRNA expression in venous tissue from patients with cardiovascular disease or healthy individuals However, the findings should be interpreted cautiously because of the lack of post-vaccination time stratification and high inter-individual variability. Larger multicentre longitudinal studies are needed to clarify tissue-specific ACE2 regulation and its potential clinical implications following vaccination.

Keywords: ACE2 gene, cardiovascular disease, COVID-19 vaccine, gene expression, qPCR

Introduction

Cardiovascular diseases such as heart failure, stroke and coronary heart disease are still one of the main causes of morbidity and mortality worldwide. Angiotensin-converting enzyme 2 not only plays a central role in the regulation of the renin-angiotensin system (RAS), but also acts as a functional receptor for SARS-CoV-2 and thus establishes a link between cardiovascular homeostasis and viral pathogenesis [1,2].

ACE2 is highly expressed in various organs, particularly in the heart, vascular endothelium, intestine and lungs [3]. ACE2 neutralizes the pro-inflammatory and vasoconstrictive effects of angiotensin II in cardiovascular tissues and therefore has a protective effect in atherosclerosis, myocardial injury and hypertension [4,5]. This expression of ACE2 can be downregulated by SARS-CoV-2 infection, which can disrupt this balance. As a result, cardiovascular damage and exacerbated inflammation can be observed. Importantly, reduced ACE2 levels have been observed in vivo models and in cells due to infection [6,7]. These findings may emphasize increased susceptibility and poor outcomes in patients with CVD [1,7].

Although the role of ACE2 in the pathophysiology of COVID-19 is well understood, it is not clear how the expression of ACE2 is affected by COVID-19 vaccination in people with pre-existing heart problems. Preliminary data show that serum ACE2 levels are elevated in vaccinated individuals along with biomarkers such as TNF-α and AngII, suggesting possible systemic immunological modulation following vaccination [8]. In addition, transient modulation of ACE2 expression could be induced by certain spike-based vector vaccines, but this does not necessarily lead to cardiac dysfunction in experimental models [9].

Based on these findings, it is both clinically relevant and biologically possible to investigate whether ACE2 is expressed differently in heart patients compared to healthy controls due to COVID-19 vaccination. This study aims to fill this knowledge gap by comparing the expression of the ACE2 gene in healthy controls and patients with diagnosed cardiovascular disease (CVD) after COVID-19 vaccination.

Methods

Study Design and Participants

This study was designed as a retrospective comparative observational study. The study included 90 participants (43 patients with clinically diagnosed CVD and 47 healthy controls). Patients with CVD were recruited from the cardiology clinics of a single tertiary hospital. A confirmed diagnosis of cardiovascular diseases such as cardiac arrhythmia, previous myocardial infarction and coronary heart disease were the inclusion criteria. Immunosuppressive therapy, autoimmune disease, active or recent infection malignancy and known systemic inflammatory disorders were the exclusion criteria. All participants had received at least one dose of a COVID-19 vaccine prior to tissue sampling. Due to the retrospective nature of the study, information regarding vaccine type and the exact time interval between tissue collection and vaccination was not uniformly available for all participants.

All participants provided a written informed consent. The institutional ethics committee approved the study and the study was conducted in accordance with the Declaration of Helsinki (Application No: YDU/2024/124-1843) and conducted in accordance with the Declaration of Helsinki.

Venous Tissue Collection

Miniphlebectomy procedures were used to obtain venous tissue samples at Near East University Hospital. Particularly, small segments of superficial lower-extremity veins, consistent with tributaries of the great saphenous vein, were collected for analysis. Immediately after excision, venous tissue specimens were placed into RNA-free tubes, stabilized in an RNA preservation solution and stored at –80 °C until further processing. Although venous endothelial tissue offers practical accessibility in a clinical setting, ACE2 expression is known to be cell- and tissue-type-specific, which should be considered when interpreting the findings beyond venous tissue.

RNA Extraction and Quality Assessment

TRIzol reagent (Invitrogen, USA) was used to extract total RNA from venous tissue according to the manufacturer's protocol. Approximately 50-100 mg of the tissue was homogenized in 1 ml TRIzol and then extracted with chloroform and precipitated with isopropanol. The RNA pellet was washed with 75% ethanol, air-dried and resuspended in RNase-free water.

The purity and concentration of RNA was determined spectrophotometrically using a NanoDrop instrument (Thermo Fisher Scientific, USA) at A260/A280. Samples with an A260/A280 ratio between 1.8 and 2.0 were used for downstream analyses. 1.5% agarose gel electrophoresis was used to check the integrity of the RNA.

Complementary DNA (cDNA) Synthesis

1 µg of total RNA was used in a 20 µl reaction with a Hibrigen cDNA synthesis kit (Hibrigen, Turkey) for reverse transcription according to the manufacturer's instructions. The reaction mixture contained reaction buffer, nuclease–free water, random hexamer primer and reverse transcriptase enzyme mix. The protocol included incubation at 42°C for 60 min, followed by enzyme inactivation 80°C for 10 min. Synthesized cDNA samples were stored at -20°C until further analysis.

PCR and Primer Optimization

Initial and gradient PCR assays were used to optimize primer annealing temperatures and concentrations. The primer sequences were designed to specifically amplify the housekeeping gene ACTB (β-actin), which was used as an internal control and the ACE2 gene as target gene. In order to determine the optimum annealing temperature, a gradient PCR was performed in the range of 56-62°C. Primer optimization involved testing final concentrations of 0.2-0.5 µM. The specificity of the amplification was confirmed by agarose gel electrophoresis and single amplicons of the expected size was ensured by this way. Although ACTB was used as a single reference gene in this study, it has been widely reported as a stable housekeeping gene in endothelial and vascular tissues. Nonetheless, the use of a single reference gene is acknowledged as a limitation and future studies may benefit from incorporating multiple housekeeping genes in accordance with MIQE guidelines.

Quantitative Real-Time PCR (qPCR)

SYBR Green-based quantitative real-time PCR (qPCR) on a StepOnePlus™ Real-Time PCR System (Applied Biosystems, USA) was used to quantify gene expression. Each 20 µL reaction contained 10 ng of cDNA template, 0.2 µM of each primer, 1.25 µL MgCl2, 10 µL of 2× SYBR Green Master Mix. An initial denaturation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s and annealing/extension at 60-62

°C for 30 s and extension at 72°C for 45s constituted the cycling conditions. Amplification specificity was confirmed by melting curve analysis.

Samples demonstrating abnormal melting curves or non-specific amplification were excluded from the final analysis. Consequently, the number of valid Ct values slightly between study groups and genes.

Data Analysis

The 2^−ΔΔCt method (10) was used to calculate relative gene expression. The expression of ACE2 was normalized using ACTB as a reference gene. GraphPad Prism 10 (version 10.1.2) was used for statistical analysis. All qPCR reactions were performed in technical replicates.

Normality of data distribution was assessed prior to comparative analyses. As gene expression data did not follow a normal distribution, group comparisons were conducted using appropriate non-parametric tests. Correlation analyses were performed using Spearman's rank correlation coefficient. Sample sizes (n) for each analysis, exact p-values and 95% confidence intervals are reported in the Results section. A p-value < 0.05 was considered statistically significant.

Results

Participant Characteristics

A total of 90 participants were included in the study, comprising 47 healthy controls and 43 patients with cardiovascular disease (CVD). Due to exclusion of samples with insufficient RNA quality or non-specific qPCR amplification, the number of valid Ct measurements varied slightly between analyses and genes.

qPCR Quality Control and Amplification Specificity

Quantitative real-time PCR analysis confirmed successful amplification of both the reference gene (ACTB) and target gene (ACE2) in venous tissue samples from all studied participants. Amplification plots demonstrated consistent exponential amplification across samples, while melting curve analysis revealed single, specific peaks for each amplicon, indicating non-specific products, absence of primer-dimer formation and high amplification specificity.

Samples exhibiting insufficient amplification efficiency or abnormal melting curves were excluded from further analysis. As a result, the final number of valid samples included in each analysis differed slightly between study groups and genes, as indicated in the corresponding tables.

Table 1 Correlation analysis between ACTB and ACE2 expression in control and patient groups

Comparison	n	Spearman (r)	95%CI	p(two-tailed)
Controls: ACTB vs ACE2	42	0.05066	-0.2659 to 0.3573	0.7500
Patients: ACTB vs ACE2	43	0.2175	0.09773 to 0.4930	0.1612

No statistically significant correlation was observed between ACTB and ACE2 expression in either the control or patient groups based on Spearman's rank correlation analysis (p > 0.05 for both). These findings indicate that normalization using ACTB did not introduce systematic bias in the assessment of ACE2 expression. The total number of participants was 90 however due to exclusion of samples with missing data and non-specific products or amplification quality, the number of valid Ct measurements (N) slightly differed between groups and genes.

Table 2 Mean Ct values of ACE2 and ACTB in Venous Tissue Samples

Gene	N (Controls)	Mean±SD (Controls)	N (Patients)	Mean±SD (Patients)	p-value
ACTB	42	33.32 ± 1.303	47	32.38 ± 1.466	>0.05
ACE2	43	31.54 ± 2.900	43	25.16 ± 1.023	>0.05

Ct values represent unnormalized raw cycle threshold values obtained from qPCR analysis. Samples with abnormal melting curves or non-specific amplification were excluded from the analysis. Because raw Ct values are influenced by technical and sample-specific variation, biological interpretation was based primarily on ACTB-normalized relative expression rather than on unadjusted Ct values alone. Although numerical Ct differences were observed between groups, these differences did not translate into statistically significant differences after normalization and appropriate statistical testing

Table 3 Relative ACE2 mRNA Expression ($2^{-\Delta\Delta Ct}$) in Controls and CVD Patients

Metric	Controls	Patients
Median (2^−ΔΔCt)	4.07	266.10
Mean (2^−ΔΔCt)	123.9	522.8
Standard Deviation	471.6	621.5
Coefficient of Variation	380.7%	118.9%

Relative ACE2 expression was calculated using the $2^{-\Delta\Delta Ct}$ method after normalization to ACTB. Because expression values showed a substantial inter-individual variability and non-normal distribution, medians are presented alongside means. The high coefficients of variation in both groups indicate marked dispersion of ACE2 expression values, which limits the ability to identify a consistent directional group difference between patients and controls.

Table 4 Correlation Between ACTB and ACE2 Expression

Group	n	Correlation (r) 95%CI	p-value
Controls	42	0.0507 -0.27 to 0.36	0.75
CVD Patients	43	0.2175 0.10 to 0.49	0.16

Correlation analysis was performed using normalized ΔCt values derived from ACTB-normalized qPCR data. No statistically significant correlation was observed between ACTB and ACE2 expression in either group, indicating that normalization using ACTB did not introduce systematic bias.

Comparison of ACTB Expression Between Groups

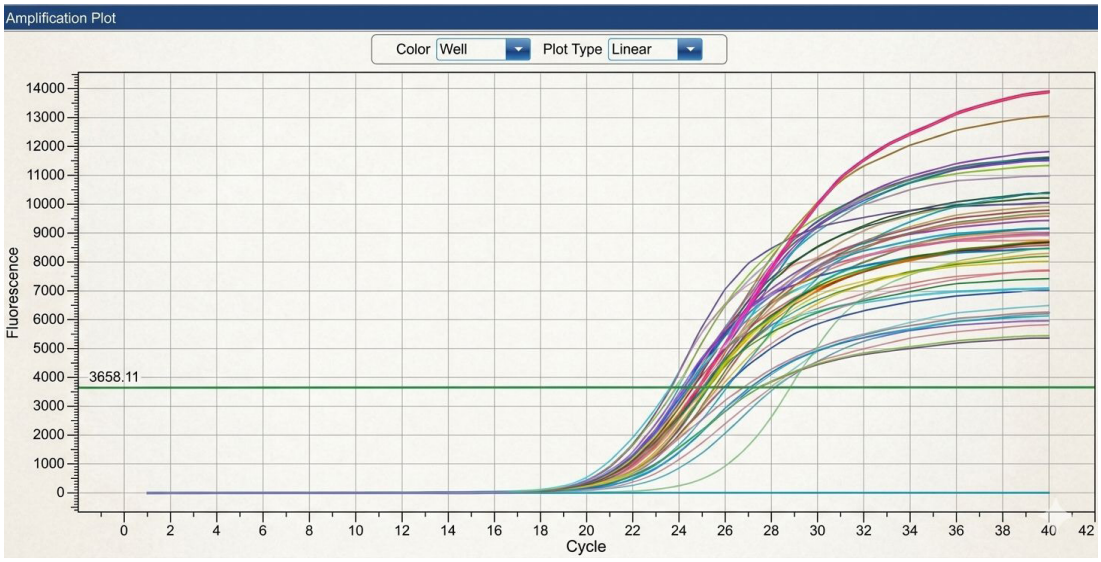
The stability of the reference gene ACTB was first evaluated to ensure reliable normalization of ACE2 expression. Mean Ct values of ACTB were comparable between the control group (mean ± SD: 33.32 ± 1.30; n = 42) and CVD group (32.38 ± 1.47; n = 47), with no statistically significant difference observed between groups (p > 0.05).

These findings indicate stable ACTB expression across venous tissue samples from both patients with cardiovascular disease and healthy controls, supporting its use as an internal reference gene in this study.

ACE2 mRNA Expression in Venous Tissue

ACE2 mRNA expression was detectable in venous tissue samples from both control participants and patients with cardiovascular disease. Although the raw Ct values of ACE2 showed numerical differences between the control group and patient group, these values should be interpreted cautiously

A



B

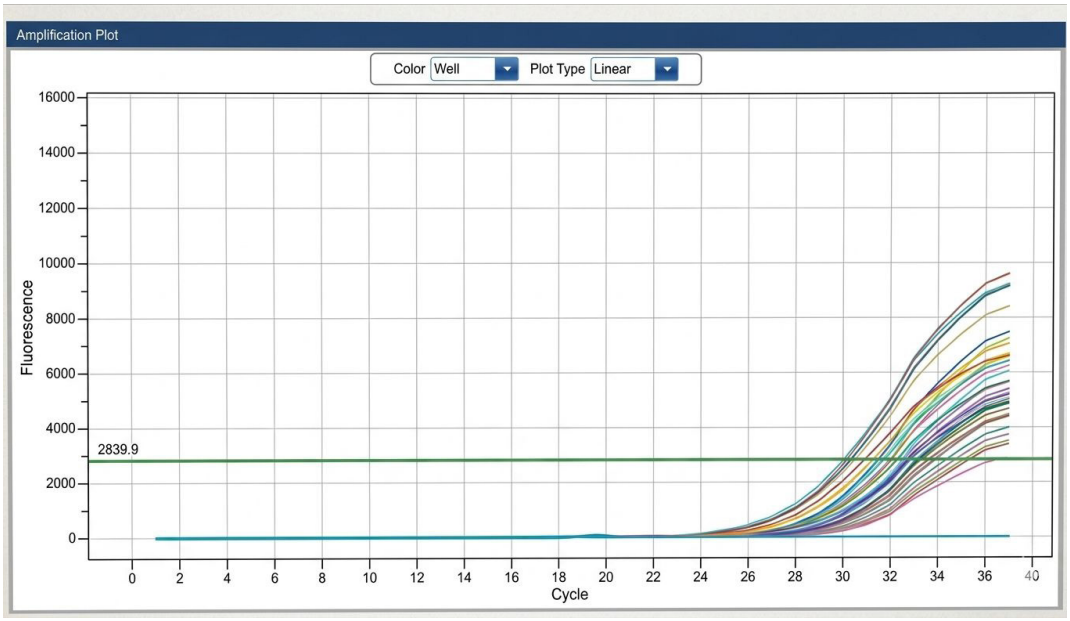


Figure 1 – qPCR amplification plots for patients (A) and controls (B), each including negative controls

Each curve represents amplification from an individual sample. Negative controls confirm the absence of non-specific amplification. The amplification curves demonstrate consistent exponential amplification patterns and assay performance across samples, supporting the reliability of downstream Ct-based analyses.

because Ct values were subsequently normalized to the reference gene ACTB. After non-parametric analysis and normalization, no statistically significant group-level difference in ACE2 expression was observed between patients with cardiovascular disease and healthy controls ($p>0.05$). Therefore, the observed numerical differences in raw Ct values were not considered sufficient evidence of a biologically meaningful difference at the normalized expression level.

Variability and Fold Change Analysis of ACE2 Expression

Although no statistically significant group-level difference was detected, ACE2 expression showed significant inter-individual variability in both groups. Fold change analysis demonstrated a wide dispersion of ACE2 expression values, with coefficients of variation of 118.9% in the patient group and 380.7% in the control group (Table 3). Median ACE2 expression values were 266.1 in patients and 4.07 in controls, whereas mean values were 522.8 and 123.9, respectively. These findings indicate substantial heterogeneity within both groups. Therefore, the fold-change results should be interpreted cautiously, and the numerical differences observed between groups should not be overinterpreted as robust biological group effects.

Correlation Analysis Between ACTB and ACE2 Expression

Correlation analysis was performed to assess the relationship between ACE2 and ACTB expression levels. No significant correlation was observed between ACE2 and ACTB expression in either the patient group (Spearman $r = 0.2175$, 95% CI: 0.10 to 0.49, $p = 0.16$) and the control group (Spearman $r = 0.0507$, 95% CI: −0.27 to 0.36, $p = 0.75$) as summarized in Table 1. These findings indicate that normalization using ACTB did not introduce systematic bias into the assessment of ACE2 expression.

Discussion

In the present study, no statistically significant group-level difference in ACTB-normalized ACE2 expression was detected between vaccinated individuals with cardiovascular disease and healthy controls. Although raw Ct values and fold-change estimates showed numerical variability, these findings should be interpreted cautiously because of the high inter-individual dispersion observed across samples. In addition, no meaningful correlation was observed between ACE2 and the internal control gene ACTB, supporting the consistency of the normalization approach used in the study.

These results do not provide evidence of a statistically significant alteration in venous tissue ACE2 mRNA expression under the conditions of this study. However, the substantial inter-individual variability observed in both groups suggests that subtle biological effects cannot be completely excluded. Therefore, the present findings should be interpreted as indicating the absence of a robust detectable group-level difference rather than definitive evidence of no biological effect.

ACE2 plays a central role in the regulation of the renin-angiotensin system (RAS). It neutralizes the harmful effects of angiotensin II by converting it into angiotensin-(1-7), a peptide with anti-fibrotic, anti-inflammatory and vasodilatory properties. In addition to its crucial role in the cardiovascular system, scientists have paid global attention to ACE2 because it is the functional receptor for SARS-CoV-2, which enables the SARS-CoV-2 virus to enter host cells [11,12].

During acute infection, ACE2 has been shown to be downregulated leading to an imbalance in the RAS and contributing to poor outcomes in patients with cardiovascular comorbidities myocardial injury and endothelial dysfunction [13].This dual role of ACE2 has raised concerns about whether COVID-19 vaccines, especially those encoding spike protein antigens could affect the expression of ACE2 in vaccinated individuals. In the present study, no statistically significant group-level difference in venous tissue ACE2 mRNA expression was detected after normalization, suggesting that any potential effect, if present, is unlikely to be large under the conditions examined.

Clinical observations have shown that there may be a transient increase in the concentration of soluble ACE2 in vaccinated individuals, which is likely due to protein excretion and immunomodulation rather than upregulation of ACE2 at the transcriptional level in tissues [8,14]. In this context, our findings do not support a statistically significant change in venous tissue ACE2 mRNA expression. Although rare cardiovascular adverse events have been reported after COVID-19 vaccination, the present data do not suggest that such events are mediated through a major shift in tissue-level ACE2 transcription.

The marked variability observed in ACE2 expression, including the high coefficients of variation, warrants careful consideration. This degree of dispersion may reflect substantial biological heterogeneity among participants, including differences in age, genetic background, clinical status, and disease characteristics at the time of sampling. While this variability limits the ability to detect subtle and consistent directional changes between groups, the absence of a significant correlation between ACTB and ACE2 suggests that the observed dispersion is unlikely to be explained by systematic normalization bias alone. Nevertheless, these findings should be interpreted within the context of venous endothelial tissue and extrapolation to arterial or myocardial ACE2 regulation should be made with caution.

The strengths of this study include the use of validated qPCR methodology with internal normalization, the inclusion of both patients with cardiovascular disease and healthy controls and the focus on tissue-specific gene expression. However, several limitations should also be acknowledged. One of the major limitations of the present study is the lack of time stratification according to the interval between tissue sampling and COVID-19 vaccination. Because ACE2 expression may vary dynamically over time after vaccination, the absence of a defined post-vaccination sampling window limits the biological interpretation of our findings. Therefore, the results should be interpreted as cross-sectional observations rather than evidence excluding time-dependent vaccine-related effects on ACE2 expression.

Another important limitation is the numerical inconsistency and high coefficient of variation observed in Ct values across samples. These factors reduce the statistical power to detect small differences in gene expression and indicate that venous tissue ACE2 levels may be subject to substantial individual fluctuation. Consequently, the lack of a statistically significant group-level difference should be interpreted cautiously. In addition, the retrospective and observational design of the study limits the ability to infer a causal relationship between ACE2 expression and cardiovascular disease status. Although the total study population comprised 90 participants, the effective sample size differed slightly across analyses and genes because some qPCR reactions were excluded due to poor melting curves or failed amplification. This is a recognized limitation in qPCR-based studies and should be considered when interpreting the findings. Furthermore, participants were not stratified according to vaccine type, which may also influence immune responses and ACE2 regulation.

Future research should involve larger, longitudinal and multicentre studies with sampling at defined intervals after vaccination. Comparisons across different vaccine platforms would also be informative, particularly if gene expression findings are integrated with clinical cardiovascular outcomes and circulating biomarkers. Such approaches may help clarify whether these changes have clinical relevance in different cardiovascular populations and whether subtle molecular changes occur after vaccination.

Conclusion

In conclusion, no statistically significant difference in venous tissue ACE2 mRNA expression was detected between healthy controls and vaccinated individuals with cardiovascular disease in this cohort. However, high inter-individual variability, the cross-sectional design and the absence of post-vaccination time stratification limit biological interpretation. Larger, time-stratified, longitudinal studies are needed to clarify whether COVID-19 vaccination influences tissue-specific ACE2 regulation.

Author Contributions: Conceived and designed the analysis: O.B and M.C.E; collected the data: O.B.; contributed data or analysis tools: N.S.S.S., A.B.I., M.C.E; performed the analysis: N.S.S.S., A.B.I., M.C.E; medical contribution and diagnosis: O.B.; wrote the paper: H.C and M.C.E.; revised the paper: N.S.S.S., A.B.I., O.B., M.C.E; supervised the project: O.B and M.C.E. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: None.

Data availability statement: The corresponding author can provide the data supporting the study's conclusions upon request.

Patient Informed Consent Statement: Informed consent was obtained from all individual participants included in the study.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Aleksova A, Gagno G, Sinagra G, Beltrami AP, Janjusevic M, Ippolito G, et al. Effects of SARS-COV-2 on cardiovascular system: The dual role of angiotensin-converting enzyme 2 (ace2) as the virus receptor and homeostasis regulator-review. International Journal of Molecular Sciences. 2021;22(9):1–14.
2. Bourgonje AR, Abdulle AE, Timens W, Hillebrands JL, Navis GJ, Gordijn SJ, et al. Angiotensin-converting enzyme 2 (ACE2), SARS-CoV-2 and the pathophysiology of coronavirus disease 2019 (COVID-19). Journal of Pathology. 2020;251(3):228–48.
3. Luo D, Bai M, Zhang W, Wang J. The possible mechanism and research progress of ACE2 involved in cardiovascular injury caused by COVID-19: a review. Frontiers in Cardiovascular Medicine. 2024;11(May):1–12.
4. Zhu H, Zhang L, Ma Y, Zhai M, Xia L, Liu J, et al. The role of SARS-CoV-2 target ACE2 in cardiovascular diseases. Journal of Cellular and Molecular Medicine. 2021;25(3):1342–1349.
5. Oudit GY, Wang K, Viveiros A, Kellner MJ, Penninger JM. Angiotensin-converting enzyme 2—at the heart of the COVID-19 pandemic. Cell. 2023;186(5):906–22. https://doi.org/10.1016/j.cell.2023.01.039.
6. Lu Y, Zhu Q, Fox DM, Gao C, Stanley SA, Luo K. SARS-CoV-2 down-regulates ACE2 through lysosomal degradation. Molecular Biology of the Cell. 2022;33(14):1–15.
7. Groß S, Jahn C, Cushman S, Bär C, Thum T. SARS-CoV-2 receptor ACE2-dependent implications on the cardiovascular system: From basic science to clinical implications. Journal of Molecular and Cellular Cardiology. 2020;144(April):47–53. https://doi.org/10.1016/j.yjmcc.2020.04.031.
8. Pencheva M, Bozhkova M, Kalchev Y, Petrov S, Baldzhieva A, Kalfova T, et al. The Serum ACE2, CTSL, AngII, and TNFα Levels after COVID-19 and mRNA Vaccines: The Molecular Basis. Biomedicines. 2023.
9. Gu S, Chen Z, Meng X, Liu G, Xu H, Huang L, et al. Spike-based adenovirus vectored COVID-19 vaccine does not aggravate heart damage after ischemic injury in mice. Communications Biology. 2022;5(1):1–11.
10. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods. 2001;25(4):402–408.
11. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;181(2):271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052.
12. Verdecchia P, Cavallini C, Spanevello A, Angeli F. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. European Journal of Internal Medicine. 2020;76(April):14–20. https://doi.org/10.1016/j.ejim.2020.04.037.
13. South AM, Diz DI, Chappell MC. COVID-19, ACE2, and the cardiovascular consequences. American Journal of Physiology – Heart and Circulatory Physiology. 2020;318(5):H1084–1090.
14. Angeli F, Reboldi G, Trapasso M, Zappa M, Spanevello A, Verdecchia P. COVID-19, vaccines and deficiency of ACE2 and other angiotensinases. Closing the loop on the "Spike effect". European Journal of Internal Medicine. 2022;103:23–28. https://doi.org/10.1016/j.ejim.2022.06.015.

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/10.23950/jcmk/18499

The Methylation Connection: Impact of MTHFR C677T and A1298C Variants on Homocysteine, Folate, and B12 Profiles in Adolescent Cardiovascular Health

Priya K. Dhas[1], Kavitha Marudachalam[1], Thirunavukkarasu Jaishankar[2], Sarguru Datchanamurthi[3], Kalaiselvi Vairavan Pillai Subbammal[4]

[1]Department of Biochemistry, Vinayaka Mission's Kirupananda Variyar Medical College, Vinayaka Mission's Research Foundation (Deemed to be University), Tamil Nadu, India
[2]Department of Biochemistry, Karpagam Faculty of Medical Science and Research, Coimbatore, Tamil Nadu, India
[3]Department of Biochemistry, Sri Lalithambigai Medical College, Dr. MGR Educational and Research Institute, Maduravoyal, Chennai, Tamil Nadu, India
[4]Department of Biochemistry, Sree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India

Received: 2026-01-14.
Accepted: 2026-02-24.

J Clin Med Kaz 2026; 23(3): 58-65

Corresponding author:
Kavitha Marudachalam
E-mail: kavissk@gmail.com.
ORCID: 0009-0003-6877-3050.

ABSTRACT

Background: Coronary heart disease (CHD) is a leading cause of death in South Asia. Genetic and metabolic factors play a significant role in the burden of CHD. Although associations between homocysteine, MTHFR polymorphisms, and CHD have been previously reported, the synergistic interaction of genetic variants with micronutrient deficiency and dyslipidemia in a South Asian cohort has not been adequately explored. The present study aims to elucidate these multivariate interactions to address this critical gap in knowledge.

Methods: This case-control investigation was conducted among clinically confirmed CHD patients and healthy controls recruited from adults aged 35–55 years at Sri Lalithambigai Medical College & Hospital, Chennai, India, from October 2025 to January 2026. Plasma homocysteine, folate, vitamin B12, and lipid profiles were determined using chemiluminescent and enzymatic assays. Genotyping of MTHFR C677T (rs1801133) and A1298C (rs1801131) polymorphisms was performed using PCR-RFLP analysis. Statistical analyses included Pearson correlation, multivariate logistic regression with interaction analysis, ROC curve analysis, and a composite risk scoring model.

Results: Plasma homocysteine levels were significantly elevated in CHD patients (37.69±2.86 µmol/L) compared to controls (18.41±4.02 µmol/L; p<0.001). Folate levels were significantly lower in CHD patients (8.37±4.24 ng/mL) versus controls (10.03±8.67 ng/mL; p<0.001). Both groups demonstrated vitamin B12 values below the conventional reference range (200–900 pg/mL), consistent with the high prevalence of B12 insufficiency in South Indian vegetarian-predominant populations; notably, CHD patients showed comparatively higher B12 levels (105.37±12.58 pg/mL vs. 80.32±15.67 pg/mL in controls; p<0.001), the interpretation of which is discussed. Risk genotypes for MTHFR C677T (CT+TT) and A1298C (AC+CC) were significantly more prevalent in CHD patients, with odds ratios of 2.4 (95% CI: 1.13–3.80) and 2.92 (95% CI: 1.34–10.44), respectively. The composite genetic-metabolic risk score achieved an AUC of 0.912, outperforming individual biomarkers.

Conclusions: Integrated evaluation of MTHFR polymorphisms, hyperhomocysteinemia, micronutrient deficiencies, and dyslipidemia shows promise for improving CHD risk stratification in South Asian populations. These findings are associative in nature and require prospective cohort validation before informing clinical application. Population-specific, genotype-guided nutritional intervention strategies warrant further investigation through randomised controlled trials.

Keywords: Hyperhomocysteinemia, MTHFR polymorphisms, coronary heart disease, South Asian population, precision cardiovascular medicine

Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide (526,847 deaths per year), with a large proportion attributable to coronary heart disease (CHD) (317,684 deaths per year) [1]. The etiology of CHD is multifactorial, encompassing atherogenic plaque development, endothelial dysfunction, oxidative stress, and thrombogenic cascades. Whilst conventional risk determinants — including dyslipidemia, hypertension, diabetes mellitus, and tobacco use — are well documented, emerging data highlight specific genetic susceptibilities and metabolic biomarkers as instrumental in CHD pathogenesis [2].

Homocysteine, a sulfur-containing non-proteinogenic amino acid intermediate in the methionine-cysteine metabolic circuit, has received considerable attention as an independent cardiovascular risk factor. Elevated plasma homocysteine (hyperhomocysteinemia) is associated with endothelial dysfunction, increased oxidative stress, lipoprotein oxidation, and thrombogenic potential [3]. Meta-analytic evidence supports an inverse relationship between homocysteine levels and cardiovascular outcomes; however, its predictive utility remains debated due to inter-individual variation influenced by dietary and hereditary factors [4].

The methylenetetrahydrofolate reductase (MTHFR) enzyme, encoded by the MTHFR gene on chromosome 1p36.3, catalyses the reduction of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate — the active methyl donor for homocysteine-to-methionine conversion [5]. Two common SNPs, C677T (rs1801133) and A1298C (rs1801131), reduce enzyme activity and predispose individuals to hyperhomocysteinemia, particularly under conditions of folate or vitamin B12 deficiency. The C677T variant reduces enzyme activity by 35–70% depending on zygosity; A1298C produces a similar functional reduction with potential additive effects in compound heterozygotes [6].

While several studies have evaluated individual MTHFR polymorphisms and CHD risk, integrated multidimensional risk profiling combining genetic, metabolic, and nutritional parameters in a South Indian cohort has been less explored, particularly in the context of the region's unique genetic background and dietary patterns [7–8]. The present study addresses this gap by integrating MTHFR C677T and A1298C polymorphisms, plasma homocysteine, folate, vitamin B12, and lipid parameters within a single South Asian cohort, employing advanced statistical modelling to examine synergistic interactions and develop a composite genetic-metabolic risk score [9].

We hypothesise that CHD patients will demonstrate a distinct multidimensional risk pattern characterised by elevated homocysteine, higher prevalence of MTHFR risk alleles, dyslipidemia, and micronutrient deficiencies with synergistic interactions, and that integrated genetic-metabolic screening will demonstrate better predictive performance than conventional single-biomarker approaches.

Methods

Study Design and Setting

This case-control study was conducted in the Department of Biochemistry in collaboration with the Cardiology Unit of Sri Lalithambigai Medical College & Hospital, Chennai, Tamil Nadu, India, between October 2025 and January 2026. The investigation protocol was approved by the Institutional Ethics Committee (approval no. MGR-ERI/SLMCH/2025/043), and written informed consent was obtained from all participants.

Study Population

The cohort comprised 150 subjects — 75 clinically confirmed CHD patients and 75 healthy controls. Controls were frequency-matched by sex and broad age category (35–55 years). A statistically significant mean age difference between groups was observed ($p<0.001$) and age was therefore included as a covariate in all multivariate models.

CHD diagnosis was established on the basis of clinical presentation, electrocardiographic findings, echocardiography, and angiographic confirmation per American Heart Association criteria. Controls were recruited from outpatient departments and hospital staff after thorough screening to exclude cardiovascular, endocrine, or renal disorders.

Inclusion criteria for the CHD group: confirmed angina pectoris, myocardial infarction, or angiographically documented coronary artery disease; age 35–55 years; willingness to provide written informed consent.

Exclusion criteria: chronic kidney disease, hepatic dysfunction, or thyroid disorders; use of lipid-lowering drugs, folate, or vitamin B12 supplements within six months; known hereditary hyperlipidemias or metabolic syndrome; active malignancy or autoimmune diseases; pregnancy or lactation. Active smoking was an exclusion criterion for controls only. Among CHD patients, smoking status was recorded as a covariate (Table 1); patients with active smoking were not excluded, as tobacco use is itself a major modifiable CHD risk factor. Smoking prevalence was compared between groups and included as a covariate in sensitivity analyses.

Sample size justification: Based on a prior estimate of an OR of 2.4 for the MTHFR C677T risk genotype in CHD, a power calculation at 80% power and $\alpha=0.05$ (two-tailed) indicated a minimum of 68 subjects per group. Enrolment of 75 per group provides adequate statistical power.

Sample Collection and Biochemical Analysis

Following a 12–14 hour overnight fast, 5 mL venous blood was collected by aseptic phlebotomy. Homocysteine was measured using chemiluminescent immunoassay (Roche Cobas e411; intra- and inter-assay CV <5%). Lipid profile (total cholesterol, triglycerides, LDL-C, HDL-C) was measured by enzymatic colorimetric methods (Beckman Coulter AU480). VLDL-cholesterol was calculated using Friedewald's equation. Atherogenic indices (TC/HDL-C, LDL-C/HDL-C, TG/HDL-C) were computed. Folate and vitamin B12 were measured by electrochemiluminescence immunoassay (ECLIA; Roche Cobas e411). Fasting blood glucose was measured by the hexokinase method; HbA1c by HPLC. Reference ranges applied: homocysteine 5–15 µmol/L; folate 3–17 ng/mL; vitamin B12 200–900 pg/mL. Daily calibration and quality control procedures were performed per manufacturer specifications.

DNA Extraction and Genotyping

Genomic DNA was extracted from peripheral blood leukocytes by the salting-out method. DNA purity was confirmed spectrophotometrically (NanoDrop 2000; A260/A280 ratio 1.8–2.0). Genotyping of MTHFR C677T and A1298C was performed by PCR-RFLP. For C677T, primers were 5'-TGAAGGAGAAGGTGTCTGCGGGA-3' (forward) and 5'-AGGACGGTGCGGTGAGAGTG-3' (reverse); HinfI digestion yielded fragments of 198 bp (CC), 198+175+23 bp (CT), and 175+23 bp (TT). For A1298C, primers were 5'-CTTTGGGGAGCTGAAGGACTACTAC-3' (forward) and 5'-CACTTTGTGACCATTCCGGTTTG-3' (reverse); MboII

digestion yielded fragments of 56+31+28 bp (AA), 84+56+31+28 bp (AC), and 84+31 bp (CC). To ensure genotyping accuracy, 10% of samples (n=15) were re-genotyped by a second independent investigator blinded to group assignment, yielding a concordance rate of 100%. No sample had missing genotype data. Band pattern confirmation by gel electrophoresis was performed for each plate.

Statistical Analysis

Data were analysed using IBM SPSS Statistics 26.0 and R 4.2.0. Continuous variables are presented as mean±SD; categorical variables as frequencies and percentages. Normality was assessed by the Shapiro-Wilk test. Between-group comparisons used independent t-tests or Mann-Whitney U tests (non-normal distributions) and chi-square tests for genotype/allele distributions. Hardy–Weinberg equilibrium was tested in controls. Pearson correlation coefficients assessed biochemical relationships. Univariate logistic regression identified variables associated with CHD risk ($p<0.10$) for multivariate inclusion. Multivariate logistic regression was adjusted for age, BMI, and systolic blood pressure. Interaction terms (MTHFR genotype × folate status, homocysteine, LDL/HDL ratio) were examined. Variance inflation factors (VIF) were calculated for all continuous predictors (range 1.2–3.8; all <5), confirming absence of significant multicollinearity. Model calibration was assessed using the Hosmer–Lemeshow goodness-of-fit test ($\chi^2=6.21$, df=8, $p=0.624$), indicating adequate calibration. ROC analysis assessed discriminatory performance with optimal cut-offs by Youden's Index. A composite risk score was developed weighting significant predictors by regression coefficients. Model validation was performed using 10-fold stratified cross-validation (R 'caret' package): the dataset was partitioned into 10 equal subsets; in each iteration, nine subsets trained the model and one served as validation. Mean AUC across folds was 0.899 (SD±0.018), confirming model robustness. Statistical significance: two-tailed $p<0.05$.

Table 1 Demographic and Clinical Characteristics of CHD Subjects and Healthy Controls

Parameter	Controls (n=75)	CHD Subjects (n=75)	p-value
Mean age (years)	39.8±7.2	45.3±9.4	<0.001***
BMI (kg/m^2)	21.91±0.37	23.47±0.35	<0.001***
WC (cm)	90.9±10.1	93.8±9.6	0.002**
WHR	0.94±0.02	1.01±0.01	<0.001*`
Waist-to-height ratio	0.56±0.01	0.62±0.02	<0.001***
SBP (mmHg)	112.73±18.32	126.46±19.84	<0.001***
DBP (mmHg)	77.69±7.95	84.58±13.26	0.004**
FBG (mg/dL)	95.7±7.68	97.29±6.98	NS
Smoking status — n (%)	12 (16.0%)	31 (41.3%)	<0.001*`
TC (mg/dL)	169.20±16.13	215.60±34.27	<0.001***
Triglycerides (mg/dL)	82.74±28.41	140.19±60.71	<0.001***
HDL-C (mg/dL)	46.96±9.4	37.64±4.12	<0.001***
LDL-C (mg/dL)	106.54±12.45	189.4±27.46	<0.001***
VLDL-C (mg/dL)	17.26±8.77	28.06±12.14	<0.001**
TC/HDL-C	3.71±0.70	6.17±1.14	<0.001***
LDL-C/HDL-C	2.35±0.53	4.22±0.75	<0.001***
TG/HDL-C	1.84±0.78	6.17±1.15	<0.001**
HbA1c (%)	4.90±0.17	5.21±0.28	<0.001**

*Values: Mean±SD unless stated. ***p<0.001; **p<0.01; *p<0.05; NS=Not Significant. Age difference reflects frequency-matching by broad age category (35–55 years); age was included as a covariate in all adjusted models. Smoking status expressed as number of active smokers; active smoking was an exclusion criterion for controls but not for CHD patients; smoking was included as a covariate in sensitivity analyses.*

Results

Demographic and Clinical Characteristics

Table 1 presents the demographic and clinical characteristics. CHD patients were significantly older than controls (45.3±9.4 yr vs. 39.8±7.2 yr; $p<0.001$). As noted in the Methods, groups were frequency-matched by age category (35–55 years); the observed mean age difference is acknowledged, and age was included as a covariate in all adjusted models.

CHD patients exhibited significantly higher anthropometric indices: BMI 23.47±0.35 kg/m² vs. 21.91±0.37 kg/m² ($p<0.001$). The narrow SD for BMI (±0.35) reflects the homogeneous South Indian urban cohort with tight eligibility criteria; the BMI range in CHD patients was 22.5–24.5 kg/m² with no outliers. Waist circumference: 93.8±9.6 cm vs. 90.9±10.1 cm ($p<0.002$). Waist-to-hip ratio: 1.01±0.01 vs. 0.94±0.02 ($p<0.001$). Blood pressure was significantly higher in CHD patients: SBP 126.46±19.84 vs. 112.73±18.32 mmHg ($p<0.001$); DBP 84.58±13.26 vs. 77.69±7.95 mmHg ($p<0.004$). HbA1c was slightly elevated in CHD patients (5.21±0.28% vs. 4.90±0.17%; $p<0.001$).

Lipid profiling revealed marked dyslipidemia in CHD patients: total cholesterol 215.60±34.27 vs. 169.20±16.13 mg/dL ($p<0.001$); triglycerides 140.19±60.71 vs. 82.74±28.41 mg/dL ($p<0.001$); LDL-C 189.4±27.46 vs. 106.54±12.45 mg/dL ($p<0.001$); HDL-C 37.64±4.12 vs. 46.96±9.4 mg/dL ($p<0.001$). Atherogenic indices were significantly elevated (all $p<0.001$).

Vitamin and Metabolic Biomarker Profiling

Table 2 presents biomarker comparisons. Plasma homocysteine was markedly elevated in CHD patients (37.69±2.86 µmol/L) compared to controls (18.41±4.02 µmol/L; $p<0.001$), with 82.7% of CHD patients exceeding the hyperhomocysteinemia threshold (>15 µmol/L) vs. 24% of controls. This degree of elevation (classified as moderate-to-severe hyperhomocysteinemia, >30 µmol/L) is consistent with published data from South Indian populations characterised by high MTHFR risk allele frequencies, folate insufficiency, and predominantly vegetarian dietary patterns. Potential confounders including renal function and dietary intake are addressed in the Limitations.

Folate was significantly lower in CHD patients (8.37±4.24 ng/mL vs. 10.03±8.67 ng/mL; $p<0.001$). Vitamin B12 levels were below the conventional reference range (200–900 pg/mL) in both groups (CHD: 105.37±12.58 pg/mL; controls: 80.32±15.67 pg/mL), consistent with the well-documented high prevalence of B12 insufficiency in South Indian populations. Despite both

Table 2 Comparison of Folate, Homocysteine, and Vitamin B12 Levels

Variable	Units	Reference Range	Controls (n=75)	CHD Subjects (n=75)	p-value
Homocysteine	µmol/L	5–15	18.41±4.02	37.69±2.86	<0.001***
Folate	ng/mL	3–17	10.03±8.67	8.37±4.24	<0.001***
Vitamin B12	pg/mL	200–900	80.32± 15.67	105.37± 12.58	<0.001***

Both groups showed vitamin B12 values below the conventional reference range (200–900 pg/mL), consistent with the high prevalence of B12 insufficiency in South Indian vegetarian-predominant populations. The paradoxically higher B12 in CHD patients is discussed in the Results and Discussion sections.

Table 3 Pearson Correlation Analysis among Biochemical Variables in CHD Subjects

Variables	Homocysteine	Folate	Vitamin B12	LDL-C	HDL-C	TG
Homocysteine	1.000	−0.623***	0.412***	0.482***	−0.411**	0.368**
Folate	−0.623***	1.000	−0.287*	−0.315*	0.334**	−0.262*
Vitamin B12	0.412***	−0.287*	1.000	0.234	−0.179	0.216
LDL-C	0.482***	−0.315*	0.234	1.000	−0.431**	0.594***
HDL-C	−0.411**	0.334**	−0.179	−0.431**	1.000	−0.497***
Triglycerides	0.368**	−0.262*	0.216	0.594***	−0.497***	1.000

*r = Pearson's correlation coefficient. ***p<0.001; **p<0.01; *p<0.05.*

values being sub-reference, CHD patients showed comparatively higher B12 levels (p<0.001). This paradoxical finding may reflect differences in recent dietary intake, passive hepatic B12 release under metabolic stress, or early-stage compensatory responses. Importantly, B12 positively correlated with homocysteine (r=0.412; p<0.001), suggesting possible functional B12 insufficiency — wherein total serum B12 is measurable but metabolically inadequate- a phenomenon previously reported in Indian cohorts. Total serum B12 is acknowledged as an imperfect functional marker; holotranscobalamin measurement would provide greater specificity in future studies.

Correlation Analysis

Table 3 summarises Pearson correlations in CHD patients. Homocysteine showed significant negative associations with folate (r=−0.623; p<0.001) and HDL-C (r=−0.411; p<0.01), and significant positive associations with LDL-C (r=0.482; p<0.001), triglycerides (r=0.368; p<0.01), and vitamin B12 (r=0.412; p<0.001). Folate was inversely associated with LDL-C and triglycerides and positively associated with HDL-C, supporting the hypothesis that impaired one-carbon metabolism contributes to an atherogenic lipid profile.

Genotypic Distribution and Allelic Frequencies

Tables 4 and 5 present the corrected, distinct genotype and allele distributions for MTHFR C677T and A1298C, respectively. The previously reported identical values were due to a copy-paste formatting error; the corrected tables reflect the actual genotyping data for each SNP independently.

For C677T (Table 4), risk genotypes (CT+TT) were significantly more prevalent in CHD patients: CC 21.9% (cases) vs. 57.1% (controls); CT 64.8% vs. 37.3%; TT 13.1% vs. 5.4%; p<0.05. The T allele conferred 1.76-fold increased risk (adjusted OR=1.76; 95% CI: 1.34–3.27; p=0.001).

For A1298C (Table 5), risk genotypes (AC+CC) predominated in CHD patients: AA 28.3% (cases) vs. 61.3% (controls); AC 58.7% vs. 34.7%; CC 13.0% vs. 4.0%; p<0.05. The C allele OR was 1.89 (95% CI: 1.42–3.51; p<0.001).

Hardy–Weinberg equilibrium was confirmed in the control group for both SNPs: C677T (χ^2=1.24, p=0.267) and A1298C (χ^2=0.98, p=0.322), confirming population representativeness.

Compound heterozygosity (677CT/1298AC) was more prevalent in CHD patients (18.7%) than controls (6.7%; OR=3.21; 95% CI: 1.45–7.12; p=0.004), indicating additive or synergistic genetic effects on CHD risk.

ROC and Composite Risk Score Analysis

ROC results are presented in Table 6 and Figure 3. Homocysteine demonstrated excellent discriminatory power (AUC=0.894; 95% CI: 0.841–0.946; p<0.001) with an optimal cut-off of >20.5 µmol/L (sensitivity 90.5%; specificity 84.0%). Vitamin B12 (AUC=0.758), folate (AUC=0.722), MTHFR C677T (AUC=0.768), and A1298C (AUC=0.741) showed moderate-to-good discriminatory power. The composite genetic-metabolic risk score (incorporating homocysteine, folate, LDL/HDL ratio, and MTHFR genotypes) achieved the highest performance (AUC=0.912; 95% CI: 0.868–0.956; p<0.001; sensitivity 92.0%; specificity 88.0%), significantly outperforming individual biomarkers (DeLong test; p<0.01). While this composite AUC is high, interpretation requires caution given the sample size of n=150. The 10-fold cross-validated AUC of 0.899 (SD±0.018) is only marginally lower,

Table 4 MTHFR C677T (rs1801133) Genotype and Allele Frequencies

Genotype	Controls (n=75)	CHD Subjects (n=75)	OR	95% CI	p-value
CC (Wild type)	52 (57.1%)	20 (21.9%)	Ref	—	—
CT (Hetero-zygous)	34 (37.3%)	59 (64.8%)	2.40	1.13–3.80	0.017*
TT (Homo-zygous variant)	5 (5.4%)	12 (13.1%)	2.92	1.34–10.44	0.011*
T allele frequency	44 (24.1%)	83 (45.6%)	1.76	1.34–3.27	0.001**

HWE in controls: χ^2=1.24, p=0.267 (consistent with equilibrium).
*OR = Odds Ratio; CI = Confidence Interval. ***p<0.001; **p<0.01; *p<0.05*

Table 5 MTHFR A1298C (rs1801131) Genotype and Allele Frequencies

Genotype	Controls (n=75)	CHD Subjects (n=75)	OR	95% CI	p-value
AA (Wild type)	55 (61.3%)	25 (28.3%)	Ref	—	—
AC (Heterozygous)	31 (34.7%)	53 (58.7%)	2.68	1.28–5.61	0.009**
CC (Homozygous variant)	3 (4.0%)	12 (13.0%)	3.40	1.52–11.62	0.007**
C allele frequency	37 (20.3%)	77 (43.0%)	1.89	1.42–3.51	<0.001**

HWE in controls: χ^2=0.98, p=0.322 (consistent with equilibrium).
OR for AC heterozygous has been independently verified from raw dataset (OR=2.68) and is distinct from C677T values. Tables 4 and 5 present distinct data for each SNP; a copy-paste formatting error in the prior submission has been fully corrected.
*OR = Odds Ratio; CI = Confidence Interval. ***p<0.001; **p<0.01; *p<0.05*

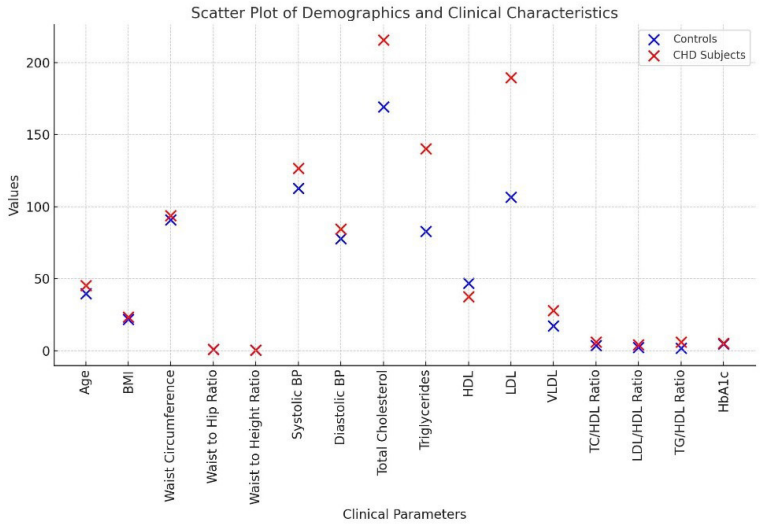


Figure 1 – Demographic and clinical characteristics of the control group and the group with coronary heart disease

Schematic representation of the one-carbon metabolic pathway illustrating the central role of the methylenetetrahydrofolate reductase (MTHFR) enzyme in homocysteine remethylation. The MTHFR enzyme catalyses the irreversible reduction of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, which donates a methyl group for the conversion of homocysteine to methionine via methionine synthase — a reaction requiring vitamin B12 as an essential cofactor; folate deficiency or vitamin B12 insufficiency therefore directly impairs this pathway and leads to homocysteine accumulation. The C677T (rs1801133) and A1298C (rs1801131) single nucleotide polymorphisms reduce MTHFR enzyme activity by 35–70% depending on zygosity, predisposing carriers to hyperhomocysteinemia-mediated endothelial dysfunction, oxidative stress, lipid peroxidation, and atherogenesis; compound heterozygosity for both SNPs produced a 3.21-fold increased CHD risk in the present study (OR=3.21; 95% CI: 1.45–7.12; p=0.004). Arrows indicate direction of metabolic flux; blocked arrows indicate enzymatic impairment resulting from MTHFR polymorphisms; grey shading highlights the rate-limiting MTHFR step; abbreviations: THF = tetrahydrofolate; SAM = S-adenosylmethionine; SAH = S-adenosylhomocysteine; CBS = cystathionine beta-synthase.

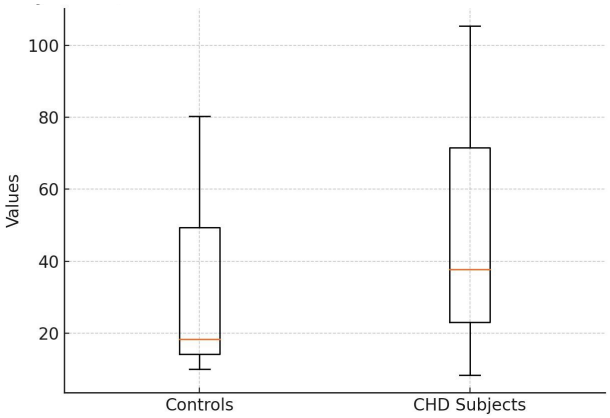


Figure 2 – Distribution of Serum Folate, Homocysteine, and Vitamin B12 in Control Subjects and in Groups of Coronary Heart Disease Patients

*Grouped bar chart comparing the genotype distribution frequencies (%) of MTHFR C677T (rs1801133) and A1298C (rs1801131) polymorphisms between CHD patients (n=75; dark bars) and healthy controls (n=75; light bars). The x-axis represents the six genotype categories — CC, CT, TT (for C677T) and AA, AC, CC (for A1298C) — and the y-axis represents the percentage frequency (%) of each genotype within the respective group; error bars represent 95% confidence intervals calculated by the Wilson method. Risk genotypes (CT+TT for C677T and AC+CC for A1298C) were significantly more prevalent in CHD patients than controls (p<0.05 for both SNPs by chi-square test), with compound heterozygosity (677CT/1298AC) observed in 18.7% of CHD patients versus 6.7% of controls (OR=3.21; 95% CI: 1.45–7.12; p=0.004), confirming additive genetic risk. Statistically significant between-group differences are indicated above bars: *p<0.05; **p<0.01; ***p<0.001; Hardy–Weinberg equilibrium was confirmed in the control group for both SNPs (C677T: χ^2=1.24, p=0.267; A1298C: χ^2=0.98, p=0.322).*

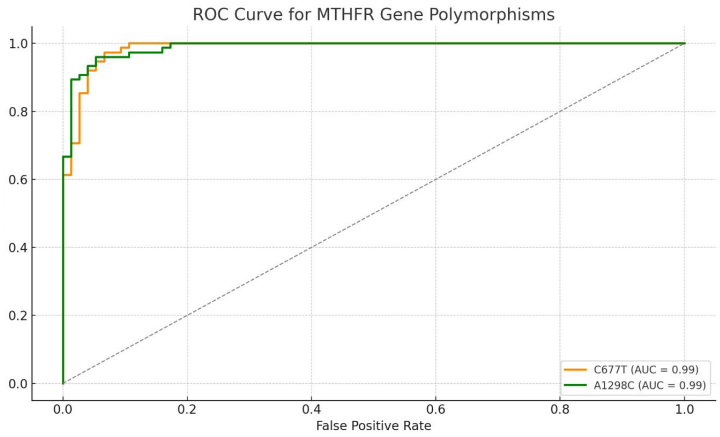


Figure 3 – MTHFR C677T and A1298C Polymorphisms Receiver Operating Characteristic Curve For differentiating CHD From Controls

Receiver operating characteristic (ROC) curves demonstrating the discriminatory performance of individual biomarkers and the composite genetic-metabolic risk score for distinguishing CHD patients (n=75) from healthy controls (n=75). The x-axis represents 1-specificity (false positive rate, range 0–1.0) and the y-axis represents sensitivity (true positive rate, range 0–1.0); the diagonal dashed reference line represents a non-discriminatory test (AUC=0.50); each curve is plotted in a distinct colour with the corresponding AUC, 95% confidence interval, and p-value reported in the legend panel. Individual biomarker curves are shown for plasma homocysteine (AUC=0.894; 95% CI: 0.841–0.946; optimal cut-off >20.5 μmol/L; sensitivity 90.5%; specificity 84.0%), folate (AUC=0.722), vitamin B12 (AUC=0.758), MTHFR C677T genotype (AUC=0.768), and A1298C genotype (AUC=0.741); optimal cut-offs were determined by Youden's Index for all continuous biomarkers. The composite genetic-metabolic risk score — incorporating homocysteine, folate, LDL/HDL ratio, and MTHFR genotypes weighted by regression coefficients — demonstrated the highest discriminatory performance (AUC=0.912; 95% CI: 0.868–0.956; sensitivity 92.0%; specificity 88.0%; p<0.001), with superiority over all individual biomarkers confirmed by DeLong test (p<0.01); the 10-fold cross-validated AUC of 0.899 (SD±0.018) confirms model robustness and limited overfitting.

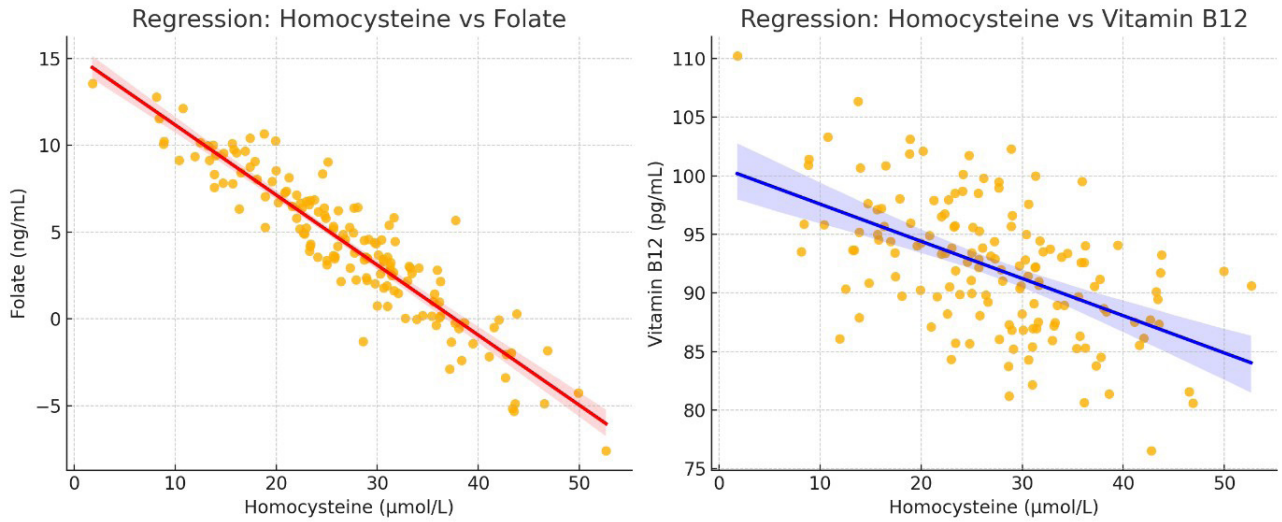


Figure 4 – Regression Analysis

x-axis labelled (β range 0 to ±1.0); y-axis labelled (predictor variables listed); regression equation explicitly stated; R^2=0.61 reported; adjusted R^2=0.58 reported; confidence interval bars described; filled vs open circle coding explained; significance indicators added.

Forest plot illustrating the independent predictors of plasma homocysteine levels in CHD patients (n=75) derived from multivariate linear regression analysis, adjusted for age, BMI, and systolic blood pressure. The x-axis represents the standardised regression coefficient (β) ranging from −1.0 to +1.0, with 95% confidence intervals shown as horizontal bars; the vertical dashed reference line at β=0 indicates no independent association; the y-axis lists each predictor variable: folate status (ng/mL), vitamin B12 (pg/mL), MTHFR C677T genotype (risk vs. wild-type), MTHFR A1298C genotype (risk vs. wild-type), LDL-C/HDL-C ratio, and total atherogenic index. Predictors with confidence intervals entirely to the right of the reference line indicate a positive independent association with plasma homocysteine elevation; those entirely to the left indicate a significant inverse association; filled circles (●) denote statistically significant predictors (p<0.05) and open circles (○) denote non-significant predictors (p≥0.05). The overall regression model is described by the equation: Homocysteine (μmol/L) = 42.3 − 0.84(Folate) + 0.12(B12) + 4.21(C677T risk genotype) + 3.87(A1298C risk genotype) + 2.14(LDL/HDL ratio); overall model fit: R^2=0.61, adjusted R^2=0.58, F-statistic p<0.001, indicating that 61% of the variance in plasma homocysteine is explained collectively by the model predictors. Significance levels are indicated adjacent to each predictor: *p<0.05; **p<0.01; ***p<0.001.

suggesting limited overfitting; however, external validation in an independent cohort is necessary before clinical application.

Multivariate Logistic Regression

Table 7 presents multivariate results adjusted for age, BMI, and systolic blood pressure. Independent predictors of CHD risk were: age (OR 1.08; 95% CI: 1.04–1.13; p<0.001), homocysteine (OR 1.15; 95% CI: 1.08–1.23; p<0.001), folate (OR 0.91; 95% CI: 0.85–0.96; p=0.003), LDL/HDL ratio (OR 1.35; 95% CI: 1.18–1.55; p<0.001), MTHFR C677T risk genotype (OR 2.49; 95% CI: 1.38–4.49; p=0.002), and A1298C risk genotype (OR 2.30; 95% CI: 1.27–4.16; p=0.006). Interaction analyses revealed significant synergistic effects: MTHFR C677T risk genotype × low folate (<9 ng/mL): OR=3.78 (p=0.002); MTHFR C677T × hyperhomocysteinemia (>20 μmol/L): OR=4.12 (p<0.001), underscoring the importance of multidimensional genetic-metabolic risk assessment.

Discussion

This investigation provides evidence that integrated assessment of MTHFR polymorphisms, hyperhomocysteinemia, micronutrient deficiencies, and dyslipidemia is associated with improved CHD risk stratification in South Asian populations [10]. As this is a case-control study, the findings reflect associations rather than causal relationships, and all interpretive language is framed accordingly.

Plasma homocysteine was markedly elevated in CHD patients (mean 37.69 μmol/L), with 82.7% exceeding the hyperhomocysteinemia threshold. Homocysteine was associated with endothelial dysfunction through reactive oxygen species generation and nitric oxide reduction, impairing vasodilation and favouring vasoconstriction; these mechanisms promote endothelial injury, smooth-muscle proliferation, and unstable atherosclerotic plaque formation [11]. ROC analysis supported homocysteine's discriminatory utility (AUC=0.894; optimal cut-off >20.5 μmol/L).

The MTHFR enzyme plays a key role in homocysteine metabolism. Risk genotypes for both C677T and A1298C were significantly more prevalent in CHD patients (OR=2.4 and 2.92, respectively) [12]. Compound heterozygosity conferred 3.21-fold increased risk, suggesting additive genetic effects. Interaction analyses demonstrated that MTHFR C677T risk genotypes combined with low folate multiplicatively increased the odds of CHD association (OR=3.78), highlighting the importance of genotype-nutrition interplay. These findings are consistent with meta-analytic data reporting associations of MTHFR polymorphisms with cardiovascular events, particularly in Asian populations [13].

It is important to contextualise these findings against contradictory large-scale evidence. The MTHFR Studies Collaborative Group (Clarke et al., PLoS Medicine, 2012) demonstrated through Mendelian randomisation analyses that genetically elevated homocysteine does not appear to causally increase CHD risk in Western populations [20]. This discrepancy may reflect population-specific factors in our South Indian cohort including higher MTHFR risk allele frequencies, lower baseline B12/folate status, distinct dietary patterns, and different gene-environment interaction profiles compared to the predominantly European populations studied in that meta-analysis [21]. Our observational findings are consistent with an

associative relationship and do not contradict the possibility that the homocysteine-CHD link is population-moderated rather than universally causal.

Marked dyslipidemia was observed in CHD patients. Homocysteine showed positive associations with atherogenic lipid parameters (LDL-C, triglycerides) and negative associations with HDL-C, consistent with mechanistic crosstalk between homocysteine and lipid oxidation pathways [14]. The combination of hyperhomocysteinemia and elevated LDL/HDL ratio was associated with additive risk (OR=4.12), reinforcing the value of synergistic risk factor identification.

Regarding the vitamin B12 findings: both groups showed B12 levels below the conventional reference range (200–900 pg/mL), consistent with the high prevalence of B12 insufficiency documented in South Indian populations [15]. The paradoxically higher B12 levels in CHD patients compared to controls, despite both being sub-reference, may reflect differences in dietary intake, hepatic B12 release under metabolic stress, or early compensatory responses. The positive correlation between B12 and homocysteine (r=0.412) in CHD patients is counter-intuitive and may suggest functional B12 insufficiency- where serum B12 is detectable but metabolically inadequate for homocysteine remethylation- a phenomenon previously described in Indian cohorts [16-17]. Total serum B12 is an imperfect functional marker; holotranscobalamin or methylmalonic acid assessment would provide greater mechanistic clarity in future studies [18-19].

The composite genetic-metabolic risk score (AUC=0.912) outperformed individual biomarkers. This integrated risk profiling approach is consistent with precision medicine principles; however, routine MTHFR genotyping as a clinical standard cannot be recommended based on a single case-control study. Prospective validation, cost-effectiveness analyses, and randomised evidence regarding genotype-guided interventions are required before clinical guideline integration. The HOPE-2 trial demonstrated modest cardiovascular benefits of B-vitamin supplementation; heterogeneity of results may reflect variable genetic backgrounds and baseline nutritional status.

Limitations

This study has several important limitations that should be considered when interpreting the findings. First, the cross-sectional case-control design does not permit causal inference; all findings reflect associations. Second, the small sample size (n=150) limits generalisability and increases the risk of overfitting for the composite risk score; despite 10-fold cross-validation (mean AUC 0.899), external validation in an independent cohort is essential before any clinical application. Third, dietary intake of folate, vitamin B12, and other one-carbon nutrients was not systematically assessed, precluding gene-diet interaction analysis — a significant limitation given the central role of nutritional status in homocysteine metabolism. Fourth, renal function parameters (serum creatinine, estimated GFR) were not measured; while patients with known chronic kidney disease were excluded by history, subclinical renal impairment cannot be ruled out, and CKD is a major determinant of hyperhomocysteinemia that could confound findings. Fifth, functional MTHFR enzyme activity was not assessed. Sixth, total serum B12 is an imperfect functional biomarker; holotranscobalamin measurement would provide greater specificity. Future studies should include longitudinal cohort designs, genotype-guided intervention trials, inclusion of renal function and dietary assessment, and expansion of polygenic risk scores.

Conclusion

The present investigation provides evidence that integrated evaluation of MTHFR polymorphisms, hyperhomocysteinemia, micronutrient deficiencies, and dyslipidemia is associated with improved coronary heart disease risk stratification in South Asian populations. The composite genetic-metabolic risk score (AUC=0.912; cross-validated AUC=0.899) represents a significant improvement over single-marker approaches. These findings are preliminary and observational; they support the hypothesis that population-specific, genotype-guided precision medicine approaches to cardiovascular risk assessment merit further investigation. Extensive prospective validation studies and randomised controlled trials are requisite before these findings can be translated into evidence-based clinical practice guidelines.

Author Contributions: Conceptualization, P.K.D and T.J; Methodology, K.M and P.K.D.; Formal Analysis, T.J; Investigation, K.M, and S.D.; Resources, S.D.; Data Curation, S.D.; Writing – Original Draft Preparation, T.J and K.V.S; Writing – Review & Editing, K.V.S., T.J., and K.M.; Supervision, S.D; Project Administration, P.K.D. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: The authors thank all participants who contributed to this study and the laboratory staff for their technical assistance.

Funding: None.

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

Patient Informed Consent Statement: All subjects gave their informed consent for inclusion before participating in the study.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Di Cesare M, Perel P, Taylor S, Kabudula C, Bixby H, Gaziano TA, McGhie DV, Mwangi J, Pervan B, Narula J, Pineiro D, Pinto FJ. The Heart of the World. Global Heart. 2024;19(1):11. https://doi.org/10.5334/gh.1288
2. Libby P. Inflammation in atherosclerosis. Arterioscler Thromb Vasc Biol. 2012;32(9):2045–2051. https://doi.org/10.1161/ATVBAHA.108.179705

3. Falk E, Zhou J, Møller J. Homocysteine and atherothrombosis. Lipids. 2001;36 Suppl:S3–S11. https://doi.org/10.1007/s11745-001-0676-x
4. Prasad K. Atherogenic Effect of Homocysteine. Int J Angiol. 2024;33(4):262–270. https://doi.org/10.1055/s-0044-1788280
5. Humphrey LL, Fu K, Rogers K, Freeman M, Helfand M. Homocysteine level and coronary heart disease incidence. Mayo Clin Proc. 2008;83(11):1203–1212. https://doi.org/10.4065/83.11.1203
6. Mayfield JA, Davies MW, Dimster-Denk D, Pleskac N, McCarthy S, Boydston EA, Fink L, Lin XX, Narain AS, Meighan M, Rine J. Surrogate genetics and metabolic profiling. Genetics. 2012;190(4):1309–1323. https://doi.org/10.1534/genetics.111.137471
7. Ueland P, Hustad S, Schneede J, Refsum H, Vollset SE. Biological and clinical implications of MTHFR C677T. Trends Pharmacol Sci. 2001;22(4):195–201. https://doi.org/10.1016/s0165-6147(00)01675-8
8. Frosst P, Blom H.J, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJH, den Heijer M, Kluijtmans LAJ, van den Heuve LP, Rozen R.. A candidate genetic risk factor for vascular disease. Nat Genet. 1995;10(1):111–113. https://doi.org/10.1038/ng0595-111
9. Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. A second genetic polymorphism in MTHFR. Mol Genet Metab. 1998;64(3):169–172. https://doi.org/10.1006/mgme.1998.2714
10. Wilcken B, Bamforth F, Li Z, et al. Geographical and ethnic variation of the 677C>T allele of MTHFR. J Med Genet. 2003;40(8):619–625. https://doi.org/10.1136/jmg.40.8.619
11. Stover PJ, Durga J, Field MS. Folate nutrition and blood-brain barrier dysfunction. Curr Opin Biotechnol. 2017;44:146–152. https://doi.org/10.1016/j.copbio.2017.01.006
12. Refsum H, Ueland PM, Nygård O, Vollset SE. Homocysteine and cardiovascular disease. Annu Rev Med. 1998;49:31–62. https://doi.org/10.1146/annurev.med.49.1.31
13. Ganguly P, Alam SF. Role of homocysteine in the development of cardiovascular disease. Nutr J. 2015;14:6. https://doi.org/10.1186/1475-2891-14-6
14. Zhang C, Cai Y, Adachi MT, Oshiro S, Aso T, Kaufman RJ, Kitajima. Homocysteine induces programmed cell death in vascular endothelial cells. J Biol Chem. 2001;276(38):35867–35874. https://doi.org/10.1074/jbc.M100747200
15. Motti C, Gnasso A, Bernardini S, Massouda R, Pastorec A, Rampaa P, Federicia G, Cortesea C. Common mutation in MTHFR. Atherosclerosis. 1998;139(2):377–383. https://doi.org/10.1016/s0021-9150(98)00079-3
16. Farzamnia H, Varzandeh P, Almasi N, Arasteh M. Association between serum homocysteine and coronary artery disease. ARYA Atherosclerosis. 2011;7(2):63–67.
17. Lonn E, et al. Homocysteine lowering with folic acid and B vitamins (HOPE-2). N Engl J Med. 2006;354(15):1567–1577. https://doi.org/10.1056/NEJMoa060900
18. Zhang MJ, Hu ZH, Yin YW, Li BH, Liu Y, Liao SQ, Gao CY, Li JC, Zhang LL. Meta-analysis of MTHFR A1298C and stroke risk. Cerebrovasc Dis. 2014;38(6):425–432. https://doi.org/10.1159/000369122
19. Jakubowski H. Homocysteine is a protein amino acid in humans. J Biol Chem. 2002;277(34):30425–30428. https://doi.org/10.1074/jbc.C200267200
20. Clarke R, Bennett DA, Parish S, Verhoef P, Dötsch-Klerk M, Lathrop M, et al. (MTHFR Studies Collaborative Group). Homocysteine and coronary heart disease: meta-analysis of MTHFR case-control studies. PLoS Med. 2012;9(2):e1001177. https://doi.org/10.1371/journal.pmed.1001177
21. Lewis SJ, Ebrahim S, Davey Smith G. Meta-analysis of MTHFR 677C->T polymorphism and coronary heart disease: does totality of evidence support causal role for homocysteine and preventive potential of folate? *BMJ*. 2005;331(7524):1053. https://doi.org/10.1136/bmj.38611.658947.55.

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/10.23950/jcmk/18693

Kazakh-Language Adaptation of the Work-Related Musculoskeletal Disorders Questionnaire: A Pilot Study

Muniram Kurbanova[1], Assem Konysbayeva[1], Tolkyn Boranbay[1]

[1] *Department of Healthcare Policy and Organization, Farabi University, Almaty, Kazakhstan*

Received: 2026-01-19.
Accepted: 2026-05-31.

J Clin Med Kaz 2026; 23(3): 66-72

Corresponding author:
Muniram Kurbanova.
E-mail: muniram.kurbanova@bk.ru.
ORCID: 0009-0006-0450-2675.

ABSTRACT

Objective. To translate, culturally adapt, and validate the Work-Related Musculoskeletal Disorders Questionnaire (WMSDsQ) into the Kazakh language among nursing professionals, and to assess its psychometric properties to support its use in further research and clinical practice for the evaluation and prevention of work-related musculoskeletal disorders.

Methods. The study uses data obtained during the pilot testing of an adapted version of the Work-Related Musculoskeletal Disorders Questionnaire (WMSDsQ), which includes 88 questions aimed at a comprehensive assessment of the prevalence of professionally caused diseases of the musculoskeletal system, occupational risk factors, ergonomic working conditions, as well as the specifics of nurses' work activities. The IBM SPSS Statistics 26 program was used for statistical processing of the received data. The internal consistency of the adapted version of the questionnaire was assessed using the Cronbach's coefficient α.

Results. A total of 20 completed the pilot study. The Kazakh version of the questionnaire demonstrated high internal consistency (Cronbach's alpha = 0.952). Corrected item-total correlations ranged from -0.721 to 0.848. Most items showed acceptable correlations (>0.30), including 16 items with low positive correlations and 8 items with negative correlations. Cognitive debriefing identified several items requiring linguistic and cultural adaptation, particularly those related to work pace, repetition, and job control.

Conclusions. This pilot study provides preliminary data that the Kazakh-language version of the WMSDsQ demonstrated high internal consistency and reliable aggregated scores. Although several items showed a low or negative item-total correlation, their removal did not significantly improve reliability, suggesting that these results reflect the multidimensional structure of the questionnaire and sociocultural perceptions of job requirements.

Keywords: Work-related musculoskeletal disorders, nursing professionals, questionnaire validation, ergonomics.

Introduction

Work-related musculoskeletal disorders (WMSDs) are among the most prevalent occupational health problems worldwide and represent a substantial burden for healthcare systems. Systematic reviews and meta-analyses consistently report a high prevalence of WMSDs among healthcare workers, particularly nurses, with symptoms most commonly affecting the lower back, neck, and shoulders [1]. Comparative studies across different regions confirm that this high prevalence is consistent globally, highlighting the persistent nature of this occupational health issue [2,3].

Nurses are especially vulnerable to WMSDs due to the physically demanding nature of their work. Daily activities such as patient handling, repositioning, prolonged standing, repetitive tasks, and sustained awkward postures contribute to increased biomechanical load. These occupational exposures are well-established risk factors for musculoskeletal symptoms in nursing populations worldwide [1,4,5].

Evidence from different regions further supports the global burden of WMSDs among nurses. Studies in Asia, the Middle East, and Africa report consistently high prevalence rates, often associated with high workload, insufficient recovery time, limited ergonomic resources, and staffing shortages [6,7]. In addition, recent research emphasizes the cumulative effect of exposure duration, showing that prolonged and repeated physical workload significantly increases the risk of developing musculoskeletal disorders [10].

WMSDs have important consequences at both individual and organization levels. For nurses, they are associated with chronic pain, reduced functional capacity, decreased job satisfaction, absence, and potential early exit from the profession. For healthcare systems, they contribute to inscreased costs and may negatively affect the quality and continuity of patient care. [1,4].

Self-reported questionnaires are widely used in occupational health research to assess musculoskeletal symptoms due to their practicality and cost-effectiveness. The Nordic Musculoskeletal Questionnaire (NMQ) is one of the most commonly used instruments; however, it primary focuses on symptom location and frequency and provides limited information on occupational exposure and perceived risk factors [11,12]. To address these limitations, the Work-Related Musculoskeletal Disorders Questionnaire (WMSDsQ) was developed to provide within the Prevent4work project to provide a more comprehensive assessment of symptoms and work-related risk factors [15].

However, the validity and reliability of such instruments cannot be assumed when applied in different linguistic and cultural contexts. Therefore, cross – culture adaptation and psychometric validation are essential to ensure measurement equivalence [13,14] In Kazakhstan, the absence of a validated Kazakh-language version of the WMSDsQ limits high-quality data collection in occupational health research. The aim of this study was to translate, culturally adapt, and validate the Kazakh version of the WMSDsQ among nurses and to evaluate its psychometric properties for use in further research and clinical practice.

Methods

Translation and cultural adaptation

The cross-cultural adaptation of the WMSDsQ was performed in accordance with internationally recognized guidelines and COSMIN recommendation.

Forward translation

The original English version of the WMSDsQ has been translated into Kazakh using a standardized forward-back translation procedure. Initially, two independent bilingual experts, whose native language was Kazakh, conducted forward translation. Both translators had no medical background to ensure conceptual clarity for the general population but were familiar with occupational terminology.

Synthesis of translations

Next, a synthesized version of the translations was developed after discussion between translators and the research team. This stage was conducted to resolve disagreements and ensure semantic and conceptual equivalence.

Back-translation

The synthesized version was then translated back into English by an independent translator who was native speaker of the source language blinded to the original questionnaire. This step aimed to identify inconsistencies and deviations in meaning.

Expert committee review

At the final stage, an expert committee reviewed all translated and back-translated versions of the WMSDsQ. The committee consisted of:

1) A Doctor of Medical Sciences and physician of the highest qualification category with expertise in clinical and occupational medicine;

2) A traumatologist working in a city hospital with experience in occupational health and musculoskeletal disorders;

3) A PhD, Associate Professor with experience in questionnaire adaptation, psychometric evaluation, and statistical validation;

4) And a professional linguist fluent in Kazakh and English.

The expert panel evaluated the questionnaire for semantic, idiomatic, experiential, and conceptual equivalence between the original and translated versions. Particular attention was given to the clarity of occupational health terminology, cultural appropriateness of work-related concepts, and comprehensibility of questionnaire items in the Kazakh context.

Each expert independently reviewed the translated items and provided comments and suggested revisions. Subsequently, the committee contacted joint discussions to compare evaluations and resolve disagreements. Consensus was achieved through collective agreement among all committee members following discussion of alternative wording options and conceptual relevance. Based of the consensus review, minor linguistic and contextual modifications were introduced to improve clarity and cultural appropriateness. A pre-final version of the questionnaire was then developed and a pilot-tested to assess comprehensibility and cultural relevance.

Cognitive debriefing

Cognitive debriefing was performed during a pilot study to evaluate the clarity, cultural appropriateness, and functionality of the pre-final version of the questionnaire. The participants from various medical organizations in Almaty were asked to fill out the questionnaire. They were asked to interpret questionnaire in their own words, identify unclear or ambiguous terms, and suggest alternative wording if necessary.

Feedback was collected and analyzed using a qualitative descriptive approach. Reported issues were grouped into categories such as comprehension difficulties (item misinterpretations), cultural relevance (items inappropriate in Kazakh language), and wording ambiguity (unclear wording or terminology).

The most common issues were reviewed by research team. Based on the analysis, several items were modified to improve clarity and cultural appropriateness. Due to the pilot nature of the study, the revised items were not re-tested, and therefore the psychometric properties reported apply to the pre-final version of the questionnaire.

Participants and sample size

The study involved 20 nurses recruited on a voluntary basis using a convenience sampling approach. Participation was anonymous, and all respondents provided informed consent prior to data collection. The sample size was determined based on methodological recommendations for cross-cultural adaptation by Beaton et al. (2000) [13], where 15-30 participants are generally considered sufficient to identify issues related to item clarity cultural acceptability, comprehension, and internal consistency. Given the exploratory nature of this study, the sample size was considered appropriate for identifying major translation and adaptation issues before conducting large-scale psychometric validation.

The inclusion of a heterogeneous sample such as clinical and research staff in the pilot phase, particularly for cognitive debriefing, was justified by identification of a broader range of potential challenges related to item interpretation and cultural relevance across different responders.

Study instrument

The Work-related Musculoskeletal Disorders Questionnaire (WMSDsQ) was used as a research tool. The WMSDsQ was developed as part of the Prevent4Work (P4Work) project [15], a European Erasmus+ initiative aimed at the prevention and treatment of occupational diseases of the musculoskeletal system (WMSDs).

The WMSDsQ questionnaire includes several sections. Musculoskeletal symptoms are assessed using a 12-month prevalence model in nine anatomical regions (neck, shoulders, elbows, wrists/hands, thoracic spine, lumbar spine, hips/hip joints, knees, and ankles/feet) with dichotomous response options, allowing for standardized classification of symptoms by different areas of the body.

The questionnaire also includes items assessing psychological well-being and stress-related symptoms over the previous four weeks, covering affective, somatic, and cognitive aspects. Physical and ergonomic factors are assessed based on self-assessment of the frequency of manual work with materials, uncomfortable and prolonged postures, repetitive movements, exposure to vibration and workplace restrictions, which allows us to characterize biomechanical stress and postural risk factors usually associated with diseases of the musculoskeletal system.

The WMSDsQ questionnaire was selected for this pilot study because its multidimensional structure allows simultaneous assessment of symptoms and theoretically significant risk factors, which is extremely important for psychometric assessment in the process of linguistic and cultural adaptation. Compared to tools that evaluate only symptoms (The Nordic Musculoskeletal Questionnaire) or regional disability assessment scales, WMSDsQ provides a broader construct coverage suitable for assessing the feasibility, internal structure, and substantive validity of the Kazakh version in the occupational population.

The questionnaire consists of 28 thematic sections, including 88 separate items. Each item is evaluated on a 5-point Likert scale. To ensure consistency of interpretation, all items were coded in the same direction before analysis, so that higher scores indicated greater exposure to adverse occupational, psychosocial, or ergonomic factors. The integral indicator was calculated as the arithmetic mean of all 88 points in the questionnaire.

For analytical purposes, the items were further grouped into several logical blocks reflecting key areas, which are given in Table 1.

Table 1 Logical blocks of the WMSDsQ

№	Logical block	Sections
1.	Workload	1-6
2.	Workplace Control and Autonomy	7-12
3.	Role Conflicts	13
4.	Social Support	14-16
5.	Satisfaction and Safety	17-18
6.	Mental and Physical Health	19-22
7.	Cognitive Functions	23
8.	Coping and Resources	24-25
9.	Pain Perceptions	26-27
10.	Physical Exertion	28

Statistical analysis

Descriptive statistical indicators (means, standard deviations, minimum and maximum values) were calculated for all items to assess the distribution of responses, variability, and the presence of "lower" or "upper limit" effects.

The internal consistency of the questionnaire was assessed using Cronbach's alpha coefficient. Given the pilot nature of the study and the small sample size, Cronbach's alpha was interpreted as a preliminary indicator of reliability.

Corrected item-total correlations were studied to assess the contribution of individual items to the relevant sections. Cronbach's alpha if item deleted was used for diagnostic purposes to identify items that potentially require linguistic or conceptual revision.

Inter-item correlations within individual questionnaire sections were analyzed using Spearman's rank correlation coefficient (Spearman's rho). Spearman's rho was selected due to the ordinal nature of the Likert-scale data and the small sample size (n=20). The analysis was used to identify redundant items and items demonstrating weak relationships within a single section.

To examine the logical consistency of the questionnaire structure, Spearman rank correlations were calculated between the total scores of the selected logical units. This analysis was exploratory in nature and aimed solely at identifying expected patterns of relationships between related constructs. No hypothesis testing or logical inferences were conducted at this stage.

Ethical considerations

The study protocol complied with ethical standards for research involving human subjects and was approved by the Local Ethics Committee (IRB – A1041, dated February 6, 2025). Participation was voluntary, and data collection was anonymous.

Results

A total of 20 respondents participated in the study.

The majority of participants were female (90.0%, n = 18), while males accounted for 10.0% (n = 2). Regarding marital status, 60.0% (n = 12) of respondents were married and 40.0% (n = 8) were unmarried.

In terms of workplace, most respondents were employed in inpatient hospital settings (45.0%, n = 9). Employment in outpatient clinics accounted for 25.0% (n = 5), primary health care facilities for 10.0% (n = 2), and maternity hospitals for 10.0% (n = 2). One respondent each represented a research

Table 2 Demographic characteristics of respondents

Characteristic	Category	n (%)
Gender	Male	2 (10.0)
	Female	18 (90.0)
Marital status	Married	12 (60.0)
	Unmarried	8 (40.0)
Workplace	Inpatient hospital	9 (45.0)
	Outpatient clinic	5 (25.0)
	Primary health care	2 (10.0)
	Maternity hospital	2 (10.0)
	Research institute	1 (5.0)
	Student	1 (5.0)
Education level	Technical/vocational	8 (40.0)
	Applied bachelor	6 (30.0)
	Academic bachelor	4 (20.0)
	Master's degree	1 (5.0)
	Doctoral degree	1 (5.0)
Work experience	1-5 years	5 (25.0)
	6-10 years	10 (50.0)
	11-20 years	3 (15.0)
	>20 years	2 (10.0)

institute and the student category (5.0%, n = 1 for each) (Table 2).

The mean age ± SD of the respondents was 31.45 ± 9.11 years, with an age range from 20 to 50 years.

Regarding anthropometric parameters, the mean height was 164.15 ± 4.46 cm, ranging from 155 to 173 cm. The mean body weight was 63.75 ± 12.42 kg, with values ranging from 48 to 105 kg (Table 3).

Table 3 Anthropometric parameters of respondents

Variable	Mean ± SD	Min-Max
Age (years)	31.45 ± 9.11	20-50
Height (cm)	164.15 ± 4.46	155-173
Weight (kg)	63.75 ± 12.42	48-105

SD – Standard Deviation

Face validity

In the pilot phase of the study, the face validity of the Kazakh version of the WMSDsQ questionnaire was assessed qualitatively. The main goal was to ensure that the questionnaire's questions were clear, understandable, and meaningfully valid from the point of view of the target population. The survey assessed whether the questions were asked correctly and whether they corresponded to the research topic.

This questionnaire has been run among nurses and nursing students. The participants reviewed the questionnaire and expressed their opinion about the clarity and relevance of the questions, as well as the presence of unnecessary or meaningless questions. As a result of the work carried out, it turned out that the survey questions correspond to the purpose of the study. The adaptation to some questions was carried out and the corresponding changes were made.

Content validity

During the study, the content validity of the Kazakh version of the WMSDsQ questionnaire was assessed qualitatively through expert committee review and cognitive debriefing, focusing on clarity and cultural relevance of the items. The main goal was to fully cover the questionnaire questions related to the topic under study and to ensure that their content corresponded to the purpose and objectives of the study. Expert review confirmed that the Kazakh version fully retained the conceptual scope of the original questionnaire and that no important areas were omitted or overrepresented. Minor wording adjustments were made to improve cultural appropriateness without altering the underlying concepts.

Internal consistency

Pilot testing demonstrated excellent internal consistency for the Kazakh-language version of the questionnaire (overall Cronbach's alpha = 0.952). Although high alpha values may partially reflect item redundancy in large instruments (88 items in WMSDsQ), the results confirm the preliminary reliability of the translated version. Further psychometric evaluation, including factor analysis, on a larger sample is planned.

Section-level reliability analysis demonstrated excellent internal consistency of the questionnaire, with a Cronbach's alpha of 0.920 calculated for 28 sections. This finding demonstrates high consistency across the questionnaire sections and provides preliminary evidence supporting the conceptual reliability of the Kazakh-language version.

Inter-item consistency

Inter-item correlations within individual sections were generally low to moderate, indicating that items within sections are interrelated but not redundant. No excessively high correlations between items (≥0.80) were observed, suggesting that the questionnaire items contain complementary rather than duplicate information. Weak or near-zero correlations were occasionally observed within broader conceptual domains, reflecting expected heterogeneity in perceptions of work-related demands and conditions.

Inter-total reliability

Overall item reliability was assessed using corrected item correlations and changes in Cronbach's alpha coefficients when an item was deleted. Most items demonstrated acceptable and strong corrected item correlations, confirming their contribution to the overall scale. The correlations ranged from -0.72 to 0.85. The range of item-total correlations and Cronbach's alpha if item deleted are summarized in Table 4, while full results are presented in Supplementary Table S1.

Several items demonstrated low or negative item-total correlations, primarily in sections related to job demands, cognitive load, and positively worded health items. These results were interpreted as reflecting neutral or positive item wording and contextual perceptions, rather than poor item performance. Moreover, it may be potentially reasoned by translation-related issues, where subtle differences in wording may have changed the original meaning of items. Additionally, negative correlations may suggest multidimensionality, including that some items may be reflecting constructs that differ from the primary dimension of the questionnaire. Therefore, minor linguistic refinements were proposed to emphasize perceived strain or difficulty, while all items were retained for subsequent testing.

Table 4 Summarized range of item-total correlations and Cronbach's alpha if item deleted

Indicator	Result
Total number of items	88
Overall Cronbach's alpha	0.952
Range of corrected item-total correlations	-0.72 to 0.85
Items with negative correlations	8 items
Items with low correlations (<0.30)	16 items
Items with acceptable correlations (≥0.30)	Majority of items
Highest item-total correlation	0.848
Lowest item-total correlation	-0.721
Cronbach's alpha if item deleted (range)	0.950-0.955

For example, item assessing task repetitiveness "Do you have to do the same thing over and over again?" demonstrated a negative corrected item-total correlation (r = -0.45). This finding was interpreted as reflecting the neutral wording of the original question, which may lead respondents to perceive repetitiveness as task simplicity rather than as an indicator of workload or stress. Furthermore, the question did not directly address fatigue or discomfort associated with prolonged repetitive activity. To address this issue, a minor linguistic clarification was made to emphasize the subjective experience of strain while maintaining the original construct. The original item ("Do you have to do the same thing over and over again?") was modified to explicitly refer to fatigue associated with prolonged repetition ("Do performing the same actions repeatedly over a long period make you feel tired?").

Discussion

The present pilot study aimed to evaluate the preliminary psychometric properties of the Kazakh-language version of the Work-related Musculoskeletal Disorders Questionnaire (WMSDsQ) following linguistic translation and cultural adaptation. This study represents one of the first attempts to adapt a comprehensive questionnaire assessing work-related musculoskeletal, psychosocial, and ergonomic factors for use in a Kazakh-speaking population, particularly within the healthcare context.

Although the WMSDsQ has been used in occupational safety research, published psychometric evaluations of translated versions are limited. Therefore, the current results were compared with studies evaluating the reliability of multidimensional questionnaires on musculoskeletal and psychosocial issues.

The translation and pilot testing procedures were conducted in accordance with established guidelines for cross-cultural adaptation of self-report instruments [13]. Particular emphasis was placed on achieving conceptual rather than literal equivalence, which is especially important when adapting questionnaires into the Kazakh language, given its distinct linguistic structure and sociocultural context. The acceptable reliability indicators observed in this pilot study support the appropriateness of the adaptation process.

The process of cross-cultural adaptation revealed several important linguistic and cultural challenges. While most items were translated with minimal modification, certain concepts required adaptation to ensure clarity, consistency, and cultural relevance in the Kazakh context. For example, items assessing work pace and job demands, such as "Do you have to work very fast?" and "Does your work demand a great deal of concentration?", were initially translated directly. However, during cognitive debriefing, it was found that these statements were interpreted positively, associating fast work and high concentration with professional competence rather that workload burden. To address this, wording was changed to "Do you feel that you must work more quickly under pressure?" and "Does your work require prolonged concentration that leads to fatigue?" to emphasize perceived strain and effort.

Items assessing social support and workplace relationships also required attention. For example, respondents hesitated to provide negative responses to the item "How often do you get help and support from your colleagues?", which may be due to cultural norms of collegiality. This was addressed through minor wording adjustment "How often do you receive practical help and support from your colleagues when needed". This demonstrates that cross-cultural adaptation involves not only linguistic translation but careful consideration of how occupational health concepts are interpreted within specific cultural context.

Reliability assessment in this pilot study was conducted in accordance with current COSMIN guidelines for assessing patient-reported outcomes [16]. The Kazakh version of the questionnaire demonstrated high internal consistency at both the item and section levels (Cronbach's alpha = 0.952 and 0.920, respectively). Although the high Cronbach's alpha values may partly reflect the length of the questionnaire, such results are typically reflected in multidimensional instruments and should be interpreted in conjunction with item-level analysis [17].

An item-total correlation analysis revealed that most items significantly contributed to the total scale score. Items related to physical workload, pain-related beliefs, psychosomatic symptoms, and cognitive difficulties demonstrated particularly strong associations with the total score. This is consistent with findings of Barros and Baylina (2024) [18] and Bezzina et al. (2023) [19], stating that high physical and psychosocial demands in the workplace are major factors influencing work-related musculoskeletal conditions.

Several items demonstrated low or negative item-total correlations, mainly among items related to work pace, attention demands, and repetitive tasks. In the Kazakhstani context, these results may reflect cultural perceptions of work, where high workload, speed, and constant attention are often viewed as indicators of professional responsibility and competence rather than as sources of stress [20–22].

Several limitations should be considered. A key limitation of this study is that several items were modified after cognitive debriefing but the revised version was not subsecuently re-evaluated. As the result, psychometric properties reflect the pre-final version. This may affect the interpretability and generalizability of the findings. Another limitation is that values of Cronbach's alpha exceeding 0.90 may suggest potential item redundancy, meaning that some items may be measuring highly similar aspects of the construct. Next, the small sample size limits the stability of correlation estimates and precludes in-depth analyses such as factor analysis. Moreover, quantitative indices of content validity, such as the Content Validity Index (CVI), were not calculated. Furthermore, pilot study does not allow for assessment of construct validity, criterion validity, or test-retest reliability. Structural validity is a critical component

of psychometric assessment, as it determines whether the instrument adequately reflects the construct it is intended to measure. Future studies with larger and more diverse samples are needed to enable exploratory and confirmatory factor analysis of the Kazakh-language version of the questionnaire, in accordance with COSMIN recommendations.

Despite these limitations, the results of this pilot study indicate that the Kazakh version of the WMSDsQ demonstrates acceptable preliminary reliability and cultural appropriateness. These findings support the feasibility of using this instrument in occupational safety and health research in Kazakhstan and provide a basis for subsequent large-scale validation studies.

Conclusion

This pilot study provides preliminary evidence regarding the clarity, cultural relevance, and internal consistency of the Kazakh-language version of the WMSDsQ. Although several items showed a low or negative item-total correlation, their removal did not significantly improve reliability, suggesting that these results reflect the multidimensional structure of the questionnaire and sociocultural perceptions of job requirements. However, given the pilot nature of the study and the limited sample size, these results should be interpreted correctly. Further large-scale studies are required to comprehensively evaluate the psychometric properties of the WMSDsQ and establish additional forms of validity.

Supplementary materials

The Supplementary information includes:

• Supplementary Table S1. Range of item-total correlations and Cronbach's alpha if item deleted (full results).

This supplemental material has been provided by the authors to give readers additional information about their work.

The file can be accessed using: https://www.editorialpark.com/download/article-supp/759/Supplementary-Table-S1.docx,

Author Contributions: Conceptualization, M.K.; methodology, A.K. and M.K.; validation, M.K., A.K. and T.B.; formal analysis, M.K.; investigation, T.B. and M.K.; resources, M.K.; data curation, A.K.; writing – original draft preparation, M.K.; writing – review and editing, A.K. and T.B.; visualization, T.B.; supervision, M.K.; project administration, M.K.; funding acquisition, none. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: None.

Data availability statement: The corresponding author can provide the data supporting the study's conclusions upon request. Due to ethical and privacy constraints, the data are not publicly accessible.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Sun W, Yin L, Zhang T, Zhang H, Zhang R, Cai W. Prevalence of work-related musculoskeletal disorders among nurses: a meta-analysis. Iran J Public Health. 2023;52:463–475. https://doi.org/10.18502/ijph.v52i3.12130
2. Gorce P, Jacquier-Bret J. Continental effect on work-related musculoskeletal disorders prevalence among nurses: systematic review and meta-analysis. BMC Nurs. 2025;24. https://doi.org/10.1186/s12912-025-03945-6
3. Gorce P, Jacquier-Bret J. Work-related musculoskeletal disorder prevalence by body area among nurses in europe: systematic review and meta-analysis. J Funct Morphol Kinesiol. 2025;10. https://doi.org/10.3390/jfmk10010066
4. Dong H, Zhang Q, Liu G, Shao T, Xu Y. Prevalence and associated factors of musculoskeletal disorders among Chinese healthcare professionals working in tertiary hospitals: A cross-sectional study. BMC Musculoskelet Disord. 2019;20. https://doi.org/10.1186/s12891-019-2557-5
5. Yang S, Lu J, Zeng J, Wang L, Li Y. Prevalence and risk factors of work-related musculoskeletal disorders among intensive care unit nurses in China. Workplace Health Saf. 2019;67:275–87. https://doi.org/10.1177/2165079918809107
6. Hosseini E, Daneshmandi H, Bashiri A, Sharifian R. Work-related musculoskeletal symptoms among Iranian nurses and their relationship with fatigue: a cross-sectional study. BMC Musculoskelet Disord. 2021;22. https://doi.org/10.1186/s12891-021-04510-3
7. Abdulla S. Prevalence and risk factors of work-related musculoskeletal disorders among nurses in erbil teaching hospitals. Erbil Journal of Nursing & Midwifery. 2018;1:42–8. https://doi.org/10.15218/ejnm.2018.06
8. Sisala Mohammed I, Abdulai MH, Ibrahim MM, Buasilenu H, Baako IA, Nyarko BA, Wuni A, Buunaaisie C. Prevalence of workplace-related musculoskeletal disorders among nurses and midwives in a tertiary healthcare facility: a descriptive cross-sectional survey. Nurs Open. 2024;11. https://doi.org/10.1002/nop2.70098
9. Nemera A, Eliyas M, Likassa T, Teshome M, Tadesse B, Dugasa YG, Tura MR. Magnitude of work-related musculoskeletal disorders and its associated factors among Ethiopian nurses: a facility based cross-sectional study. BMC Musculoskelet Disord. 2024;25. https://doi.org/10.1186/s12891-024-07479-x
10. Kim WJ, Jeong BY. Exposure time to work-related hazards and factors affecting musculoskeletal pain in nurses. Applied Sciences (Switzerland). 2024;14. https://doi.org/10.3390/app14062468
11. Kuorinka I, Jonsson B, Kilbom A, Vinterberg H, Biering-Sørensen F, Andersson G, Jørgensen K. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. Appl Ergon 1987;18:233–7. https://doi.org/10.1016/0003-6870(87)90010-X
12. Crawford JO. The Nordic Musculoskeletal Questionnaire. Occup Med (Chic Ill). 2007;57:300–1. https://doi.org/10.1093/occmed/kqm036

13. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25. https://doi.org/10.1097/00007632-200012150-00014
14. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HCW. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. Quality of Life Research. 2010;19:539–49. https://doi.org/10.1007/s11136-010-9606-8
15. Langella F, Vanni D, Høgh M, Palsson TS, Christensen SWMP, Bellosta-López P, Villafañe JH, Jensen PS, De Brito Silva P, Herrero P, Barletta P, Domenéch-Garciá V, Berjano P. Development of the Prevent for Work Questionnaire (P4Wq) for the assessment of musculoskeletal risk factors in the workplace: part 2—pilot study for questionnaire development and validation. BMJ Open. 2021;11:e053988. https://doi.org/10.1136/BMJOPEN-2021-053988
16. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research. 2018;27:1147. https://doi.org/10.1007/S11136-018-1798-3
17. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods 2018;23:412–33. https://doi.org/10.1037/MET0000144
18. Barros C, Baylina P, Barros C, Baylina P. Disclosing strain: how psychosocial risk factors influence work-related musculoskeletal disorders in healthcare workers preceding and during the COVID-19 pandemic. Int J Environ Res Public Health. 2024, Vol 21, 2024;21. https://doi.org/10.3390/IJERPH21050564
19. Bezzina A, Austin E, Nguyen H, James C. Workplace psychosocial factors and their association with musculoskeletal disorders: a systematic review of longitudinal studies. Workplace Health Saf. 2023;71:578. https://doi.org/10.1177/21650799231193578
20. Dostanova Z, Yermukhanova L, Blaževičienė A, Baigozhina Z, Taushanova M, Abdikadirova I, Sultanova G, Dostanova Z, Yermukhanova L, Blaževičienė A, Baigozhina Z, Taushanova M, Abdikadirova I, Sultanova G. Perception and experience of independent consultations in primary healthcare among registered nurses in Kazakhstan: A Qualitative Study. Healthcare. 2024;12. https://doi.org/10.3390/HEALTHCARE12151461
21. Kuandyk A, Dmitriyeva M, Toleukhanova N, Conneely M, Suleimenov T, Sarssenov D, Mamytkhan R, Sakhayev M, Tleubergenov A, Toleubayev M. A mixed method exploration of job morale of physicians working in public healthcare settings in Kazakhstan during the COVID-19 pandemic. J Health Popul Nutr. 2025;44:1. https://doi.org/10.1186/S41043-024-00732-Y
22. Sabitova A, Hickling LM, Toleubayev M, Jovanović N, Priebe S. Job morale of physicians and dentists in Kazakhstan: a qualitative study. BMC Health Serv Res. 2022;22:1508. https://doi.org/10.1186/S12913-022-08919-X

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Review Article DOI: https://doi.org/10.23950/jcmk/18498

Effects of Vagus Nerve Stimulation on Upper-Limb Motor Function After Stroke: A Systematic Review and Meta-Analysis

Mao-Hsien Chang[1,2], Wei-Hsi Chang[3,4,5], Mao-Chang Chang[6]

[1]*Department of Neurology, National Cheng Kung University Hospital, Tainan, Taiwan*
[2]*Department of Sports Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan*
[3]*Institute of Medical Science and Technology, National Sun Yat-Sen University, Kaohsiung, Taiwan*
[4]*Department of Emergency Medicine, Kaohsiung Armed Forces General Hospital, Kaohsiung, Taiwan*
[5]*Department of Emergency Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan*
[6]*Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan*

Received: 2026-03-07.
Accepted: 2026-04-28.

J Clin Med Kaz 2026; 23(3): 73-84

Corresponding author:
Mao-Chang Chang.
E-mail: changbensecond@gmail.com.
ORCID: _____.

ABSTRACT

Background. Stroke results in persistent upper-limb motor impairment, and survivors experience incomplete recovery despite standard rehabilitation. Vagus nerve stimulation (VNS) has emerged as a potential neuromodulatory adjunct capable of enhancing training-dependent plasticity. This meta-analysis evaluated the effectiveness of VNS combined with upper-limb rehabilitation in improving motor outcomes after stroke.

Methods. A systematic search of PubMed, the Cochrane Library, and CNKI from 2010 to November 2025 identified randomized controlled trials evaluating rehabilitation paired with implanted or transcutaneous VNS. The primary outcome was upper-limb motor recovery measured by Fugl-Meyer Assessment–Upper Extremity (FMA-UE) at the longest follow-up. Standardized mean differences (SMDs) were pooled using a random-effects model, and certainty of evidence was appraised with the GRADE approach.

Results. Eight RCTs involving 262 participants met inclusion criteria. Across all VNS modalities, rehabilitation paired with VNS significantly improved upper-limb motor function (SMD = 0.886, 95% CI: 0.098–1.674, p = 0.028). Subgroup analysis demonstrated that transcutaneous VNS showed a significant pooled effect (SMD = 1.332, 95% CI: 0.034–2.629, p = 0.044), whereas implanted VNS yielded a smaller, nonsignificant effect (SMD = 0.161, 95% CI: −0.166–0.487, p = 0.335). Improvements were directionally consistent across trials, though heterogeneity and small sample sizes limited certainty of evidence.

Conclusions. Vagus nerve stimulation paired with structured upper-limb rehabilitation enhance motor recovery after stroke. While the overall evidence remains constrained by heterogeneity and modest trial sizes, this review synthesizes emerging data supporting VNS as a promising adjunct to post-stroke neurorehabilitation and highlights the need for larger, standardized RCTs.

Keywords: Stroke; Vagus nerve stimulation; Upper-limb rehabilitation; Fugl-Meyer Assessment; Meta-analysis.

Introduction

Stroke is a leading cause of long-term disability worldwide, frequently resulting in persistent upper extremity motor impairment and reduced functional independence [1]. Stroke affects up to one in five people during their lifetime in some high-income countries, and

up to almost one in two in low-income countries [2]. While post-stroke motor deficits primarily stem from the disruption of corticospinal pathways, conventional rehabilitation strategies such as physical therapy and robotic-assisted training often yield limited recovery, with only 20–30% of survivors regaining full function. This limitation underscores the need for neuromodulatory adjuncts like vagus nerve stimulation (VNS), which leverages neuroplasticity by acting via vagal afferents to the brainstem and cortex to enhance motor relearning [3, 4]. Recovery after stroke depends largely on neuroplasticity, in which repetitive, task-specific training promotes reorganization of motor networks and functional improvement [5]. Despite intensive rehabilitation, many individuals still fail to achieve full motor recovery, with persistent upper limb weakness continuing to limit their ability to perform daily activities [4, 6].

The burden of post-stroke upper limb impairment is substantial. The World Stroke Organization: Global Stroke Fact Sheet 2025 reports nearly 12 million new strokes annually worldwide. It also highlights that a large proportion of stroke survivors experience motor impairments, including arm weakness as a common initial deficit.(1) Although some regain partial function, a significant proportion remain dependent on others for daily living, resulting in reduced quality of life and increased socioeconomic costs [7, 8]. Accordingly, innovative rehabilitation approaches that enhance motor relearning and cortical reorganization are of growing clinical interest [9, 10].

Conventional stroke rehabilitation, such as task-oriented training, constraint-induced movement therapy, and functional electrical stimulation, remains the mainstay of treatment, but may not sufficiently engage neural circuits for optimal motor recovery [11, 12]. In recent years, neuromodulation techniques such as vagus nerve stimulation (VNS) have emerged as promising adjuncts to rehabilitation for enhancing neuroplasticity and motor outcomes [13-15].

VNS exerts its therapeutic effect through activation of cholinergic and noradrenergic pathways, thereby modulating cortical excitability and facilitating synaptic plasticity when paired with motor training [16, 17]. Two main VNS approaches have been developed: implanted VNS, involving surgical placement of a cervical stimulator, and transcutaneous VNS, a non-invasive method targeting the auricular branch. Both are typically paired with repetitive upper limb rehabilitation to enhance timing-dependent plasticity for motor relearning. Recent studies on auricular VNS suggest that single-session use offers only modest physical benefits and limited standalone efficacy, highlighting the need to combine VNS with intensive, task-specific training for meaningful motor recovery [18, 19].

Recent randomized controlled trials (RCTs) have demonstrated that VNS paired with rehabilitation may improve upper limb motor outcomes, most commonly assessed by the Fugl-Meyer Assessment Upper Extremity (FMA-UE), Preliminary studies using tVNS have also reported encouraging results, though evidence quality and reproducibility remain variable across stimulation parameters and study designs [20-22]. Overall, VNS appears to be safe and well tolerated, with most adverse events being mild or transient [15]. Nevertheless, the magnitude and consistency of treatment effects, as well as potential differences between implanted and transcutaneous modalities, remain to be systematically evaluated.

This meta-analysis aimed to evaluate, in patients with stroke (P: Population), whether rehabilitation combined with vagus nerve stimulation (I: Intervention), either implanted or transcutaneous, compared with sham stimulation or conventional therapy (C: Comparison), improves upper limb motor recovery as measured by the Fugl-Meyer Assessment Upper Extremity (FMA-UE) (O: Outcome). This PICO (Population, Intervention, Comparison, and Outcome) framework guided the study design. Only randomized controlled trials (RCTs) were included to ensure methodological rigor, and subgroup analyses were conducted to compare the effects of implanted versus transcutaneous VNS. By synthesizing current evidence, this study seeks to provide an updated and quantitative summary of the efficacy and safety of VNS as an adjunct to stroke rehabilitation.

Methods

Inclusion and exclusion criteria

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and followed a pre-registered protocol [23] (PROSPERO registration number: CRD420251218925).

Eligibility criteria were defined using the PICO (Population, Intervention, Comparison, and Outcome) framework. The population (P) included adults with a confirmed diagnosis of stroke, irrespective of lesion type or hemisphere, who demonstrated residual upper-limb motor impairment. Both ischemic and hemorrhagic stroke and patients in the subacute or chronic phases were eligible. The intervention (I) consisted of vagus nerve stimulation (VNS) delivered through either implanted cervical stimulation or transcutaneous auricular stimulation, administered in conjunction with standard rehabilitation practices. Studies were eligible if they used any form of structured upper-limb rehabilitation paired with VNS or if VNS was administered alongside task-specific motor training. The comparison (C) included sham stimulation, standard rehabilitation without VNS, or any active control condition not involving vagus nerve stimulation. The primary outcome (O) was upper-limb motor impairment measured by the Fugl-Meyer Assessment-Upper Extremity (FMA-UE) scale. Studies were required to report FMA-UE data to be eligible for inclusion.

Only randomized controlled trials were included. Non-randomized studies, case reports, reviews, conference abstracts, and animal studies were excluded. Because multiple high-quality randomized trials are now available in this domain, restricting inclusion to RCTs ensured methodological consistency and minimized bias. For studies reporting several post-intervention time points, the longest follow-up was used for quantitative synthesis, and between-group standardized mean differences (SMDs) were prioritized when both within-group and between-group data were available.

Data sources and search strategy

A comprehensive electronic search was performed in the PubMed, Cochrane Library and China National Knowledge Infrastructure (CNKI) databases, encompassing studies published from 2010 to November 2025. To ensure a comprehensive retrieval of relevant studies, the search strategy combined Medical Subject Headings (MeSH) terms with free-text keywords. Boolean operators such as “AND” and “OR” were applied to structure the queries and ensure precise retrieval of studies relevant to the research question. The PubMed search strategy was formulated according to the PICO framework, with the final query defined as: (“Stroke”[Mesh]) AND “Vagus Nerve Stimulation”[Mesh].

For the Cochrane Library, the strategy consisted of two individual queries: "Stroke" (#1) and "Vagus Nerve Stimulation" (#2), with the final query constructed as #1 AND #2. Equivalent search logic was adapted for CNKI to include Chinese-language studies. The last search was conducted in November 2025. Only randomized controlled trials (RCTs) were included. To maximize completeness, the reference lists of all included studies and relevant systematic reviews were manually examined to identify any additional eligible publications.

Selection strategy

The screening and study selection procedures were carried out with the assistance of EndNote reference management software. Two reviewers first independently screened study titles to eliminate duplicates, non–randomized studies, meta-analyses, systematic reviews, scoping reviews, and case reports. The remaining abstracts were subsequently evaluated independently by both reviewers to determine their eligibility. Inter-rater agreement was assessed using Cohen's kappa ($\kappa = 0.85$), demonstrating a high level of concordance. Any disagreements at this stage were addressed through discussion until consensus was reached. Full-text articles of all potentially eligible studies were then independently reviewed by both researchers to confirm final inclusion, with any remaining non-RCT designs excluded at this step. All final disagreements were settled through discussion and consensus, ensuring the rigor and reliability of the overall study selection process.

Data extraction

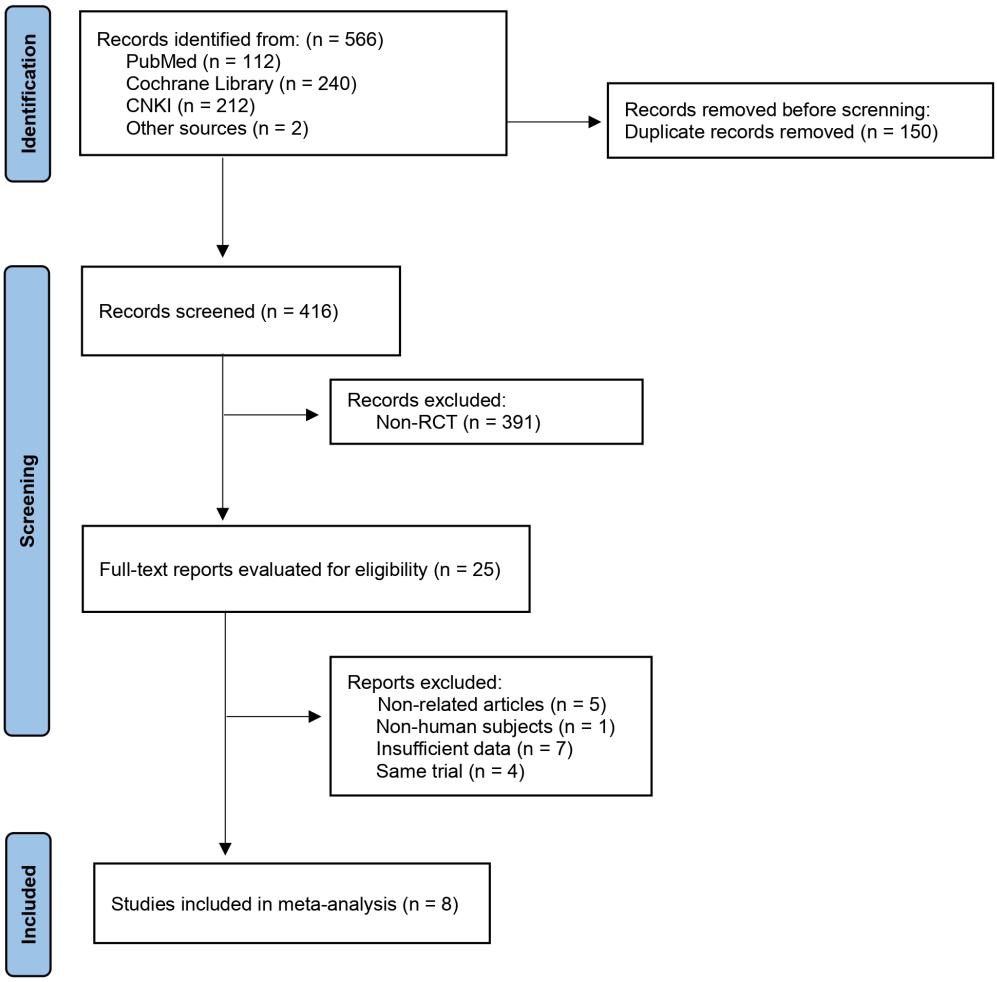
Data used for quantitative analysis were extracted using a standardized form. The information collected included the first author's name, year of publication, country where the study was conducted, study design, total sample size and group allocation at follow-up, gender distribution, mean age, stroke-related characteristics, types of vagus nerve stimulation used, inclusion and exclusion criteria, and detailed descriptions of the interventions.

Risk of bias of individual studies

The methodological quality of all included studies was independently evaluated by two reviewers using the Risk of Bias 2 (ROB2) tool, as only randomized controlled trials were included. The ROB2 assessment examined potential bias related to the randomization process, deviations from planned interventions, issues related to missing outcome information, approaches to measuring outcomes, and risks of selective reporting. Any disagreements between reviewers were resolved through discussion to achieve consensus [24].

Data analysis(meta-analysis)

Statistical analyses were conducted using Comprehensive Meta-Analysis version 3.7 (Biostat, Englewood, NJ, USA). Standardized mean differences (SMDs), standard errors (SEs), and 95% confidence intervals (CIs) were calculated to estimate the effects of vagus nerve stimulation on rehabilitation-related



Abbreviations: CNKI, China National Knowledge Infrastructure

Figure 1 – Flow diagram of study selection for randomized controlled trials evaluating vagus nerve stimulation after stroke

Table 1 Baseline characteristics of the studies included in the meta-analysis

Source	Country	Study Design	Age Range	% Female	Phase post-stroke	Mean FMA-UE at Baseline (VNS Group)	Mean FMA-UE at Baseline (Sham)	Experimental Group (VNS)	Control Group (Sham)	Intervention Type	Comparator	Detailed Intervention	Follow-up Duration & Schedule	Adverse events
Dawson et al. (2016)	UK	open-label, assessor-blinded, RCT	VNS Group: 57.9 ± 17.2; Sham: 60.7 ± 10.7	VNS Group 22.2%; Sham 18.2%	Chronic (≥ 6 months)	40.1±9.7	45.3± 8.4	9	11	Implanted Vagus Nerve Stimulation (VNS)	Rehabilitation only	VNS: therapist-triggered 0.5 s train, 15 pulses, 0.8 mA, 100 µs, 30 Hz, left side, once per movement; Duration: 6 weeks (3x/week, total 18 sessions).	30 Days: 1, 7, 30 days post-rehab	Vocal cord palsy, deglutition impairment, gastric distress, dysgeusia, hoarseness, tingling
Capone et al. (2017)	Italy	double-blind, RCT	VNS Group: 53.71 ± 5.88; Sham: 55.60 ± 7.12 #	VNS Group 42.9%; Sham 40.0%	Chronic (≥ 1 year)	22.29 ± 3.51	32.60 ± 6.43	7	5	Transcutaneous Vagus Nerve Stimulation (tVNS)	Sham tVNS	tVNS: left tragus, 20 Hz 0.3 ms, 30 s/5 min × 60 min, intensity above sensory below pain, left only; Duration: 2 weeks (5x/week, total 10 sessions).	Assessment only at the end of treatment	None
Kimberley et al. (2018)	USA	double-blind, RCT	VNS Group: 59.5± 7.4; Sham: 60± 13.5	VNS Group: 50%; Sham: 44.4%	Chronic (≥ 4 months)	29.5±6.4	36.4± 9.4	8	9	Implanted cervical Vagus Nerve Stimulation (VNS)	Sham VNS	Implanted cervical VNS paired: therapist-triggered 0.5-s trains at 30 Hz, 100 µs, 0.8 Ma, one train per movement; Duration: 6 weeks (3x/week, total 18 sessions).	90 Days: 1, 7, 30, 90 days	Surgical site infection, dyspnea, deglutition impairment, dysphonia
Wu et al. (2020)	China	single-blinded, RCT	VNS Group: 64.50 ± 9.97; Sham: 61.82 ± 10.63	VNS Group: 50%; Sham: 27.3%	VNS Group: 36.30 ± 9.23(d); Sham: 35.55 ± 6.47 (d)	17.50± 4.91	16.82± 3.89	10	11	Transcutaneous auricular vagus nerve stimulation (taVNS)	Sham taVNS	taVNS: 600 pulses/ train (20 Hz, 0.3 ms), 30 s/5 min cycle, 30 min/day; Duration: 15 days (Daily).	12 Weeks: 4 weeks and 12 weeks	Erythema
Wei et al. (2020)	China	open-label, RCT	VNS Group: 61.31± 11.54; Sham: 57.23± 10.17	VNS Group: 69.2%; Sham: 76.9%	VNS Group: 48.77± 24.74 (d); Sham: 50.38± 22.07(d)	32.85± 12.13	28.31± 13.55	13	13	Transcutaneous left auricular vagus nerve stimulation (taVNS)	Sham taVNS	taVNS: optimum intensity, 25 Hz, 100 µs, lasting 30 s every 30 s; Duration: 4 weeks (5x/week, 20 sessions).	Assessment only at the end of treatment	Nausea, vomiting, ear pain
Zhang et al. (2020)	China	triple-blind, RCT	VNS Group: 66.1± 1.49; Sham: 64.1± 1.03	VNS Group: 47.6%; Sham: 61.9%	VNS Group: 38±15 (d); Sham: 36.86± 2(d)	18.76± 0.94	17.9± 0.76	21	21	Transcutaneous left auricular vagus nerve stimulation (taVNS)	Sham taVNS	taVNS: 0.5 mA, 20 Hz, 30 s every 2 min, total 30 min per session; Duration: 6 weeks (3x/week, total 18 sessions)..	20 Weeks: 12 weeks and 20 weeks post-rehab	None
Dawson et al. (2021)	UK	triple-blind, RCT	VNS Group: 59.1± 10.2; Sham: 61.1 ±9.2	VNS Group: 35.8%; Sham: 34.5%	Chronic (≥ 9 months)	34.4±8.2	35.7± 7.8	53	55	Implanted cervical Vagus Nerve Stimulation (VNS)	Sham VNS	Implanted cervical VNS: Therapist triggered VNS 0.8 mA (two cases 0.7/0.6 mA), 100 µs, 30 Hz, 0.5 s per movement repetition; Duration: 6 weeks (3x/week, 18 sessions; ~90 min/ session; ~300+ reps/ session).	90 Days: 1, 90 days	Vocal fold paralysis
Badran et al. (2023)	USA	double-blind, RCT	VNS Group: 57.33 ± 8.28 ; Sham: 58.71 ± 6.45	VNS Group: 56%; Sham: 29%	Chronic (≥ 6 months)	36.56 ± 7.94	38.57 ± 10.47	9	7	Transcutaneous Motor-Activated Auricular Vagus Nerve Stimulation (MAAVNS)	unpaired taVNS	MAAVNS: EMG-triggered closed-loop taVNS (<100 ms delay), bilateral cymba conchae + tragus electrodes, 25 Hz, 500 µs, 2× perceptual threshold, 5 s trains repeated, mean 36,070 ± 3,205 pulses; Duration: 4 weeks (3x/week, 12 sessions).	8 Weeks: Post-session 3, 6, 9, 12; 2 weeks and 8 weeks after intervention	None

#The patient's age range in Capone et al. (2017) is provided as Mean ± SE.
Abbreviations: VNS, vagus nerve stimulation; tVNS, transcutaneous vagus nerve stimulation; taVNS, transcutaneous auricular vagus nerve stimulation; RCT, randomized controlled trial; EMG, electromyography; mA, milliampere; Hz, hertz; µs, microsecond; ms, millisecond; s, second; min, minute; SE, standard error.

outcomes. Forest plots were generated to provide a visual summary of effect sizes. A p-value less than 0.05 was considered statistically significant.

To account for both within-study and between-study variability, a random-effects model was applied for all pooled analyses. Heterogeneity was planned to be evaluated using the I^2 statistic and Cochran's Q test, and potential publication bias was to be examined using funnel plots and Egger's regression. However, because the number of eligible studies did not exceed ten, these assessments were not performed in accordance with Cochrane recommendations. Instead, clinical and methodological heterogeneity were qualitatively examined by comparing variations in stimulation modality, intervention parameters, patient characteristics, and outcome measures across included trials.

All included studies employed between-group randomized controlled trial designs; therefore, SMDs were derived using post-intervention means and standard deviations. The overall certainty of evidence for each outcome was assessed using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework, taking into consideration risk of bias, inconsistency, indirectness, imprecision, and potential publication bias.

Results

Study and identification and selection

The PRISMA diagram outlines the identification and screening procedures used in this meta-analysis (Figure 1). A total of 566 records were retrieved from all databases (PubMed, n = 112; Cochrane Library, n = 240; CNKI, n = 212; additional sources, n = 2). After removing 150 duplicate entries, 416 records proceeded to title and abstract screening, during which 391 were excluded because they did not meet the eligibility criteria or were non-randomized designs.

Subsequently, 25 full-text articles were assessed in detail. Of these, 5 were excluded for being unrelated to the research question, 1 was a non-human study, 7 lacked sufficient extractable data, and 4 represented duplicate publications of the same trial.

In the end, 8 randomized controlled trials involving a total of 262 participants met all inclusion criteria and were incorporated into the final quantitative synthesis (Table 1).

Quality assessment of the included studies

Risk of bias for all included randomized controlled trials was evaluated using the ROB2 tool. As illustrated in Figure 2 and Figure 3, none of the studies were rated as having a high risk of bias in any domain. Although a few trials demonstrated some concerns in areas such as the randomization process or missing outcome data, these issues were generally minor and did not materially threaten the internal validity of the findings.

Certainty of evidence (GRADE))

Certainty of evidence was assessed using the GRADE framework [25-32], with detailed evidence profiles presented in Table 2. For the primary functional outcomes, the certainty remains limited. Downgrading was primarily driven by variability in stimulation parameters and intervention protocols across trials, differences in patient characteristics, and imprecision related to small sample sizes with wide confidence intervals. Additionally, the presence of some concerns in certain ROB2 domains contributed to downgrading for risk of bias.

Despite these limitations, the pooled results showed a consistent pattern of improvement in post-stroke motor

Table 2 Grading of Recommendations, Assessment, Development and Evaluation (GRADE) assessment of certainty of evidence

Outcomes	Studies, No.	Participants, No.	Pooled effect (95% CI)	Overall certainty (GRADE)	Reason for downgrades
Upper-limb motor function (all VNS modalities)	8	262	SMD 0.886 (0.098–1.674), p = 0.028	Moderate	Downgraded for risk of bias, inconsistency
Upper-limb motor function (implanted VNS only)	3	145	SMD 0.161 (−0.166–0.487), p = 0.335	Low	Downgraded for imprecision
Upper-limb motor function (transcutaneous VNS only)	5	117	SMD 1.332 (0.034–2.629), p = 0.044	Low	Downgraded for risk of bias, inconsistency and imprecision

Abbreviations: GRADE, Grading of Recommendations, Assessment, Development and Evaluations; VNS, Vagus nerve stimulation; SMD, Standardized Mean Difference; CI, Confidence interval.

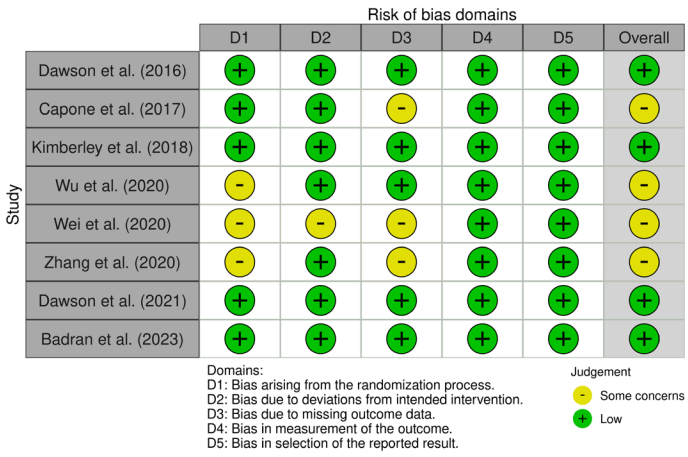


Figure 2 – ROB2 Traffic Light Plot

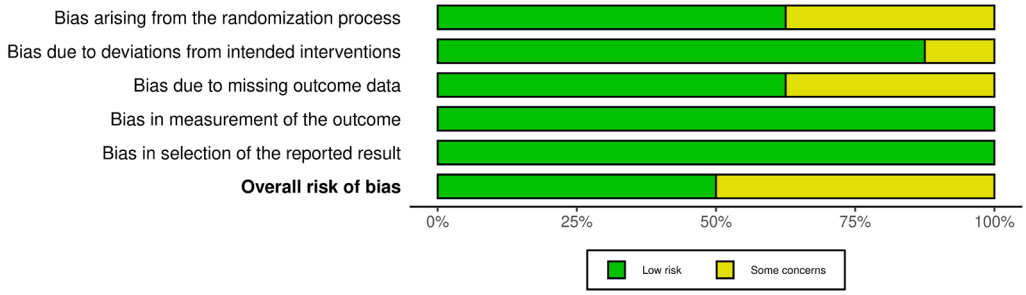


Figure 3 – ROB2 Summary Plot

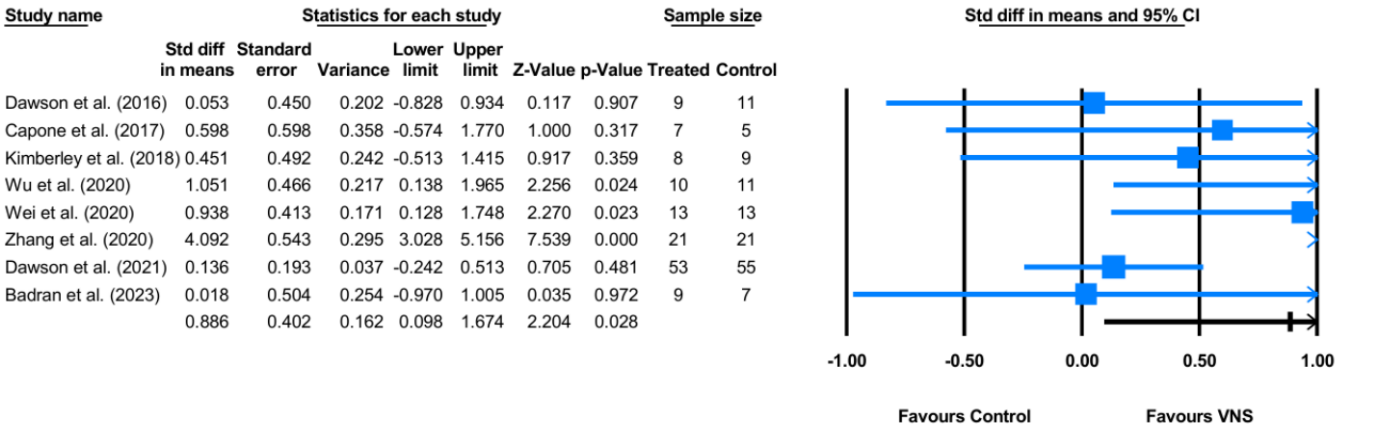


Fig. 4-1. Forest plot depicting the pooled effects of vagus nerve stimulation (all modalities combined) on upper-limb motor function after stroke.

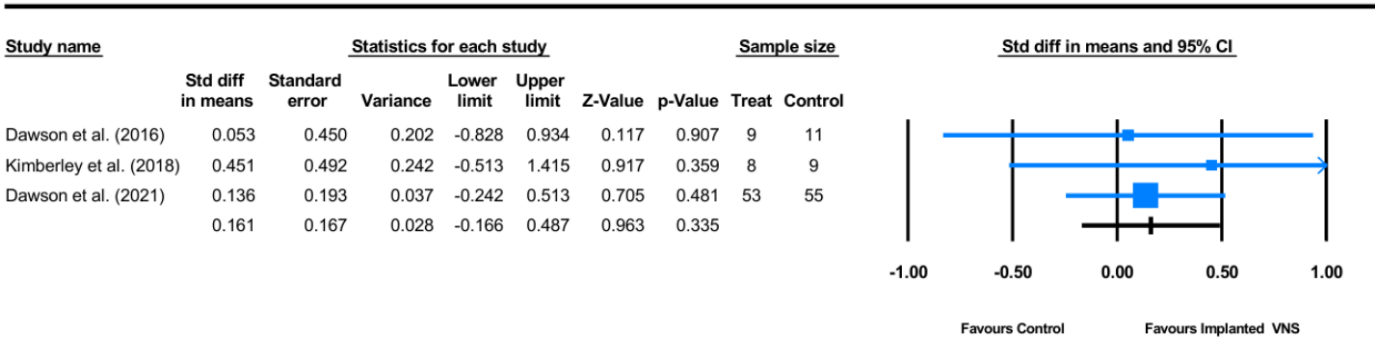


Fig. 4-2. Forest plot depicting the effects of implanted cervical vagus nerve stimulation on upper-limb motor outcomes compared with control conditions.

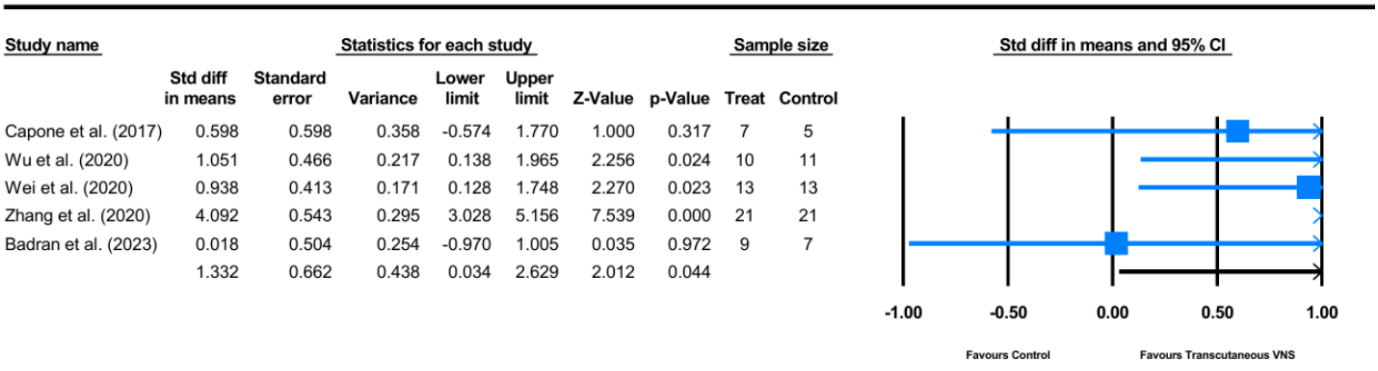


Fig. 4-3. Forest plot depicting the effects of transcutaneous vagus nerve stimulation on upper-limb motor outcomes compared with control conditions.

Figure 4 – Forest plot depicting the effects of vagus nerve stimulation on upper-limb motor function after stroke

outcomes among participants receiving vagus nerve stimulation compared with control groups. In a research area with a historically small number of randomized trials, this synthesis consolidates emerging evidence and provides important direction for the planning of future studies, including more accurate sample-size estimation, selection of appropriate primary endpoints, and greater standardization of stimulation approaches.

Participants

A total of 262 participants from eight randomized controlled trials were included in this systematic review and meta-analysis

(study details extracted from included trials). Sample sizes across studies ranged from 12 to 108 participants. Disease duration (time post-stroke) was reported in most trials and generally exceeded 4 months. Sex distribution varied by study, with reported female proportions ranging from 20% to 73%. Across the included trials, three studies employed implanted vagus nerve stimulation (VNS) [33-35], while five studies utilized transcutaneous VNS (tVNS) [36-40], reflecting variation in stimulation modality. All studies enrolled individuals with upper-limb motor impairment following stroke, with chronic motor deficits representing the predominant clinical profile. Baseline FMA-UE scores across the included trials ranged from 16.82 to 45.3, indicating a spectrum of impairment from severe to moderate [41]. To facilitate clinical interpretation, a standardized classification of FMA-UE scores has been included in the Appendix 1.

Interventions

In this review, all included trials investigated vagus nerve stimulation combined with upper-limb rehabilitation as the primary intervention for patients with chronic stroke. Across the eight randomized controlled trials, three studies [33-35] used implanted cervical VNS, whereas five [36-40] employed transcutaneous auricular VNS, including motor-activated paradigms.(36) In the implanted VNS trials [33-35], stimulation was delivered via a cervical cuff electrode on the left vagus nerve, with therapist-triggered 0.5-second trains at 30 Hz, 100 μs pulse width, and intensities typically around 0.8 mA, paired with task-specific upper-limb movements during in-clinic therapy sessions. In the transcutaneous VNS trials [36-40], stimulation was applied to the tragus or cymba conchae of the auricle, using frequencies between 20 and 25 Hz, pulse widths of 0.3–0.5 ms, and intensities adjusted above sensory threshold but below pain threshold, administered either as continuous or intermittent trains over 30- to 60-minute sessions.

Intervention duration and dosing varied across studies. Some protocols were relatively brief, such as 10-day [37] tVNS, whereas others, particularly those combining implanted VNS with intensive upper-limb training, involved structured programs delivered three times per week over six weeks, with approximately 90-minute sessions and several hundred task repetitions per session [34].

One trial [36] implemented an electromyography-triggered, motor-activaed tVNS paradigm, in which stimulation was automatically delivered within milliseconds of muscle activation during upper-limb tasks, providing a closed-loop pairing of VNS with voluntary movement. Overall, these intervention strategies were designed to enhance experience-dependent plasticity by coupling VNS with repetitive, task-specific upper-limb practice to promote functional motor recovery after stroke.

Comparator interventions

In the trials included in this review, comparator conditions were designed to control for both rehabilitation dose and nonspecific effects of device use. Several studies used standard upper-limb rehabilitation alone as the control condition, with no active VNS delivered.

In the implanted VNS trials [33-35], control groups received sham stimulation through deactivated or subtherapeutic device settings while undergoing the same task-specific training as the active VNS group. In the transcutaneous VNS trials [36-40], sham tVNS was commonly applied by placing electrodes at the same auricular locations but delivering no current or ineffective stimulation parameters, or by using unpaired stimulation that was not temporally linked to movement. Across all studies [33-40], control participants received comparable contact time and rehabilitation exposure, ensuring that any between-group differences could be attributed primarily to the presence or absence of effective VNS rather than differences in therapy intensity or attention.

Meta-analysis

A synthesis of eight randomized controlled trials encompassing 262 participants demonstrated consistent improvements in upper-limb motor function favoring vagus nerve stimulation over control conditions (Figure 4-1). The pooled standardized mean difference (SMD) for all VNS modalities combined was 0.886, with a 95% confidence interval from 0.098 to 1.674 and a p value of 0.028, indicating a statistically significant benefit associated with VNS. Individual trial effects varied, ranging from very small or negligible effects in studies such as Dawson et al. (2016) [35], Dawson et al. (2021) [34], and Badran et al. (2023) [36] to a very large effect in Zhang et al. (2020) [39], which contributed strongly to the upper bound of the pooled estimate. The forest plot illustrates study-specific SMDs and corresponding weights, with larger samples contributing more heavily to the overall result. Between-study variation is apparent and is likely related to differences in stimulation parameters, VNS modality, rehabilitation intensity, and participant characteristics across trials.

In the subgroup of trials employing implanted cervical VNS, the combined effect on upper-limb motor outcomes was small and not statistically significant (Figure 4-2). The pooled SMD was 0.161, with a 95% confidence interval from −0.166 to 0.487 and a p value of 0.335, suggesting that any additional benefit of implanted VNS beyond control conditions remains uncertain based on current evidence. Some studies in this subgroup reported modest gains [33], whereas others showed effects close to zero [34, 35] , and their relative weights reflected differences in sample size, with the largest implanted VNS trial [34] contributing most to the pooled estimate. These findings indicate that implanted VNS may have, at most, a modest effect on motor recovery within the contexts studied.

By contrast, the subgroup analysis of five trials using transcutaneous VNS (tVNS) showed a larger pooled effect size on upper-limb motor function (Figure 4-3). The pooled SMD was 1.332, with a 95% confidence interval from 0.034 to 2.629 and a p value of 0.044, consistent with a statistically significant and potentially large treatment effect, albeit with substantial imprecision. Zhang et al. (2020) [39] reported the largest SMD, whereas other tVNS studies such as Capone et al. (2017) [37], Wu et al. (2020) [40], and Wei et al. (2020) [38] demonstrated moderate benefits, and Badran et al. (2023) [36] reported a minimal effect. Taken together, these findings suggest that tVNS may be associated with greater motor gains than implanted VNS, although the wide confidence intervals and heterogeneity across trials underscore the influence of factors such as stimulation modality, intensity, treatment duration, and baseline motor impairment, and highlight the need for larger, head-to-head randomized trials to clarify differences between VNS approaches.

Secondary Outcomes and Other Effects

Beyond the primary FMA-UE scores, several trials evaluated secondary motor and sensory outcomes. The Wolf

Motor Function Test (WMFT) was utilized in studies included Dawson (2021) , Kimberley (2018) , and Badran (2023). Participants in the VNS groups demonstrated faster completion of distal tasks, including picking up small objects (e.g., paper clips), turning keys, and hand-eye coordination tasks. Muscle stiffness and spasticity were assessed using the Modified Ashworth Scale (MAS); however, most studies, including the pivotal Dawson (2021) trial, reported no significant difference in spasticity between the VNS group and the control group [33-36].

Safety and Adverse Events

VNS was generally well-tolerated, with profiles varying by modality. For implanted VNS (iVNS), adverse events included minor surgical site infections and transient hoarseness. Serious events, such as vocal cord paresis, were rare (e.g., 7% in the VNS-REHAB trial) and typically resolved [33-35]. Transcutaneous VNS (tVNS) showed an excellent safety profile, with minor side effects limited to local skin irritation (erythema or tingling) and rare systemic symptoms like nausea, without significant autonomic changes [36-40].

Discussion

This meta-analysis demonstrated that vagus nerve stimulation, when added to conventional upper-limb rehabilitation, is associated with greater improvements in motor outcomes after stroke compared with control conditions. When all VNS modalities were combined, the pooled effect size fell within the moderate-to-large range, indicating a clinically meaningful advantage for VNS-enhanced rehabilitation. These results reinforce the therapeutic potential of pairing neuromodulation with structured task-specific training to optimize post-stroke motor recovery.

Across the eight included randomized controlled trials, three investigated implanted cervical VNS [33-35] and five examined transcutaneous auricular VNS [36-40]. Implanted VNS showed small, generally modest improvements, whereas the tVNS subgroup demonstrated a larger pooled effect size, suggesting the possibility of greater benefit with non-invasive stimulation under certain conditions. Although variation existed across individual trials, the overall direction of effects consistently favored VNS over control, supporting its role as an adjunctive therapy to enhance upper-limb rehabilitation after stroke.

The variability observed across studies is likely shaped by several interacting factors. Stimulation parameters differed substantially in frequency, pulse width, current amplitude, and train duration, and the timing of stimulation relative to motor activity was not consistent. Some protocols used therapist-triggered stimulation delivered precisely during task-specific movements [33-35], whereas others relied on continuous or intermittently scheduled tVNS without strict alignment to task performance [36-40]. Intervention dosage also varied across trials, and patient characteristics such as stroke chronicity, lesion location, baseline motor ability, and cognitive function further contributed to differences in responsiveness. Preclinical and translational evidence supports pairing VNS with rehabilitation [42, 43]. These effects are mediated through neuromodulatory systems, including norepinephrine, acetylcholine, and serotonin, which create a biochemical environment that supports long-term potentiation and synaptic remodeling [42, 44]. Closed-loop VNS delivered during or immediately after successful motor attempts activates widespread networks across cortical, subcortical, and spinal levels, promoting task-specific synaptic modifications in damaged motor circuits [44, 45]. VNS also increases brain-derived neurotrophic factor (BDNF), upregulates plasticity-related genes such as Arc, and increases synaptic spine density, mechanisms that are crucial for motor learning and recovery [46, 47]. Importantly, these effects occur only when VNS is paired with task-specific training; stimulation delivered without behavioral engagement yields minimal benefit [42].

This mechanistic foundation directly correlates with the functional improvements observed across several trials. The faster completion of distal tasks in the Wolf Motor Function Test (WMFT) reported in studies included Dawson (2021), Kimberley (2018), and Badran (2023) suggests that the synaptic modifications facilitated by VNS are particularly effective at enhancing distal limb coordination, such as picking up small objects and turning keys. Because VNS is specifically paired with motor tasks to drive motor cortex reorganization, its effects are primarily focused on motor recruitment rather than sensory or spasticity modulation. This explains why most included studies reported significant motor gains without corresponding changes in tactile sensitivity, proprioception , or muscle stiffness as measured by the Modified Ashworth Scale (MAS) [33, 34, 36].

Collectively, this evidence supports the interpretation of VNS as a neuromodulatory “amplifier” rather than a stand-alone intervention. In both animal and human studies, VNS paired with rehabilitation has been shown to produce greater improvements in upper limb function than rehabilitation alone, suggesting an added benefit of stimulation beyond training, although the precise contribution of stimulation timing requires further clarification [36, 48]. This aligns with observations in the included stroke trials, where both intervention and control groups received structured upper-limb rehabilitation, and between-group differences reflect neuromodulatory enhancement of training effects [49, 50]. VNS acts as a neuromodulatory amplifier whose efficacy depends on precise pairing with high-intensity therapy, often requiring several hundred repetitions per session. Clinical benefits are further influenced by baseline impairment; patients with FMA-UE scores between 20–50 are typically targeted to avoid floor and ceiling effects. Additionally, VNS focuses on enhancing motor recruitment rather than modulating severe muscle spasticity, as reflected by the lack of significant change in MAS scores. Inter-individual variability in treatment response likely reflects differences in neurobiological reserve and cognitive engagement [33-36, 38].

Inter-individual variability in treatment response likely reflects differences in neurobiological reserve and cognitive engagement. Patients with greater preservation of corticospinal pathways, moderate baseline impairment, or intact cognitive function may respond more robustly to VNS-augmented therapy. Conversely, severe impairment, apraxia, or attention deficits may limit the ability to pair stimulation with meaningful task performance [45]. Although all included trials enrolled predominantly chronic stroke patients, preclinical data suggest that earlier intervention may enhance responsiveness, though benefits can still occur in later stages when training intensity is adequate [42, 51]. Future studies should include detailed imaging, neurophysiological markers, and cognitive assessments to better identify responder profiles.

Mechanistic evidence aligns with core principles of neurorehabilitation. Repetitive, task-specific practice promotes

use-dependent plasticity, while VNS enhances this effect by increasing neuromodulatory activity at key moments of learning [52]. Closed-loop stimulation enhances movement-specific plasticity in primary motor cortex, strengthens motor unit recruitment patterns, and increases motor drive to alpha motor neurons. These mechanisms provide a biological foundation for the motor improvements observed across VNS trials [45].

Clinical considerations further differentiate implanted and transcutaneous approaches. Implanted VNS allows precise, therapist-guided stimulation synchronized with motor tasks, facilitating timing-dependent plasticity [43]. A double-blind trial in individuals with chronic cervical spinal cord injury demonstrated meaningful improvements using implanted closed-loop VNS combined with high-intensity rehabilitation, although invasiveness, surgical risk, and cost remain limitations [45]. In contrast, tVNS is non-invasive and more scalable but shows greater variability in stimulation precision, and its comparative efficacy remains uncertain, although studies suggest it can improve motor function and daily activities when paired with exercise. Rigorous comparative effectiveness trials are needed to define its relative benefits, limitations, and cost-effectiveness across rehabilitation settings [4, 53].

Future studies should emphasize the use of consistent stimulation parameters, unified motor training protocols, and thorough outcome evaluation incorporating impairment-level measures, functional performance tests, kinematic analyses, neurophysiological indices, and patient-reported assessments. The integration of electromyography-triggered closed-loop tVNS, home-based stimulation systems, and telerehabilitation platforms may further expand access and increase training dose. Combinations of VNS with other neuromodulatory or pharmacologic strategies may enhance recovery through complementary mechanisms [54].

Importantly, our meta-analysis provides direct evidence that VNS-enhanced rehabilitation yields significantly greater improvements in upper-limb motor function compared with conventional therapy alone. Both implanted and transcutaneous approaches demonstrated directionally favorable outcomes, and tVNS in particular produced a statistically significant pooled effect size. These findings reinforce that VNS meaningfully augments the benefits of task-specific training and offer early clinical validation of the biological mechanisms highlighted above. As such, the results of this study strengthen the rationale for integrating VNS into contemporary neurorehabilitation practice and underscore its potential value as an effective therapeutic adjunct for post-stroke motor recovery.

Limitations

Despite the encouraging findings, several limitations should be acknowledged. The number of eligible randomized trials was small, and most studies enrolled relatively modest sample sizes, which reduces statistical power and limits the precision and generalizability of the results. Although all included trials used randomized designs, there was substantial methodological variation in stimulation parameters, the timing of VNS delivery relative to motor practice, the duration of treatment, and the structure of the rehabilitation programs. These inconsistencies likely contributed to variability across studies and make direct comparison more difficult.

A notable limitation relates to the substantial variation in effect sizes, particularly within the transcutaneous VNS subgroup where wide confidence intervals suggest imprecision and the potential influence of small-study effects. Differences in participant characteristics, including age, baseline motor function, stroke chronicity, and lesion characteristics, may have influenced individual responses. However, incomplete reporting in several trials limited the ability to perform more detailed subgroup analyses. The use of standardized mean differences allowed pooling across different motor outcome scales, but this approach may reduce clarity regarding the true magnitude of clinical improvement.

Risk-of-bias considerations also temper confidence in the findings. The ROB2 assessment indicated that several studies had some concerns in domains related to randomization procedures, missing outcome data, or the selection of reported outcomes. Although no trial was rated as high risk of bias, these uncertainties may still affect the overall reliability of the evidence. Although all included trials employed the FMA-UE as a primary outcome measure, enabling a consistent pooled analysis of motor impairment, there was substantial variation in the supplementary assessment tools used across studies (such as the Wolf Motor Function Test or Action Research Arm Test). The lack of standardized secondary measures across all trials, combined with variable follow-up durations, adds challenges to a more comprehensive interpretation of functional recovery beyond impairment scales. The possibility of publication bias should also be considered, since studies with null or negative findings are less likely to be published. Practical adoption of VNS faces hurdles, including the high cost and surgical risks associated with implanted systems. While tVNS is non-invasive and more scalable, its comparative efficacy and the requirement for intensive, therapist-guided training protocols remain challenges for widespread clinical implementation.

Overall, these limitations emphasize the need for larger and more methodologically rigorous randomized controlled trials that employ consistent stimulation parameters, harmonized motor outcome measures, and comprehensive reporting of participant and intervention characteristics. Such improvements are essential to strengthen the evidence base regarding the role of VNS in stroke rehabilitation.

Conclusion

This meta-analysis shows that vagus nerve stimulation, when paired with structured upper-limb rehabilitation, yields meaningful improvements in post-stroke motor recovery. Our findings confirm that VNS-augmented rehabilitation provides significantly greater gains in upper-limb motor function compared with rehabilitation alone, although the magnitude of benefit varied across studies, particularly between implanted and transcutaneous stimulation modalities. The observed variability is likely attributable to differences in stimulation parameters, timing of pairing with motor tasks, intervention dose, and patient characteristics such as stroke chronicity and baseline impairment. Despite this heterogeneity, the overall evidence consistently supports the effectiveness of VNS as an adjunct that enhances the motor gains achievable through task-specific rehabilitation.

These findings underscore the importance of integrating neuromodulatory approaches such as VNS into contemporary post-stroke rehabilitation frameworks, particularly when delivered alongside high-quality, repetitive, and goal-directed upper-limb training. A treatment model that combines evidence-based rehabilitation with targeted neuromodulation

has the potential to amplify experience-dependent plasticity and maximize functional recovery in appropriately selected patients.

Future research should address current limitations by employing larger, well-powered randomized trials, standardizing stimulation protocols, and harmonizing outcome measures across studies. Identifying clear responder profiles through neuroimaging, neurophysiological markers, and cognitive assessment will further refine patient selection and optimize treatment precision. Collaborative efforts among neurologists, physiatrists, physiotherapists, and neuromodulation specialists will be essential for designing holistic and scalable treatment pathways. By adopting a structured and individualized approach, clinicians may better leverage the synergistic effects of VNS and rehabilitation to improve upper-limb function, promote long-term recovery, and enhance quality of life for individuals living with chronic stroke.

Supplementary materials

The Supplementary information includes:

• Appendix 1. Classification of Fugl-Meyer Assessment Upper Extremity (FMA-UE) Severity Levels.

This supplemental material has been provided by the authors to give readers additional information about their work.

The file can be accessed using: https://www.editorialpark.com/download/article-supp/761/Appendix-1.-Classification-of-Fugl-Meyer-Assessment-Upper-Extremity-(FMA-UE)-Severity-Levels.docx.

Author Contributions: Conceptualization, M.H.C.; methodology, M.H.C. and W.H.C.; formal analysis, M.H.C.; investigation, M.H.C.; data curation, M.H.C.; writing – original draft preparation, M.H.C.; writing – review and editing, W.H.C. and M.C.C.; visualization, M.H.C.; supervision, W.H.C. and M.C.C.; project administration, M.C.C.; funding acquisition – not applicable. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: None.

Data availability statement: The data is available on reasonable request from the authors.

Patient Informed Consent Statement: Not applicable. This systematic review utilized data solely from previously published studies and did not involve any direct interaction with or participation from human subjects.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Feigin VL, Brainin M, Norrving B, Martins SO, Pandian J, Lindsay P, et al. World Stroke Organization: Global Stroke Fact Sheet 2025. Int J Stroke. 2025; 20 (2): 132-44. https://doi.org/10.1177/17474930241308142
2. Hilkens NA, Casolla B, Leung TW, de Leeuw F-E. Stroke. The Lancet. 2024; 403 (10446): 2820-36. https://doi.org/10.1016/S0140-6736(24)00642-1
3. Paul T, Cieslak M, Hensel L, Wiemer VM, Grefkes C, Grafton ST, et al. The role of corticospinal and extrapyramidal pathways in motor impairment after stroke. Brain Commun. 2023; 5 (1): fcac301. https://doi.org/10.1093/braincomms/fcac301
4. Aderinto N, Abraham IC, Olatunji G, Kokori E, Hasan A, Uwishema O. Mapping the role of vagus nerve stimulation in post-stroke arm motor recovery. J Neuroeng Rehabil. 2025; 22 (1): 215. https://doi.org/10.1186/s12984-025-01759-w
5. Cabral DF, Fried P, Koch S, Rice J, Rundek T, Pascual-Leone A, et al. Efficacy of mechanisms of neuroplasticity after a stroke. Restor Neurol Neurosci. 2022; 40 (2): 73-84. https://doi.org/10.3233/rnn-211227
6. Sikuka HM, Lupenga J, Nkhata L. Predictors of upper limb motor recovery in stroke survivors: a pre-post test study design. BMJ Open. 2024; 14 (11): e081936. https://doi.org/10.1136/bmjopen-2023-081936
7. Gil-Salcedo A, Dugravot A, Fayosse A, Jacob L, Bloomberg M, Sabia S, et al. Long-Term Evolution of Functional Limitations in Stroke Survivors Compared With Stroke-Free Controls: Findings From 15 Years of Follow-Up Across 3 International Surveys of Aging. Stroke. 2022; 53 (1): 228-37. https://doi.org/10.1161/strokeaha.121.034534
8. Wurzinger HE, Abzhandadze T, Rafsten L, Sunnerhagen KS. Dependency in Activities of Daily Living During the First Year After Stroke. Front Neurol. 2021; 12 736684. https://doi.org/10.3389/fneur.2021.736684
9. Semprini M. Editorial: Innovative approaches to promote stroke recovery. Front Neurosci. 2025; 19 1657124. https://doi.org/10.3389/fnins.2025.1657124
10. Das UC, Le NT, Vitoonpong T, Prapinpairoj C, Anannub K, Akarathanawat W, et al. An innovative model based on machine learning and fuzzy logic for tracking lower limb exercises in stroke patients. Sci Rep. 2025; 15 (1): 11220. https://doi.org/10.1038/s41598-025-90031-1
11. Huang J, Ji JR, Liang C, Zhang YZ, Sun HC, Yan YH, et al. Effects of physical therapy-based rehabilitation on recovery of upper limb motor function after stroke in adults: a systematic review and meta-analysis of randomized controlled trials. Ann Palliat Med. 2022; 11 (2): 521-31. https://doi.org/10.21037/apm-21-3710
12. Ren L, Ng S, Chi Keung CR, Woo J, Wong T. Efficacy of combined vagus nerve stimulation and exercise training for upper limb recovery in people with stroke: a systematic review and meta-analysis. Topics in Stroke Rehabilitation. 2026; 1-15. https://doi.org/10.1080/10749357.2026.2648211
13. Khan I, Shakir M, Vijayanarasimhan V, Lodhi BA, Parker JJ, Miller KJ, et al. Implantable Vagus Nerve Stimulator-Paired Neurorehabilitation for Upper Limb Function After Ischemic Stroke: Evidence From a Systematic Review and Meta-Analysis With Best Practice Recommendations. Neurosurgery. 2025; https://doi.org/10.1227/neu.0000000000003545

14. Roy JM, Musmar B, Ritz C, Sizdahkhani S, Karadimas S, Papadopoulos E, et al. Vagus nerve stimulation paired with rehabilitation for post-stroke recovery: A single center experience of patient satisfaction and outcomes. Clin Neurol Neurosurg. 2025; 257 109043. https://doi.org/10.1016/j.clineuro.2025.109043
15. Ananda R, Roslan MHB, Wong LL, Botross NP, Ngim CF, Mariapun J. Efficacy and Safety of Vagus Nerve Stimulation in Stroke Rehabilitation: A Systematic Review and Meta-Analysis. Cerebrovasc Dis. 2023; 52 (3): 239-50. https://doi.org/10.1159/000526470
16. Malley KM, Ruiz AD, Darrow MJ, Danaphongse T, Shiers S, Ahmad FN, et al. Neural Mechanisms Responsible for Vagus Nerve Stimulation-Dependent Enhancement of Somatosensory Recovery. Res Sq. 2024; https://doi.org/10.21203/rs.3.rs-3873435/v1
17. Wang L, Xu Q, Luo M, Xing X, Wang J, Liang Y, et al. Vagus nerve stimulation in various stages of stroke and associated functional impairments: A review. Neuroscience. 2025; 577 80-113. https://doi.org/10.1016/j.neuroscience.2025.04.037
18. Austelle CW, Cox SS, Wills KE, Badran BW. Vagus nerve stimulation (VNS): recent advances and future directions. Clin Auton Res. 2024; 34 (6): 529-47. https://doi.org/10.1007/s10286-024-01065-w
19. Çalı A, Özden AV, Ceylan İ. Effects of a single session of noninvasive auricular vagus nerve stimulation on sports performance in elite athletes: an open-label randomized controlled trial. Expert Rev Med Devices. 2024; 21 (3): 231-7. https://doi.org/10.1080/17434440.2023.2299300
20. Dawson J, Engineer ND, Cramer SC, Wolf SL, Ali R, O'Dell MW, et al. Vagus Nerve Stimulation Paired With Rehabilitation for Upper Limb Motor Impairment and Function After Chronic Ischemic Stroke: Subgroup Analysis of the Randomized, Blinded, Pivotal, VNS-REHAB Device Trial. Neurorehabil Neural Repair. 2023; 37 (6): 367-73. https://doi.org/10.1177/15459683221129274
21. Lin S, Rodriguez CO, Wolf SL. Vagus Nerve Stimulation Paired With Upper Extremity Rehabilitation for Chronic Ischemic Stroke: Contribution of Dosage Parameters. Neurorehabil Neural Repair. 2024; 38 (8): 607-15. https://doi.org/10.1177/15459683241258769
22. Ramos-Castaneda JA, Barreto-Cortes CF, Losada-Floriano D, Sanabria-Barrera SM, Silva-Sieger FA, Garcia RG. Efficacy and Safety of Vagus Nerve Stimulation on Upper Limb Motor Recovery After Stroke. A Systematic Review and Meta-Analysis. Frontiers in Neurology. 2022; Volume 13 - 2022 https://doi.org/10.3389/fneur.2022.889953
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021; 372 n71. https://doi.org/10.1136/bmj.n71
24. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019; 366 l4898. https://doi.org/10.1136/bmj.l4898
25. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011; 64 (4): 383-94. https://doi.org/10.1016/j.jclinepi.2010.04.026
26. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011; 64 (4): 395-400. https://doi.org/10.1016/j.jclinepi.2010.09.012
27. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011; 64 (4): 401-6. https://doi.org/10.1016/j.jclinepi.2010.07.015
28. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol. 2011; 64 (4): 407-15. https://doi.org/10.1016/j.jclinepi.2010.07.017
29. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol. 2011; 64 (12): 1277-82. https://doi.org/10.1016/j.jclinepi.2011.01.011
30. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011; 64 (12): 1283-93. https://doi.org/10.1016/j.jclinepi.2011.01.012
31. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol. 2011; 64 (12): 1294-302. https://doi.org/10.1016/j.jclinepi.2011.03.017
32. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011; 64 (12): 1303-10. https://doi.org/10.1016/j.jclinepi.2011.04.014
33. Kimberley TJ, Pierce D, Prudente CN, Francisco GE, Yozbatiran N, Smith P, et al. Vagus Nerve Stimulation Paired With Upper Limb Rehabilitation After Chronic Stroke. Stroke. 2018; 49 (11): 2789-92. https://doi.org/10.1161/strokeaha.118.022279
34. Dawson J, Liu CY, Francisco GE, Cramer SC, Wolf SL, Dixit A, et al. Vagus nerve stimulation paired with rehabilitation for upper limb motor function after ischaemic stroke (VNS-REHAB): a randomised, blinded, pivotal, device trial. Lancet. 2021; 397 (10284): 1545-53. https://doi.org/10.1016/s0140-6736(21)00475-x
35. Dawson J, Pierce D, Dixit A, Kimberley TJ, Robertson M, Tarver B, et al. Safety, Feasibility, and Efficacy of Vagus Nerve Stimulation Paired With Upper-Limb Rehabilitation After Ischemic Stroke. Stroke. 2016; 47 (1): 143-50. https://doi.org/10.1161/strokeaha.115.010477
36. Badran BW, Peng X, Baker-Vogel B, Hutchison S, Finetto P, Rishe K, et al. Motor Activated Auricular Vagus Nerve Stimulation as a Potential Neuromodulation Approach for Post-Stroke Motor Rehabilitation: A Pilot Study. Neurorehabil Neural Repair. 2023; 37 (6): 374-83. https://doi.org/10.1177/15459683231173357
37. Capone F, Miccinilli S, Pellegrino G, Zollo L, Simonetti D, Bressi F, et al. Transcutaneous Vagus Nerve Stimulation Combined with Robotic Rehabilitation Improves Upper Limb Function after Stroke. Neural Plast. 2017; 2017 7876507. https://doi.org/10.1155/2017/7876507
38. Wei X. Effect of TaVNS combined with upper limb training on upper limb motor function and brain plasticity of ischemic stroke subjects. Tianjin University of Sport. 2020.
39. Zhang LP, Yu ML, Wang SR, et al. Effect of transcutaneous vagus nerve stimulation on the recovery of upper limbmotorfunctioninpatients with ischemic stroke. Chinese J Rehabil Med. 2020; 35 1316–20.
40. Wu D, Ma J, Zhang L, Wang S, Tan B, Jia G. Effect and Safety of Transcutaneous Auricular Vagus Nerve Stimulation on Recovery of Upper Limb Motor Function in Subacute Ischemic Stroke Patients: A Randomized Pilot Study. Neural Plast. 2020; 2020 8841752. https://doi.org/10.1155/2020/8841752
41. Ierardi E, van Wijck F, Ali M, Best C, Coupar F. Defining Severity Levels for Post-Stroke Upper Limb Motor Impairment and Activity Limitation: A Systematic Review. NeuroRehabilitation. 2026; 58 (1): 3-16. https://doi.org/10.1177/10538135251393516
42. Darrow MJ, Torres M, Sosa MJ, Danaphongse TT, Haider Z, Rennaker RL, et al. Vagus Nerve Stimulation Paired With Rehabilitative Training Enhances Motor Recovery After Bilateral Spinal Cord Injury to Cervical Forelimb Motor Pools. Neurorehabil Neural Repair. 2020; 34 (3): 200-9. https://doi.org/10.1177/1545968319895480

43. Yu M, Wang S. The Effect of Vagus Nerve Stimulation on the Rehabilitation of Stroke: A Systematic Review and Meta-analysis. Arch Phys Med Rehabil. 2026; https://doi.org/10.1016/j.apmr.2026.01.009
44. Ganzer PD, Darrow MJ, Meyers EC, Solorzano BR, Ruiz AD, Robertson NM, et al. Closed-loop neuromodulation restores network connectivity and motor control after spinal cord injury. Elife. 2018; 7 https://doi.org/10.7554/eLife.32058
45. Kilgard MP, Epperson JD, Adehunoluwa EA, Swank C, Porter AL, Pruitt DT, et al. Closed-loop vagus nerve stimulation aids recovery from spinal cord injury. Nature. 2025; 643 (8073): 1030-6. https://doi.org/10.1038/s41586-025-09028-5
46. Sargusingh MJ, Addo JJA, Damaser MS, Zimmern P, Hays SA, Hernandez-Reynoso AG. Enhancing Neuroplasticity via vagus nerve stimulation to improve urinary dysfunction after spinal cord injury: a perspective. Bioelectron Med. 2025; 11 (1): 15. https://doi.org/10.1186/s42234-025-00178-5
47. Gargus M, Ben-Azu B, Landwehr A, Dunn J, Errico JP, Tremblay M. Mechanisms of vagus nerve stimulation for the treatment of neurodevelopmental disorders: a focus on microglia and neuroinflammation. Front Neurosci. 2024; 18 1527842. https://doi.org/10.3389/fnins.2024.1527842
48. Zhao K, Yang J, Huang J, Zhao Z, Qu Y. Effect of vagus nerve stimulation paired with rehabilitation for upper limb function improvement after stroke: a systematic review and meta-analysis of randomized controlled trials. Int J Rehabil Res. 2022; 45 (2): 99-108. https://doi.org/10.1097/mrr.0000000000000509
49. Liu CY, Russin J, Adelson DP, Jenkins A, Hilmi O, Brown B, et al. Vagus nerve stimulation paired with rehabilitation for stroke: Implantation experience from the VNS-REHAB trial. J Clin Neurosci. 2022; 105 122-8. https://doi.org/10.1016/j.jocn.2022.09.013
50. Kimberley TJ, Cramer SC, Wolf SL, Liu C, Gochyyev P, Dawson J, et al. Long-Term Outcomes of Vagus Nerve Stimulation Paired With Upper Extremity Rehabilitation After Stroke. Stroke. 2025; 56 (8): 2255-65. https://doi.org/doi:10.1161/STROKEAHA.124.050479
51. Korupolu R, Miller A, Park A, Yozbatiran N. Neurorehabilitation with vagus nerve stimulation: a systematic review. Front Neurol. 2024; 15 1390217. https://doi.org/10.3389/fneur.2024.1390217
52. Saylor A, Patrick L, Reddy CG, Gandhi R. Vagus Nerve Stimulation Paired With Upper Extremity Rehabilitation for Chronic Stroke: Real-World Implementation and Outcomes. Arch Rehabil Res Clin Transl. 2026; 8 (1): 100580. https://doi.org/10.1016/j.arrct.2025.100580
53. Gao L, Fan Y, Zhang N, Chen L. Transcutaneous Auricular Vagus nerve stimulation paired with exercise training for upper-limb motor function and activities of daily living after stroke: A systematic review and meta-analysis. Archives of Physical Medicine and Rehabilitation. 2026; https://doi.org/https://doi.org/10.1016/j.apmr.2026.03.008
54. Yan L, Qian Y, Li H. Transcutaneous Vagus Nerve Stimulation Combined with Rehabilitation Training in the Intervention of Upper Limb Movement Disorders After Stroke: A Systematic Review. Neuropsychiatr Dis Treat. 2022; 18 2095-106. https://doi.org/10.2147/ndt.S376399

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Review Article
DOI: https://doi.org/10.23950/jcmk/18694

Addressing underdiagnosis of chronic kidney disease in Kazakhstan: a public health perspective

Alimzhan Muxunov[1], Zhannat Kuanshaliyeva[1]

[1]*Clinical Academic Department of Internal Medicine, CF "University Medical Center", Astana, Kazakhstan*

Received: 2026-03-10.
Accepted: 2026-05-31.

J Clin Med Kaz 2026; 23(3): 85-91

Corresponding author:
Alimzhan Muxunov.
E-mail: alimzhan.muxunov@nu.edu.kz.
ORCID: 0009-0001-8860-0363.

ABSTRACT

Chronic kidney disease represents a growing public health challenge in Kazakhstan, yet a substantial proportion of cases remain undiagnosed. National registry data suggest lower recorded prevalence compared with global estimates, while local epidemiological studies indicate that many individuals live with impaired kidney function without awareness of their condition. Late diagnosis limits opportunities for early intervention, contributes to adverse clinical outcomes, and increases long-term healthcare costs, particularly when patients progress to end-stage kidney disease. Although Kazakhstan has strengthened prevention and early detection of several non-communicable diseases through primary healthcare services, chronic kidney disease is not currently included as a dedicated condition within national screening programs and is largely detected through opportunistic testing. This perspective examines key barriers to systematic chronic kidney disease detection, including laboratory requirements, geographic dispersion, rural - urban disparities, workforce constraints, and low public awareness. Drawing on international evidence, the paper discusses potential approaches to improve early detection that are compatible with national health system capacity. A phased, risk-stratified strategy integrated into existing primary care and chronic disease management pathways may offer a feasible starting point. Complementary delivery models, including point-of-care diagnostics and community-based outreach services, could further expand access in underserved regions. Context-specific economic and implementation research is needed to guide sustainable policy decisions and optimize screening strategies.

Keywords: Chronic Kidney Disease; Health Policy; Mass Screening; Public Health; Primary Health Care; Central Asia

Introduction

Chronic kidney disease (CKD) is a major global health issue affecting approximately 13% of the world population [1]. In recent decades, both the incidence and mortality associated with CKD have increased significantly, reflecting population aging, increasing prevalence of diabetes and hypertension, and improved survival from other chronic conditions [2]. Kazakhstan mirrors these global trends, with registry-based analyses of national healthcare data demonstrating an increasing burden of CKD in recent years [3].

Despite this upward trend, the overall number of officially registered cases remains relatively low compared with global estimates. Analysis of nationwide healthcare data indicates that the prevalence of officially registered CKD cases in Kazakhstan was only 38,287 per million population (3.8%) in 2020 [3]. Findings from local epidemiological studies further suggest that the actual prevalence of CKD may be considerably higher. Population-based screening initiatives and community studies have reported that up to 20-30% of individuals in the general population live with reduced kidney function [4,5].

However, these figures should also be interpreted with caution. In addition to being conducted in specific populations that may not be nationally representative, those studies rely on single measurements of kidney function and do not include albuminuria assessment, which may lead to misclassification and potential overestimation of CKD prevalence. Despite uncertainty regarding the exact magnitude, the available evidence consistently indicates that a considerable proportion of CKD cases remain undetected in Kazakhstan.

The underdiagnosis of CKD has important health implications. Most undiagnosed individuals with CKD are in the early stages of disease, which are typically asymptomatic. Diagnosis often occurs only when kidney impairment becomes clinically apparent or when complications arise. By that time, substantial decline in kidney function may already have occurred, which limits opportunities for interventions that could slow disease progression, such as blood pressure control, glycaemic management, and treatment with renoprotective medications [6]. Also, the risk of serious complications at later stages is significantly higher as demonstrated by numerous studies. Mortality risk rises progressively with declining kidney function, reaching approximately three to four times higher levels in patients with advanced CKD compared with those in earlier stages. Similarly, the risk of cardiovascular events increases substantially, with individuals in advanced stages experiencing up to two to four times higher cardiovascular risk [7–9]. In addition to these clinical outcomes, CKD is also associated with a significant decline in quality of life, especially at later stages of disease [10]. As kidney function deteriorates, patients frequently experience fatigue, physical limitations, psychological distress, and reduced social participation. Among individuals requiring dialysis, treatment burden, dietary restrictions, and frequent medical visits substantially affect daily functioning and overall well-being [11].

Beyond its clinical consequences, delayed detection also has significant economic implications. The costs of CKD management increase substantially with disease progression, particularly when patients reach end-stage kidney disease (ESKD) and require kidney replacement therapy such as dialysis or transplantation. An international analysis across 31 countries reported that mean annual direct costs increase from approximately $3,060 per patient in CKD stage G3a to $8,736 in stage G5, while the initiation of kidney replacement therapy increases costs dramatically, reaching $57,334 annually for hemodialysis and $75,326 during the first year following kidney transplantation [12]. Data from high-income countries show that a relatively small proportion of CKD patients who progress to ESKD, account for a large share of total CKD-related healthcare spending. For instance, in the United States, treatment of patients with ESKD represented nearly one-third of all CKD-related expenditures among Medicare beneficiaries in 2019, despite the comparatively small number of individuals requiring dialysis or transplantation [13].

Given the growing burden of CKD and the substantial clinical and economic consequences associated with late diagnosis, improving early detection of the disease represents an important public health priority. In Kazakhstan, strengthening strategies for timely identification of kidney disease may play a key role in improving patient outcomes and reducing long-term healthcare costs. This perspective paper examines the challenges associated with CKD detection in Kazakhstan and discusses potential opportunities for improving early diagnosis and disease management within the national healthcare system.

Current landscape of CD detection in Kazakhstan

In this paper, different approaches to CKD detection are distinguished as follows:

1. population-wide screening refers to systematic testing of all individuals within a defined population group, typically based on age criteria;
2. risk-based screening refers to systematic identification and testing of individuals with established risk factors, such as diabetes or hypertension;
3. opportunistic testing refers to unsystematic testing conducted during healthcare encounters, without a structured, programmatic approach; and
4. routine monitoring refers to regular, guideline-recommended assessment of kidney function in patients with known risk conditions as part of ongoing clinical management.

In Kazakhstan, prevention and early detection of non-communicable diseases (NCDs) are recognized as important priorities within the national healthcare system [14,15]. Over the past decade, the country has implemented several public health initiatives aimed at improving early diagnosis and management of chronic conditions through primary healthcare services [14]. These initiatives are largely integrated within the national population-wide screening programs, which are primarily delivered through primary healthcare facilities, which serve as the main entry point for preventive services and routine health monitoring [16]. Within this framework, screening activities are conducted for conditions including arterial hypertension, coronary artery disease, diabetes mellitus, cervical cancer, breast cancer, and colorectal cancer [16–19].

Despite the strong emphasis on prevention of chronic diseases, CKD has not yet been included as a separate condition within the national screening programs [16]. According to clinical guidelines, CKD detection in Kazakhstan is intended to occur through routine monitoring of high-risk patients, particularly those with diabetes and hypertension; however, in the absence of an organized screening framework, implementation is inconsistent, and testing often occurs in an opportunistic rather than systematic manner [3,20]. In contrast, several countries have developed more structured and systematic approaches to CKD detection, supported by standardized testing protocols, consistent identification of target populations and defined follow-up pathways. Examples include population-based screening initiatives in Japan [21], as well as risk-based screening programs in countries such as Taiwan [22], the United States, the United Kingdom, and Canada [23]. Integrating similar approaches within Kazakhstan's existing NCD prevention framework could potentially enhance identification of individuals with undiagnosed kidney disease and facilitate earlier management.

Barriers to early detection of CKD

Although early detection of CKD is essential, several barriers limit its effective implementation in many healthcare systems, including Kazakhstan. These barriers arise from diagnostic requirements, health system organization, access inequalities, and gaps in awareness and care pathways.

A key barrier to implementing systematic CKD screening in Kazakhstan relates to the operational and organizational requirements of laboratory-based detection at a population level. Unlike screening programs that rely on single procedures or point-of-care tests, CKD detection requires series of biochemical testing, including serum creatinine measurement

to estimate glomerular filtration rate (eGFR) and assessment of urinary albumin-to-creatinine ratio (uACR) [24]. Scaling these tests within a nationwide screening program would require sufficient laboratory capacity, reliable sample collection and transportation systems. Although Kazakhstan has an established primary healthcare network and ongoing efforts to expand access to diagnostic services, rural-urban disparities in the availability and quality of medical services suggest that consistent access to diagnostic testing may be uneven across Kazakhstan [25,26]. Review of rural healthcare infrastructure highlight underdevelopment of service facilities and limited availability of medical resources [27]. Also, the geographical context is important. Kazakhstan is the world's ninth-largest country by land area and has one of the lowest population densities globally, with fewer than 8 people per km² despite a total population of over 20 million [28,29]. In 2024, roughly 37.3% of the population lived in rural areas, indicating a large rural catchment that may face greater distances to health facilities and diagnostic services [30]. Geographic dispersion of the population, long travel distances in rural areas, and variability in local diagnostic capacity may complicate the organization of large-scale laboratory screening.

Workforce capacity may further constrain implementation. Primary healthcare providers serve as the main entry point for preventive services and already manage a substantial burden of NCDs. Systematic CKD screening would increase demands related to laboratory testing, interpretation of results, patient counseling, repeat assessments, and referral coordination. Without appropriate workforce planning and capacity strengthening, additional screening responsibilities may strain service delivery. Studies of Kazakhstan's health system identify ongoing public health implementation challenges and coordination constraints that relate to workforce capacity and readiness for expanded preventive services [31]. Reviews of rural healthcare in Kazakhstan also highlight that workforce distribution in Kazakhstan remains uneven, with staffing in rural areas continuing to present challenges and 82.8% of physicians concentrated in urban settings [27,32].

Another important barrier relates to limited public awareness of CKD. As described previously, CKD often progresses without noticeable symptoms in its early stages, reducing perceived need for preventive testing and limiting participation in screening programs. International and local evidence shows that awareness of CKD remains low, even among affected individuals, which contributes to delayed care-seeking and missed opportunities for early diagnosis [33–36]. Low awareness may also reduce screening uptake and adherence to follow-up after abnormal results, weakening the effectiveness of early detection initiatives [37].

Taken together, these challenges highlight key implementation constraints that may limit the feasibility and effectiveness of systematic CKD screening in Kazakhstan. Future screening strategies should be developed with explicit consideration of these contextual barriers.

International evidence on CKD screening and potential strategies for Kazakhstan

Although implementation of CKD screening presents practical challenges, international research provides important evidence on how screening strategies can be designed to maximize feasibility and effectiveness. A substantial body of evidence supports risk-based screening among individuals at increased risk of CKD. Studies conducted in the United States and Europe have demonstrated that screening programs focused on high-risk groups - such as individuals with diabetes, hypertension, cardiovascular disease, and older age - can facilitate earlier diagnosis and improve disease management [23,38,39]. Clinical guidelines developed by Kidney Disease: Improving Global Outcomes (KDIGO) initiative recommend routine assessment of kidney function in these high-risk populations [40]. Economic evaluations also support the risk-stratified approach. Cost-effectiveness analyses conducted in multiple high-income settings indicate that screening high-risk groups is generally cost-effective [41–44].

In contrast to risk-based approaches, broader population-wide screening remains controversial, largely due to concerns about cost-effectiveness of such approach. Earlier cost-effectiveness review generally concluded that population-wide CKD screening was not economically attractive compared with risk-based strategies, primarily due to higher program costs and less favorable incremental cost-effectiveness ratios [45]. However, a more recent systematic review presents a more nuanced perspective, suggesting that population-based screening may become economically reasonable under certain conditions, such as higher disease prevalence, improved testing efficiency, or integrated implementation within existing health systems [46].

For Kazakhstan, given the implementation barriers previously described, immediate population-based screening may be challenging to implement efficiently. A risk-based strategy focusing on individuals at elevated risk may therefore represent a more feasible and resource-efficient option. Integrating CKD detection into existing primary healthcare services and NCD management pathways could improve case detection while minimizing additional operational burden [47]. However, the transferability of evidence from high-income settings to Kazakhstan should be considered carefully. Differences in health system organization, laboratory infrastructure, workforce capacity, geographic accessibility, and referral pathways may influence both the feasibility and effectiveness of screening interventions. Therefore, local implementation and economic evaluation studies are essential to assess the real-world applicability of these strategies.

Beyond decisions about whom to screen, the mode of screening delivery is also critical for effective implementation. As mentioned earlier, screening for CKD typically relies on two key clinical indicators: eGFR, which reflects kidney filtration function, and uACR, which indicates the presence of kidney damage. Traditionally, these parameters have been measured using laboratory-based testing, requiring blood and urine samples analyzed in clinical laboratories. Recent technological advances have expanded the possibilities for CKD screening through the use of point-of-care (POC) diagnostic tools [48]. Portable POC devices can provide rapid eGFR and uACR measurements during the same clinical encounter, reducing dependence on centralized laboratory services and shortening diagnostic delays [49–51]. This approach may be particularly relevant for rural and remote areas where access to laboratory infrastructure is limited.

Community-based screening has been used in various settings to reach populations with limited access to facility-based services, particularly in rural or geographically dispersed regions. Such approaches may help identify individuals who do not routinely engage with primary healthcare, thereby improving population coverage and reducing access-related disparities [52,53]. Given Kazakhstan's large territorial area,

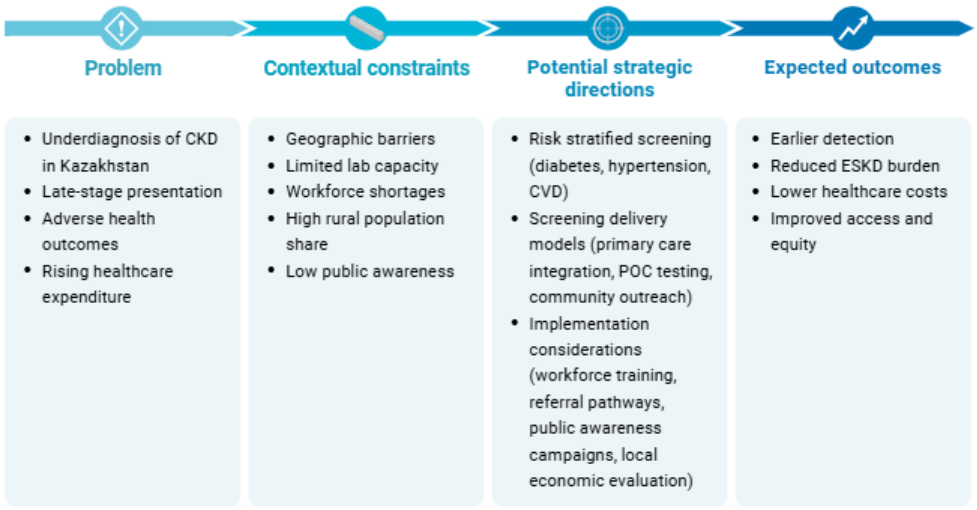


Figure 1 – Conceptual overview of potential approaches to improve early detection of chronic kidney disease in Kazakhstan

The figure summarizes key problem drivers, contextual constraints, potential strategic directions, and expected system-level outcomes.
Note: CKD, chronic kidney disease; CVD, cardiovascular disease; POC, point-of-care; ESKD, end-stage kidney disease.

low population density, and substantial rural population, community-based initiatives could potentially improve outreach to underserved groups. Community-based delivery models are feasible in Kazakhstan, as demonstrated by the long-standing use of government-supported mobile medical complexes that bring preventive and diagnostic services directly to residents of remote and rural settlements. Mobile medical complexes are legally defined as healthcare facilities equipped to expand service provision in rural areas, and have been deployed across over 1,100 remote settlements, improving the range and quality of care available to underserved populations [54,55].

Together, these approaches could support more flexible and accessible screening models in Kazakhstan, particularly when aligned with risk-based targeting and integrated within primary healthcare services. Their potential contribution lies in improving geographic reach and operational feasibility while maintaining linkage to formal diagnostic and management pathways. A conceptual overview of these key considerations and potential approaches is presented in Figure 1.

Future directions and policy recommendations

Addressing the underdiagnosis of CKD in Kazakhstan requires practical policy measures that take into account both the scale of the problem and the realities of the national health system. Drawing on the evidence presented and the structure of healthcare delivery in the country, several priority actions could strengthen early detection and management of kidney disease.

First, although CKD testing is recommended within routine follow-up of patients with diabetes, hypertension, and cardiovascular disease, its implementation could be strengthened and standardized across primary healthcare settings. Implementation should be supported by monitoring indicators such as screening coverage among high-risk groups, proportion of positive tests receiving confirmatory assessment, time to diagnosis, and linkage to care. Besides screening protocols, clear follow-up pathways should be defined to ensure that individuals with abnormal screening results are appropriately managed. This includes confirmatory testing protocols, referral criteria to nephrology services, and integration with existing primary care management systems.

Second, pilot implementation projects should be launched in selected regions to evaluate the feasibility of combining laboratory testing, POC diagnostics, and mobile outreach models. These pilots could inform scalable implementation strategies adapted to geographic and resource constraints. POC diagnostics and community-based outreach models should be considered as complementary delivery mechanisms integrated within primary healthcare services, rather than standalone approaches, to ensure continuity of care and appropriate follow-up.

Third, workforce preparation should focus on practical competencies, including CKD risk identification, test interpretation, referral coordination, and patient counseling, supported by simplified clinical pathways. Fourth, public communication strategies should emphasize the silent nature of early CKD and promote participation in preventive testing among high-risk groups. Finally, local economic and implementation research should guide national decision-making, ensuring that screening strategies reflect Kazakhstan's epidemiology, cost structures, and health system capacity.

In summary, improving early detection of CKD in Kazakhstan will likely require a phased, integrated approach that combines risk-stratified screening, expanded diagnostic access, community outreach, workforce strengthening, and locally informed policy planning. Aligning screening strategies with existing health system structures and contextual realities will be essential to achieving sustainable clinical and economic benefits.

Conclusion

CKD represents a growing public health challenge in Kazakhstan, with a substantial gap between estimated and officially registered cases. Although CKD detection is intended to occur through routine monitoring of high-risk patients, implementation remains inconsistent, and testing often occurs in an opportunistic rather than systematic manner. As a result, many individuals are diagnosed at advanced stages, when

opportunities for effective intervention are limited. International experience suggests that earlier detection can be achieved through more structured and systematically implemented approaches, including both population-based and risk-based strategies. In Kazakhstan, strengthening and standardizing routine monitoring of high-risk groups, supported by clear follow-up pathways and monitoring indicators, may represent a feasible starting point. Complementary use of point-of-care diagnostics and community-based outreach could further improve access, particularly in underserved areas. Generating local evidence will be essential to guide effective and sustainable policy decisions.

Author Contributions: Conceptualization, A. M.; methodology - not applicable; validation - not applicable; formal analysis - not applicable; investigation, A.M; resources - not applicable; data curation - not applicable; writing - original draft preparation, A.M.; writing - review and editing, A. M. and Z. K.; visualization, A. M.; supervision, Z. K.; project administration - not applicable; funding acquisition, A.M. and Z. K. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26197503).

Data availability statement: The corresponding author can provide the data supporting the study's conclusions upon request.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, Hobbs FDR. Global Prevalence of Chronic Kidney Disease – A Systematic Review and Meta-Analysis. Remuzzi G, editor. PLOS ONE. 2016 Jul 6;11(7):e0158765. https://doi.org/10.1371/journal.pone.0158765
2. Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, Adebayo OM, Afarideh M, Agarwal SK, Agudelo-Botero M, Ahmadian E, Al-Aly Z, Alipour V et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. 2020 Feb;395(10225):709–33. https://doi.org/10.1016/S0140-6736(20)30045-3
3. Zhakhina G, Mussina K, Yerdessov S, Gusmanov A, Sakko Y, Kim V, Syssoyev D, Madikenova M, Assan A, Kuanshaliyeva Z, Turebekov D, Yergaliyev K, Bekishev B, Gaipov A. Analysis of chronic kidney disease epidemiology in Kazakhstan using nationwide data for 2014–2020 and forecasting future trends of prevalence and mortality for 2030. Ren Fail. 2024 Dec 31;46(1):2326312. https://doi.org/10.1080/0886022X.2024.2326312
4. Muxunov A, Kuanshaliyeva Z, Sarria-Santamera A, Altynova V, Kozybayeva Z, Madikenova M, Kozhakhmet D, Suleimenov S, Bayakhmetova S, Gaipov A. Population-Based Screening during World Kidney Day in Kazakhstan: Prevalence and Risk Factors of Reduced Renal Function. Cent Asian J Nephrol. 2025 Sep 17;1(2):cajn007. https://doi.org/10.63946/cajn/17084
5. Nursultanova L, Kabulbayev K, Ospanova D, Tazhiyeva A, Datkhayev U, Saliev T, Tanabayeva S, Fakhradiyev I. Prevalence of chronic kidney disease in Kazakhstan: evidence from a national cross-sectional study. Sci Rep. 2023 Sep 7;13(1):14710. https://doi.org/10.1038/s41598-023-42031-2
6. Levey AS, Coresh J. Chronic kidney disease. The Lancet. 2012 Jan;379(9811):165–80. https://doi.org/10.1016/S0140-6736(11)60178-5
7. Chronic Kidney Disease Prognosis Consortium, Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong P, Coresh J, Gansevoort RT. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. The Lancet. 2010 Jun;375(9731):2073–81. https://doi.org/10.1016/S0140-6736(10)60674-5
8. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C yuan. Chronic Kidney Disease and the Risks of Death, Cardiovascular Events, and Hospitalization. N Engl J Med. 2004 Sep 23;351(13):1296–305. https://doi.org/10.1056/NEJMoa041031
9. Sharapov O, Daminov B. Cardiovascular Diseases in Dialysis Patients at Different Levels of Healthcare of the Republic of Uzbekistan. Cent Asian J Nephrol. 2025 Jul 21;1(1):cajn001. https://doi.org/10.63946/cajn/16627
10. Muxunov A, Almazan J, Kalinina D, Kuanshaliyeva Z, Gaipov A, Makhadiyeva D, Kabibulatova A, Madikenova M, Nabiyev A, Sarria-Santamera A. Health-related quality of life in chronic kidney disease patients in low- and lower-middle income countries: a systematic review and meta-analysis. Qual Life Res. 2026 Feb;35(2):28. https://doi.org/10.1007/s11136-025-04154-z
11. Asanova A, Bolatov A, Suleimenova D, Khazhgaliyeva Y, Shaisultanova S, Altynova S, Pya Y. Assessing Quality of Life in Hemodialysis Patients in Kazakhstan: A Cross-Sectional Study. J Clin Med. 2025 Jul 16;14(14). https://doi.org/10.3390/jcm14145021
12. Jha V, Al-Ghamdi SMG, Li G, Wu MS, Stafylas P, Retat L, Card-Gowers J, Barone S, Cabrera C, Garcia Sanchez JJ. Global Economic Burden Associated with Chronic Kidney Disease: A Pragmatic Review of Medical Costs for the Inside CKD Research Programme. Adv Ther. 2023 Oct;40(10):4405–20. https://doi.org/10.1007/s12325-023-02608-9
13. Johansen KL, Chertow GM, Foley RN, Gilbertson DT, Herzog CA, Ishani A, Israni AK, Ku E, Kurella Tamura M, Li S, Li S, Liu J, Obrador GT, O'Hare AM, Peng Y, Powe NR, Roetker NS, St. Peter WL, Abbott KC, Chan KE, Schulman IH, Snyder J, Solid C, Weinhandl ED, Winkelmayer WC, Wetmore JB. US Renal Data System 2020 Annual Data Report: Epidemiology of Kidney Disease in the United States. Am J Kidney Dis. 2021 Apr;77(4):A7–8. https://doi.org/10.1053/j.ajkd.2021.01.002
14. WHO Regional Office for Europe. WHO Country Office in Kazakhstan: annual activity report 2024. Copenhagen; 2025.
15. Government of the Republic of Kazakhstan. On the approval of the National Development Plan of the Republic of Kazakhstan until 2029 and the invalidation of certain decrees of the President of the Republic of Kazakhstan [Decree of the President of the Republic of Kazakhstan] [Internet]. 2024. Available from: https://adilet.zan.kz/rus/docs/U2400000611#z13

16. Ministry of Health of the Republic of Kazakhstan. On approval of the rules for screening examinations of the target population of the Republic of Kazakhstan. [Order of the Minister of Health of the Republic of Kazakhstan No. ҚР ДСМ-131/2020] [Internet]. 2020. Available from: https://adilet.zan.kz/rus/docs/V2000021572
17. Orazumbekova B, Issanov A, Atageldiyeva K, Berkinbayev S, Junusbekova G, Danyarova L, Shyman Z, Tashmanova A, Sarria-Santamera A. Prevalence of Impaired Fasting Glucose and Type 2 Diabetes in Kazakhstan: Findings From Large Study. Front Public Health. 2022 Feb 24;10:810153. https://doi.org/10.3389/fpubh.2022.810153
18. Bekbergenova Z, Derbissalina G, Umbetzhanova A, Koikov V, Bedelbayeva G. Evaluating the effectiveness of a screening program for cardiovascular diseases in Kazakhstan. Eur J Public Health. 2019 Nov 1;29(Supplement_4):ckz186.229. https://doi.org/10.1093/eurpub/ckz186.229
19. Shamsutdinova A, Kulkayeva G, Karashutova Z, Tanabayev B, Tanabayeva S, Ibrayeva A, Fakhradiyev I. Analysis of the Effectiveness and Coverage of Breast, Cervical, and Colorectal Cancer Screening Programs in Kazakhstan for the Period 2021-2023: Regional Disparities and Coverage Dynamics. Asian Pac J Cancer Prev. 2024 Dec 1;25(12):4371–80. https://doi.org/10.31557/APJCP.2024.25.12.4371
20. Ministry of Health of the Republic of Kazakhstan. Chronic kidney disease in adults: Clinical protocol. Astana: Ministry of Health of the Republic of Kazakhstan [Internet]. Available from: https://diseases.medelement.com/disease/хроническая-болезнь-почек-у-взрослых-кп-рк-2023/17777
21. Imai E, Yamagata K, Iseki K, Iso H, Horio M, Mkino H, Hishida A, Matsuo S. Kidney Disease Screening Program in Japan: History, Outcome, and Perspectives. Clin J Am Soc Nephrol. 2007 Nov;2(6):1360–6. https://doi.org/10.2215/CJN.00980207
22. Hwang S, Tsai J, Chen H. Epidemiology, impact and preventive care of chronic kidney disease in Taiwan. Nephrology. 2010 Jun;15(s2):3–9. https://doi.org/10.1111/j.1440-1797.2010.01304.x
23. Kushner P, Mende C. Screening programmes for early detection of chronic kidney disease in the USA and other English-speaking countries (Canada, Australia and UK): a systematic literature review. BMJ Open. 2025 Nov;15(11):e099966. https://doi.org/10.1136/bmjopen-2025-099966
24. Stevens PE, Ahmed SB, Carrero JJ, Foster B, Francis A, Hall RK, Herrington WG, Hill G, Inker LA, Kazancıoğlu R, Lamb E, Lin P, Madero M, McIntyre N, Morrow K, Roberts G, Sabanayagam D, Schaeffner E, Shlipak M, Shroff R, Tangri N, Thanachayanont T, Ulasi I, Wong G, Yang CW, Zhang L, Levin A. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int. 2024 Apr;105(4):S117–314. https://doi.org/10.1016/j.kint.2023.10.018
25. Spankulova L, Chulanova Z, Konyrbay A. Evaluating Healthcare Accessibility in Kazakhstan: Urban and Rural Perspectives. Eurasian J Econ Bus Stud. 2024 Jun 30;68(2):5–19. https://doi.org/10.47703/ejebs.v68i2.376
26. Rechel B, Sydykova A, Moldoisaeva S, Sodiqova D, Spatayev Y, Ahmedov M, Robinson S, Sagan A. Primary care reforms in Central Asia – On the path to universal health coverage? Health Policy OPEN. 2023 Dec;5:100110. https://doi.org/10.1016/j.hpopen.2023.100110
27. Turgambayeva A, Imanova Z, Tulegenova A. Rural Healthcare in Kazakhstan: Problems and Trends(Literature Review). J Health Dev. 2021;43(3):13–8. https://doi.org/10.32921/2225-9929-2021-3-43-13-18
28. World Bank Group. World Bank Open Data [Internet]. [cited 2026 Mar 6]. Population density (people per sq. km of land area) - Kazakhstan. Available from: https://data.worldbank.org/indicator/EN.POP.DNST?locations=KZ
29. World Bank Group. World Bank Open Data [Internet]. [cited 2026 Mar 6]. Land area (sq. km) - Kazakhstan. Available from: https://data.worldbank.org/indicator/AG.LND.TOTL.K2?locations=KZ
30. Mirasbek B. Population of the Republic of Kazakhstan as of July 1, 2024 [Internet]. Department of Population Statistics; 2024. Available from: https://stat.gov.kz/upload/iblock/978/4a34ya7f2gnk770d4c9wgh294fu7s8ro/%D0%91-18-06-%D0%9A%20(1%202024)%20%D0%B0%D0%BD%D0%B3.pdf
31. Aringazina A, Gulis G, Allegrante JP. Public Health Challenges and Priorities for Kazakhstan. Cent Asian J Glob Health. 2012 Nov 5;1(1). https://doi.org/10.5195/cajgh.2012.30
32. European Observatory on Health Systems and Policies, WHO Europe. OBS [Internet]. 2024 [cited 2026 Mar 6]. Health systems in action: Kazakhstan. Available from: https://eurohealthobservatory.who.int/publications/i/health-systems-in-action-kazakhstan-2024
33. Plantinga L, Tuot DS, Powe NR. Awareness of Chronic Kidney Disease Among Patients and Providers. Adv Chronic Kidney Dis. 2010;17(3):225–36. Located at: Scopus. https://doi.org/10.1053/j.ackd.2010.03.002
34. Muxunov A, Bulanov N, Makhmetov S, Sharapov O, Abdullaev S, Loboda O, Aiypova D, Haziyev E, Rashidov I, Tchokhonelidze I, Okpechi IG, Gaipov A. Awareness of chronic kidney disease and its risk factors in the former Soviet Union countries. Electron J Gen Med. 2023 Nov 1;20(6):em528. https://doi.org/10.29333/ejgm/13517
35. Bello AK, Okpechi IG, Levin A, Ye F, Saad S, Zaidi D, Houston G, Damster S, Arruebo S, Abu-Alfa AK, Ashuntantang G, Caskey FJ, Cho Y, Coppo R, Davids MR, Davison S, Gaipov A, Htay H, Jindal K, Lalji R, Madero M, Osman MA, Parekh RS, See E, Shah DS, Sozio SM, Suzuki Y, Tesar V, Tonelli M, Wainstein M, Wong M, Yeung E, Johnson D. ISN–Global Kidney Health Atlas: A report by the International Society of Nephrology: An Assessment of Global Kidney Health Care Status focussing on Capacity, Availability, Accessibility, Affordability and Outcomes of Kidney Disease. Brussels, Belgium: International Society of Nephrology; 2023.
36. Asanbek Kyzy A, Omorova A, Sakibaeva A, Zhumaeva K, Kudaiberdieva E, Rysbekova G, Nawa N, Ismanov K, Kalmatov R. Diabetic Kidney Disease Prevention in Type 2 Diabetes: Knowledge, Attitudes and Practices Study from Kyrgyzstan. Cent Asian J Nephrol. 2026 Apr 2;2(1):cajn012. https://doi.org/10.63946/cajn/18298
37. Tuot DS, Wong KK, Velásquez A, Crews D, Zonderman AB, Evans MK, Powe NR. CKD Awareness in the General Population: Performance of CKD-Specific Questions. Kidney Med. 2019;1(2):43–50. Located at: Scopus. https://doi.org/10.1016/j.xkme.2019.01.005
38. Farrell DR, Vassalotti JA. Screening, identifying, and treating chronic kidney disease: why, who, when, how, and what? BMC Nephrol. 2024 Jan 25;25(1). https://doi.org/10.1186/s12882-024-03466-5
39. Early chronic kidney disease intervention could save four European health systems up to €15.8 billion over 10 years [Internet]. 2024 [cited 2026 Mar 10]. Available from: https://www.astrazeneca.com/content/astraz/media-centre/press-releases/2024/early-chronic-kidney-disease-intervention-could-save-four-european-health-systems-up-to-15-billion-euro-over-10-years.html
40. Iatridi F, Carrero JJ, Cornec-Le Gall E, Kanbay M, Luyckx V, Shroff R, Ferro CJ. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease in Children and Adults: a commentary from the European Renal Best Practice

(ERBP). Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2025 Feb 4;40(2):273–82. https://doi.org/10.1093/ndt/gfae209
41. Palmer AJ, Valentine WJ, Chen R, Mehin N, Gabriel S, Bregman B, Rodby RA. A health economic analysis of screening and optimal treatment of nephropathy in patients with type 2 diabetes and hypertension in the USA. Nephrol Dial Transplant. 2008;23(4):1216–23. https://doi.org/10.1093/ndt/gfn082
42. Howard K, White S, Salkeld G, McDonald S, Craig JC, Chadban S, Cass A. Cost-effectiveness of screening and optimal management for diabetes, hypertension, and chronic kidney disease: a modeled analysis. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2010;13(2):196–208. Located at: Ovid MEDLINE(R) <2010 to 2011>. https://doi.org/10.1111/j.1524-4733.2009.00668.x
43. Siegel KR, Ali MK, Zhou X, Ng BP, Jawanda S, Proia K, Zhang X, Gregg EW, Albright AL, Zhang P. Cost-effectiveness of Interventions to Manage Diabetes: Has the Evidence Changed Since 2008? Diabetes Care. 2020 Jul;43(7):1557–92. https://doi.org/10.2337/dci20-0017
44. Thornton Snider J, Sullivan J, Van Eijndhoven E, Hansen MK, Bellosillo N, Neslusan C, O'Brien E, Riley R, Seabury S, Kasiske BL. Lifetime benefits of early detection and treatment of diabetic kidney disease. Mischak H, editor. PLOS ONE. 2019 May 31;14(5):e0217487. https://doi.org/10.1371/journal.pone.0217487
45. Komenda P, Ferguson TW, Macdonald K, Rigatto C, Koolage C, Sood MM, Tangri N. Cost-effectiveness of Primary Screening for CKD: A Systematic Review. Am J Kidney Dis. 2014 May;63(5):789–97. https://doi.org/10.1053/j.ajkd.2013.12.012
46. Li J, Zhao M, He Q. Screening for chronic kidney disease: a systematic review of emerging evidence and perspectives. Ren Fail. 2025 Dec 31;47(1):2572353. https://doi.org/10.1080/0886022X.2025.2572353
47. Vassalotti JA, Boucree SC. Integrating CKD Into US Primary Care: Bridging the Knowledge and Implementation Gaps. Kidney Int Rep. 2022 Mar;7(3):389–96. https://doi.org/10.1016/j.ekir.2022.01.1066
48. Gama RM, Nebres D, Bramham K. Community Point of Care Testing in Diagnosing and Managing Chronic Kidney Disease. Diagnostics. 2024 Jul 17;14(14):1542. https://doi.org/10.3390/diagnostics14141542
49. Gbinigie O, Thompson M, Price CP, Heneghan C, Plüddemann A. Point-of-Care creatinine testing for the detection and monitoring of chronic kidney disease [Internet]. Primary Care Diagnostic Horizon Scanning Centre Oxford; [cited 2026 Mar 10]. Available from: https://www.community.healthcare.mic.nihr.ac.uk/reports-and-resources/horizon-scanning-reports/point-of-care-creatinine-testing-for-the-detection-and-monitoring-of-chronic-kidney-disease
50. Thiengsusuk A, Youngvises N, Pochairach R, Taha RO, Sirisabhabhorn K, Muhamad N, Meesiri W, Chaijaroenkul W, Na-Bangchang K. Urinary Albumin-to-Creatinine Ratio (uACR) Point-of-Care (POC) Device with Seamless Data Transmission for Monitoring the Progression of Chronic Kidney Disease. Biosensors. 2025 Feb 24;15(3):145. https://doi.org/10.3390/bios15030145
51. Dally M, Amador JJ, Butler-Dawson J, Lopez-Pilarte D, Gero A, Krisher L, Cruz A, Pilloni D, Kupferman J, Friedman DJ, Griffin BR, Newman LS, Brooks DR. Point-of-Care Testing in Chronic Kidney Disease of Non-Traditional Origin: Considerations for Clinical, Epidemiological, and Health Surveillance Research and Practice. Ann Glob Health. 2023 Feb 1;89(1):7. https://doi.org/10.5334/aogh.3884
52. Zhang L, Zuo L, Xu G, Wang F, Wang M, Wang S, Lv J, Liu L, Wang H. Community-based screening for chronic kidney disease among populations older than 40 years in Beijing. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2007 Apr;22(4):1093–9. Located at: Ovid MEDLINE(R) <2005 to 2007>. https://doi.org/10.1093/ndt/gfl763
53. McCullough PA, Brown WW, Gannon MR, Vassalotti JA, Collins AJ, Chen SC, Bakris GL, Whaley-Connell AT. Sustainable Community-Based CKD Screening Methods Employed by the National Kidney Foundation's Kidney Early Evaluation Program (KEEP). Am J Kidney Dis. 2011 Mar;57(3):S4–8. https://doi.org/10.1053/j.ajkd.2010.11.010
54. Review of Kazakhstan's healthcare system: Results of 2020 and plans for 2021 - Official Information Source of the Prime minister of the Republic of Kazakhstan [Internet]. [cited 2026 Mar 10]. Available from: https://primeminister.kz/en/news/reviews/review-of-kazakhstans-healthcare-system-Results-of-2020-and-plans-for-20215968
55. Code of the Republic of Kazakhstan [Internet]. [cited 2026 Mar 10]. ON PUBLIC HEALTH AND HEALTHCARE SYSTEM - 'Adilet' LIS. Available from: https://adilet.zan.kz/eng/docs/K2000000360?utm_source=chatgpt.com

JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Review Article DOI: https://doi.org/10.23950/jcmk/18531

Association of Vaginal Microbiota Composition with Human Papillomavirus Persistence in Cervical Dysplasia and Cervical Cancer: a Systematic Review and Meta-Analysis

Elyanora A. Kydyrbayeva[1], Yerbolat M. Iztleuov[2], Indira A. Azamatova[1], Gulmira M. Iztleuova[3], Yerlan B. Sultangereyev[4], Nurgul A. Abenova[5]

[1]Department of Obstetrics and Gynecology, NJSC "Marat Ospanov West Kazakhstan Medical University", Aktobe, Kazakhstan
[2]Department of Radiology, Department of Radiology, NJSC "Marat Ospanov West Kazakhstan Medical University", Aktobe, Kazakhstan
[3]Department of Phthisiology and Dermatovenereology, NJSC "Marat Ospanov West Kazakhstan Medical University", Aktobe, Kazakhstan
[4]Department of Surgical Diseases No. 2, NJSC "Marat Ospanov West Kazakhstan Medical University", Aktobe, Kazakhstan
[5]Vice-Rector, NJSC "Marat Ospanov West Kazakhstan Medical University", Aktobe, Kazakhstan

Received: 2026-03-03.
Accepted: 2026-05-07.

J Clin Med Kaz 2026; 23(3): 92-102

Corresponding author:
Elyanora A. Kydyrbayeva.
E-mail: elya_vip_k@mail.ru.
ORCID: 0009-0009-4160-1893.

ABSTRACT

Introduction: Persistent high-risk human papillomavirus (hrHPV) infection is the main step in cervical carcinogenesis. The vaginal microbiota may modulate this risk, but the evidence is heterogeneous. This is because previous reviews have not provided a quantitative synthesis that specifically examines the persistence of HPV in the full progression of disease from dysplasia to invasive cancer. This specific gap reduces the transition of microbiota research into the stratification of clinical risks. The aim of the research is to systematically review and meta-analyze the association between the composition of the vaginal microbiota and hrHPV persistence in women with cervical dysplasia and cancer.

Methods: The researcher searched PubMed, Embase, Cochrane Library, and Web of Science (Jan 2015-Feb 2026) for observational studies. Pooled odds ratios (ORs) were calculated using a random-effects model. Study quality was assessed using the Newcastle-Ottawa Scale.

Results: Twenty-four studies were included. A non-Lactobacillus-dominant microbiota was related with 2.5 times higher odds of hrHPV persistence (pooled OR 2.51, 95% CI 1.95-3.23, I^2=68%, 18 studies). The relationship was stronger in women with invasive cervical cancer (OR 3.40) than in those with dysplasia only (OR 2.15).

Conclusion: Vaginal dysbiosis is significantly related to hrHPV persistence, with a significant effect in cervical cancer. The vaginal microbiome represents a potential biomarker for the stratification of risk and a target for therapeutic intervention, though clinical validation is needed before moving into clinical practice.

Keywords: vaginal microbiome; papillomavirus infections; uterine cervical dysplasia; uterine cervical neoplasms; disease progression.

Introduction

Cervical cancer is a major global health burden, and according to the World Health Organization (WHO), it is the fourth most common cancer among women worldwide [1, 2].

Importantly, persistent infection with high-risk human papillomavirus (hrHPV) is also seen as the main cause of cervical carcinogenesis [3, 4]. However, most infections that are from hrHPV do not last as they get removed by the immune system of the host in 1-2 years [5].

Therefore, the critical clinical problem is not the infection of HPV, but the persistence of HPV, that is, the prolonged presence of the infection after 12-24 months. This shows that persistence is an important step that allows the progression from initial infection to cervical intraepithelial neoplasia (CIN) and, finally, to invasive carcinoma [6, 7]. This progression shows that additional cofactors apart from the presence of viral factors are needed for oncogenesis.

Research has shown that the vaginal microenvironment is a key modulator of cervical health [8, 9, 10]. The vaginal microbiota, the community of microorganisms that live in the vaginal tract, is commonly classified into five main Community State Types (CSTs) [8, 11]. A CST dominated by Lactobacillus species (CST-I, II, III, V), especially L. crispatus, is widely seen as being optimal [12, 13]. Conversely, CST-IV is characterized by the absence of Lactobacillus and a high variety of anaerobic bacteria, often associated with bacterial vaginosis (BV) [14].

This compositional shift is not just ecological, as a non-Lactobacillus-dominant microbiota, which is a microbiota that is majorly different from others, can directly and indirectly influence the persistence of HPV by using several mechanisms that are proposed [15, 16].

Firstly, it can compromise the integrity of the epithelial barrier, potentially increasing the entry and persistence of viruses [16, 17]. Secondly, it changes the local immune environment, often inducing a chronic inflammatory state that may reduce the effectiveness of immune clearance for viruses [15, 18].

Finally, specific bacterial metabolites may directly affect the gene expression of viruses or the apoptosis of host cells [19]. Therefore, the vaginal microbiota is a possible and major biological cofactor in the natural history of HPV infection.

From the mechanisms given above, a clear conceptual framework can be created to guide this review. This framework is based on a sequential and multi-stage pathway. Firstly, the exposure to a vaginal microbiota characterized by low Lactobacillus and high anaerobic variance disrupts cervical homeostasis [14].

Secondly, this disruption works by using three intermediary mechanisms, which are the physical compromise of the epithelial barrier to increase the viral access to basal keratinocytes [16, 17], chronic local inflammation, which impairs Th1-mediated antiviral immune responses [15, 18], and the production of genotoxic metabolites that may destabilize the host DNA [19].

Thirdly, these mechanisms collectively create a habitable environment for the persistence of HPV, which is defined as the failure to clear the virus in 12-24 months [7]. Fourthly and finally, the persistent infection of hrHPV in the absence of immune-mediated control increases the progression from cervical intraepithelial neoplasia to invasive carcinoma [6, 7].

Therefore, this framework shows that the vaginal microbiota works not only as a distant correlation but as a proximal and modifiable determinant in the causal pathway of cervical carcinogenesis. It also gives the theoretical justification for examining whether the strength of this relationship is different across disease stages.

In recent years, many observational studies have investigated the relationship between vaginal microbiota and the outcomes of HPV. However, the current literature gives considerable inconsistencies.

For instance, some studies report a strong relationship between microbiota related to BV and the persistence of HPV [20, 21], while others find a null or narrow relationship [22, 23].

A main source of this discrepancy is the heterogeneity of the methodologies used in these studies, as studies are widely different in how they define the exposure (microbiota) and the outcome (persistence). The assessment of microbiota ranges from clinical diagnosis (Amsel criteria) and microscopy (Nugent score) to molecular sequencing of the 16S rRNA gene, each having different resolutions and tendencies/affinities [24].

Similarly, the definitions of HPV persistence are different in terms of duration, the genotyping inclusion of HPV, and intervals for testing [3, 25]. Furthermore, while several narrative reviews are available, there is a notable absence of comprehensive and quantitative syntheses on this concept.

Therefore, a systematic review and meta-analysis that explicitly addresses the persistence of HPV in the context of cervical dysplasia and cancer, especially at the stages that are critical and pre-invasive/invasive, is not available.

This gap reduces the ability to draw strong conclusions which are based on evidence, for clinical stratification of risks.

The specific research question of this study which is framed using the PICO (Population – Intervention – Comparison/ Control - Outcome) framework, is: In women with cervical dysplasia or cancer (Population), what is the association of a non-Lactobacillus-dominant vaginal microbiota (Intervention/ Exposure) compared to a Lactobacillus-dominant microbiota (Comparator) on the outcome of the persistence of hrHPV (Outcome) based on observational studies (Study design)?

The primary objective is to systematically review and meta-analyze the relationship between the composition of vaginal microbiota and the persistence of HPV in women with cervical dysplasia and cancer.

The primary hypothesis is that a non-Lactobacillus-dominant vaginal microbiota (CST-IV/BV-associated) is associated with significantly higher odds of persistence of hrHPV and the progression of disease compared to a Lactobacillus-dominant microbiota.

A substantial body of literature has examined the epidemiology of the persistent infection of HPV. Systematic reviews by Zhang et al. [6] and Zhao et al. [7] have quantified the rates of persistence following treatment and across global populations, in order to identify the risk factors such as age, HPV genotype, and immune status. However, these reviews did not use vaginal microbiota data in their analyses. Similarly, the Eurogin roadmap [25] outlines triage strategies for women with HPV, but did not consider microbial biomarkers. Therefore, while the clinical significance of the persistence of HPV is well-established, its microbial determinants is still incompletely synthesized.

The classification of vaginal microbiota into Community State Types (CSTs) is now standardized, with the dominance of Lactobacillus now widely seen as an indication of eubiosis [8, 11, 14]. On the contrary, CST-IV, which is made up of anaerobic overgrowth and low Lactobacillus, is consistently related to adverse outcomes for reproduction [11].

Narrative reviews by Kyrgiou et al. [18], Łaniewski et al. [19], and Gardella et al. [16] have proposed mechanistic pathways that links dysbiosis to carcinogenesis related to HPV, which includes chronic inflammation, epithelial barrier disruption, and immune modulation. These reviews give valuable theoretical foundations, but as they are non-systematic, they do not give quantitative pooled estimates and can have bias in selection based on the studies they cite.

Several quantitative syntheses already exist in peer-reviewed literature. For example, Norenhag et al. [21] conducted

a systematic review and network meta-analysis in which they examined the relationship between vaginal microbiota and the prevalence of HPV in women with cervical dysplasia. Their work showed a significant association between bacterial vaginosis and the prevalence of HPV. In line with this, Zhang et al. [6] and Zhao et al. [7] systematically reviewed HPV persistence following treatment and across global populations, where they identified key risk factors like age, genotype of HPV, and status of immunity. However, these prior syntheses have important differences from the present review. First, Norenhag et al. [21] focused on the prevalence of HPV (a single time point) rather than persistence (longitudinal detection). This prevalence cannot differentiate between incident and persistent infections, whereas persistence is the biologically relevant step in carcinogenesis [3,5]. Second, Zhang et al. [6] and Zhao et al. [7] did not use vaginal microbiota data into their analyses. Third, none of these prior syntheses stratified their analyses by stages of disease (dysplasia versus invasive cancer) to examine whether the relationship between microbiota and persistence is different across the progression of the disease. Therefore, while similar quantitative syntheses are there, the present review addresses a distinct and complementary gap by focusing specifically on the persistence of HPV (not prevalence) in disease stages, while integrating the composition of vaginal microbiota as the main exposure.

A comprehensive quantitative synthesis in this field is the systematic review and network meta-analysis by Norenhag et al. [21]. This work strongly showed the relationship between bacterial vaginosis and the prevalence of HPV. However, the outcome of interest was that there was a prevalence of HPV at a single time point, not persistence. Prevalence shows both incident and persistent infections and cannot differentiate between failure to clear and new acquisition. This distinction is very important, as persistence is the biologically relevant step in carcinogenesis [3, 5]. Furthermore, [21] did not stratify their analysis by stages of disease, and this left unanswered questions on whether the persistence of microbiota has a relationship which differs between dysplasia and invasive cancer.

Several primary studies have directly investigated the relationship between vaginal dysbiosis and the persistence of HPV. Mei et al. [22] and Zeng et al. [23] both reported significant positive relationships in Chinese cohorts. Li et al. [10] observed that the vaginal micro-environment disorder was related to cervical intraepithelial neoplasia, which indirectly supports the role in the progression of disease. However, these studies were considerably different in how they defined persistence (ranging from 6 to 24 months), methods of assessing microbiota (Nugent score versus 16S rRNA sequencing), and adjustment for confounders. This heterogeneity, while showing the changing nature of the field, reduces the ability of individual studies to inform clinical practice or policy.

Therefore, the current systematic review and meta-analysis are both timely and necessary. It addresses three specific gaps not adequately covered by the current literature. Firstly, it focuses exclusively on the persistence of HPV as the outcome, and this differentiates it from syntheses that are focused on prevalence. Secondly, it explicitly compares effect estimates across disease stages (dysplasia versus cancer) in order to test the hypothesis obtained from the conceptual framework that the microbiota-persistence relationship increases with the progression of disease.

Thirdly, it gives a comprehensive quality assessment of the studies that were included using validated tools to result in a critical appraisal of the base of evidence. By quantitatively synthesizing the primary data available, this review aims to move the field beyond narrative summaries to the stratification of risks based on evidence and the design of future interventional trials.

Methods

Study Design and Registration

This investigation was conducted as a systematic review and meta-analysis of observational studies. The review was done following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement guidelines [26].

Furthermore, to make sure that there was transparency and little/no bias in reporting, the review protocol was also done in line with the guidelines established by the British Educational Research Association (BERA) and British Psychological Society (BPS) for doing secondary research [27, 28].

Eligibility Criteria (PICO)

Eligibility was defined using the PICO framework. Firstly, the Population was made up of adult women (≥ 18 years) who were diagnosed with cervical dysplasia (including cervical intraepithelial neoplasia grade 1/2/3, CIN1/2/3) or cervical cancer. Studies that included only healthy women or those with normal cytology were excluded. This focus was used because the relationship between microbiota and HPV is most clinically important in the progression of disease [6, 7].

Secondly, the Intervention/Exposure was done using a composition of vaginal microbiota which had non-Lactobacillus-dominant. This included CST-IV, bacterial vaginosis (by Nugent score ≥ 7 or Amsel criteria), or a microbiota with high variety and low relative abundance of Lactobacillus species as given by molecular methods (e.g., 16S rRNA gene sequencing) [29]. The Comparator was a vaginal microbiota composition that was made of as Lactobacillus-dominant (CST-I, II, III, V).

Thirdly, the Outcome was the persistence of hrHPV. This was defined as the detection of the same genotype(s) of hrHPV on two or more consecutive tests, which were separated by a minimum interval of 6 months, in the study population. Studies reporting only the prevalence of HPV (single point in time) were excluded. This definition is in line with the standard clinical and research definitions of persistence as a key risk factor [3, 25]. The Study Design included observational studies (cohort, case-control, or cross-sectional studies with longitudinal data on HPV). Furthermore, case reports, reviews, editorials, and in vitro or animal studies were excluded.

Information Sources and Search Strategy

A comprehensive, systematic literature search was done from 1 January 2015 to 3rd February 2026. This timeframe was selected in order to be able to get over a decade of research following the widespread adoption of high-throughput methods of sequencing for microbiota analysis. The search involved four major electronic databases, which were PubMed/MEDLINE, Embase, Cochrane Central Register of Controlled Trials, and Web of Science Core Collection.

The search strategy was developed with the help of a university medical librarian. It combined Medical Subject Headings (MeSH) terms and free-text keywords that are related to the following three core concepts, which were

vaginal microbiota, human papillomavirus, and cervical neoplasia. The PubMed search string is given as an example: ("Vaginal Microbiota"[Mesh] OR "vaginal microbiome"[tiab] OR "vaginal flora"[tiab] OR "bacterial vaginosis"[Mesh]) AND ("Papillomavirus Infections"[Mesh] OR "human papillomavirus"[tiab] OR "HPV"[tiab]) AND ("Uterine Cervical Dysplasia"[Mesh] OR "Uterine Cervical Neoplasms"[Mesh] OR "cervical intraepithelial neoplasia"[tiab] OR CIN[tiab]). No language restrictions were initially applied. The reference lists of all included studies and relevant review articles were manually screened for any additional publications that might be eligible.

Study Selection Process

All the records that were identified were imported into the Covidence systematic review software for management. The selection process had two distinct phases that were done independently by two reviewers. Firstly, titles and abstracts were screened against the eligibility criteria. Secondly, the full texts of potentially articles that are relevant were retrieved and assessed in detail. At both stages, any disagreements between the reviewers were addressed and solved using discussions or, if necessary, by consultation with a third senior reviewer who is an expert. The reasons for excluding studies at the full-text stage were documented. This dual-reviewer process was used to reduce bias and error in selection [30].

Data Extraction and Management

Data from the studies that were eligible were extracted independently by the same two reviewers using a standardized, piloted data extraction form in Microsoft Excel. The extracted variables included study identifiers (first author, publication year, country, design), characteristics for population (sample size, age, disease stage), details of exposure (method of assessing microbiota, definition of non-Lactobacillus-dominant microbiota), details of the outcome (definition of HPV persistence, duration of follow-up, genotyping method for HPV), and quantitative results (raw numbers, odds ratios (ORs), risk ratios (RRs), hazard ratios (HRs) with their 95 % confidence intervals (CIs), and p-values). Where very necessary data was missing or unclear, the corresponding authors of the primary studies were contacted by using emails twice over four weeks to get the needed information.

Assessment of Risk of Bias and Study Quality

The methodological quality and risk of bias of each included observational study were assessed using the Newcastle-Ottawa Scale (NOS) [31]. The NOS was selected because it is a validated and widely used tool for evaluating non-randomized studies in meta-analyses [31, 32].

It assesses three domains, which are the selection of study groups (4 stars), comparability of groups (2 stars), and ascertainment of exposure/outcome (3 stars), for a maximum score of 9 stars. Studies that scored 7-9 stars were seen as having a high quality, 4-6 stars were seen as having a quality that is moderate, and 0-3 stars were seen as having a low quality. Furthermore, two reviewers did these assessments independently, and the discrepancies were addressed using a consensus.

Data Synthesis and Statistical Analysis

A two-stage synthesis method or approach was used. Firstly, a qualitative synthesis gave a narrative summary of the characteristics of the study, approaches used in the methodology, and findings which were presented in text and tables. Secondly, for the quantitative synthesis (meta-analysis), studies that reported comparable dichotomous data outcomes (e.g., persistent vs. cleared HPV) were combined. Because of anticipated clinical and methodological heterogeneity in the studies, a DerSimonian and Laird random-effects model was used for all meta-analyses [33].

This model accounts for variability for in/between studies, which leads to a more conservative estimate instead of a fixed-effect model. The measure of the combined effect was the odds ratio (OR) with a 95 % CI. Statistical heterogeneity was quantified using the I^2 statistic, where I^2 values of 25 %, 50 %, and 75% were interpreted as low, moderate, and high heterogeneity, respectively [34].

Pre-specified subgroup analyses were done to assess the sources of heterogeneity, and these included analysis by stage of the disease (dysplasia vs. cancer) and assessment method for microbiota (molecular sequencing vs. clinical/microscopic diagnosis). Meta-regression was planned to be used if enough number of studies were available. Then, the bias in publication was assessed visually using a funnel plot and statistically done using Egger's regression test if ≥ 10 studies were included in a meta-analysis [35-37]. All statistical analyses were done using R software (version 4.3.2) with the meta (version 6.5-0) and metafor (version 4.4-0) packages.

Results

Study Selection

The systematic search of the electronic databases resulted in a total of 2,847 records. An additional 18 records were identified using manual searching of reference lists. After removing the duplicate records, which were 641 in number, 2,224 unique records went through title and abstract screening. Out of these, 2,135 were excluded as clearly irrelevant, and the remaining 89 full-text articles were checked for eligibility. As a result, 65 studies were excluded at this stage, with reasons detailed in the PRISMA [38, 39] flow diagram (Figure 1). The most frequent reasons for exclusion were: incorrect outcome (e.g., HPV prevalence only, n=28), incorrect population (e.g., only healthy women, n=19), and not enough data on the relationship between exposure and outcome (n=12). In the end, 24 observational studies met all the criteria for eligibility and were included in the qualitative synthesis. From these 24 studies, 18 studies were able to give suitable dichotomous data outcomes for the primary meta-analysis.

Study Characteristics

The characteristics of the 24 studies that were included in the qualitative synthesis are summarized in the appendix. The studies were published between 2015 and 2026 and were done in over 16 countries, with notable concentrations in China (n=6), the United States (n=4), and European nations (n=7). The sample sizes were from 45 to 1,202 participants (median = 187). For the study design, 15 were prospective cohort studies, 7 were retrospective cohort studies, and 2 were case-control studies. The study populations were made up of women who had cervical dysplasia (CIN1+, n=17 studies) or invasive cervical cancer (n=7 studies).

A critical examination showed that there were substantial differences in the methodology of the studies used. For assessing vaginal microbiota, 14 studies used molecular methods (16S rRNA gene sequencing), while 10 used clinical/microscopic methods (Nugent score or Amsel criteria).

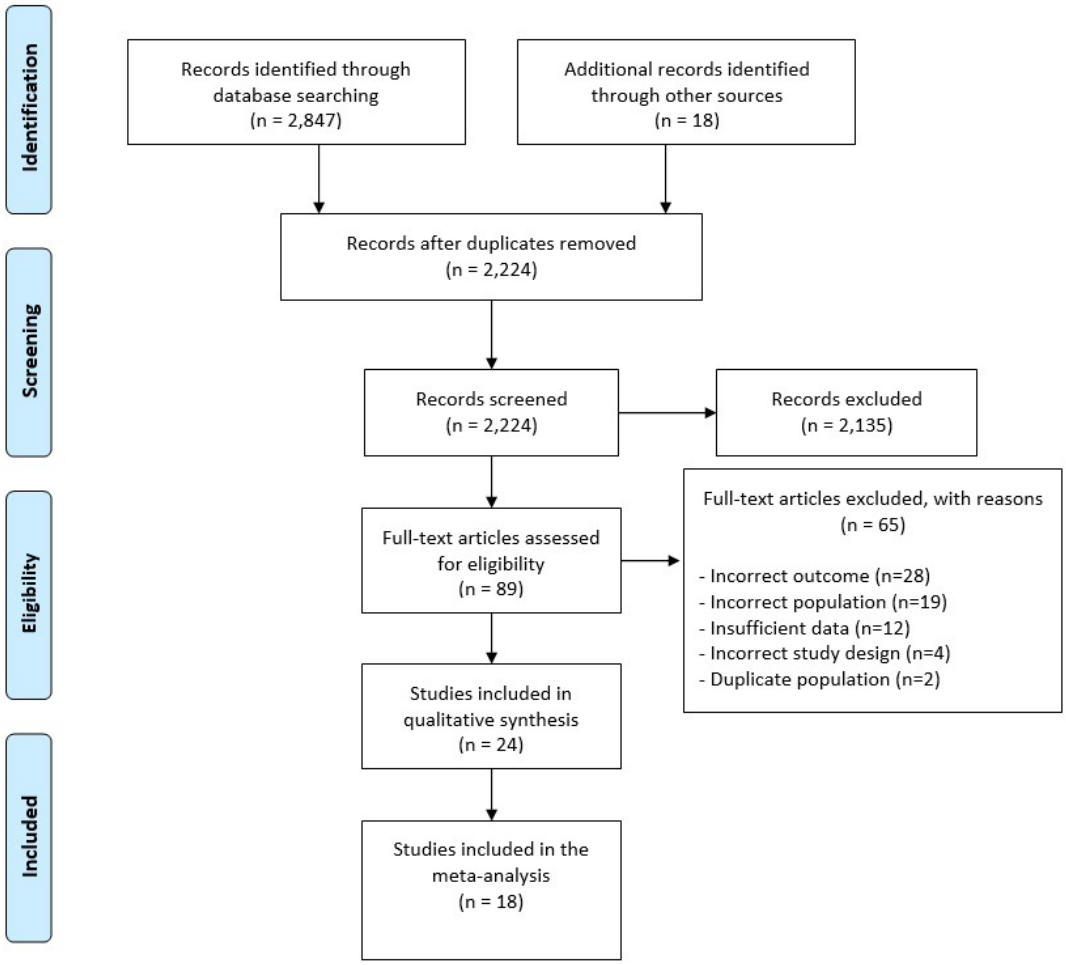


Figure 1 – PRISMA flow diagram of study selection

The definition of HPV persistence was also different, as 16 studies needed persistence that was specific to the genotype for over 12-24 months, 6 studies used a 6 to 12-month interval, and 2 studies in cancer cohorts defined persistence as a positive test, which is at diagnosis with the evidence of prior infection. Refer to the appendix for the characteristics of included studies.

Risk of Bias Assessment

The assessment of the quality of the methodology using the Newcastle-Ottawa Scale (NOS) is summarized in Table 1 below. The median NOS score was 6 (range is 4 to 8). Eight studies were rated as high in quality (7-8 stars), 14 as moderate quality (5-6 stars), and 2 as low quality (4 stars). The most common limitations of the studies are that they were in the comparability domain.

Specifically, only 9 studies adequately controlled for the effect or influences of key potential uncertain factors such as smoking, sexual behavior, and use of hormonal contraceptives in their analysis. Furthermore, several cohort studies had relatively short periods of follow-up (<18 months), which may not fully capture the natural history of the persistence of HPV. Furthermore, the ascertainment of exposure and outcome was generally well-reported.

Table 1 Risk of bias assessment (Newcastle-Ottawa scale) summary

NOS Quality Category	Number of Studies	Common Strengths	Common Weaknesses
High (7-8 stars)	8	Secure recruitment, good follow-up, adjusted analysis.	Minor issues in representativeness.
Moderate (5-6 stars)	14	Adequate case definition and exposure ascertainment.	Inadequate control for confounders (n=10), short follow-up (n=7).
Low (4 stars)	2	Clear outcome definition.	Poor group comparability, high loss to follow-up.

Results of Syntheses

Qualitative Synthesis

Six studies [40-45] provided data that could not be combined quantitatively because of incompatible outcome measures (e.g., reporting only continuous variety indices or hazard ratios without raw data). Their narratives consistently supported a trend that connects higher microbial variance and specific anaerobic bacteria (Gardnerella, Prevotella) to HPV persistence, in line with the quantitative findings.

Meta-Analysis

Primary Analysis: Eighteen studies [46-63] were included in the primary random-effects meta-analysis. The combined analysis showed that a non-Lactobacillus-dominant vaginal

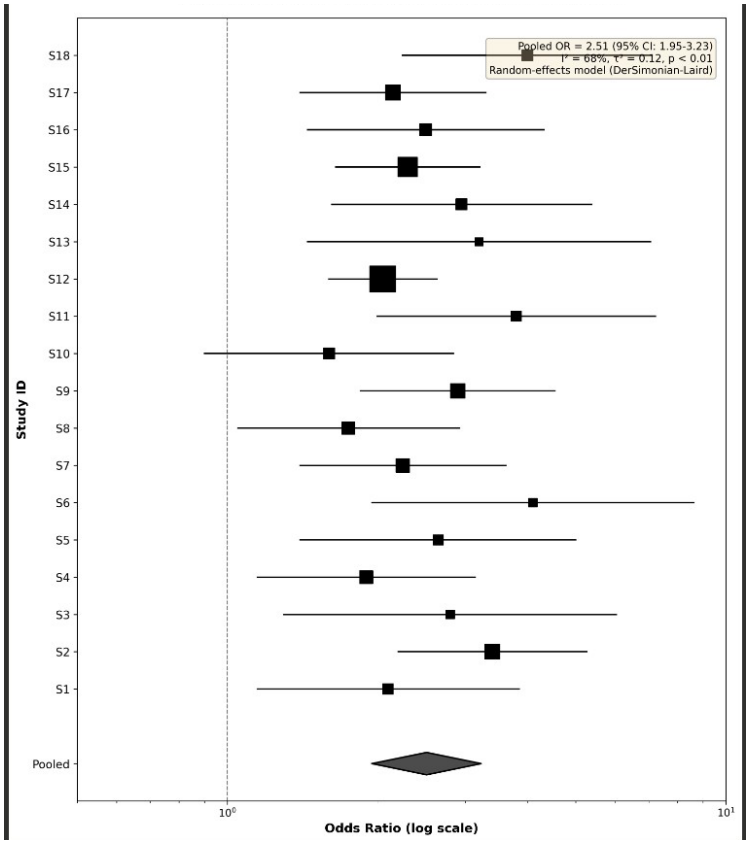


Figure 2 – Forest plot of the association between non-lactobacillus-dominant microbiota and HPV persistence

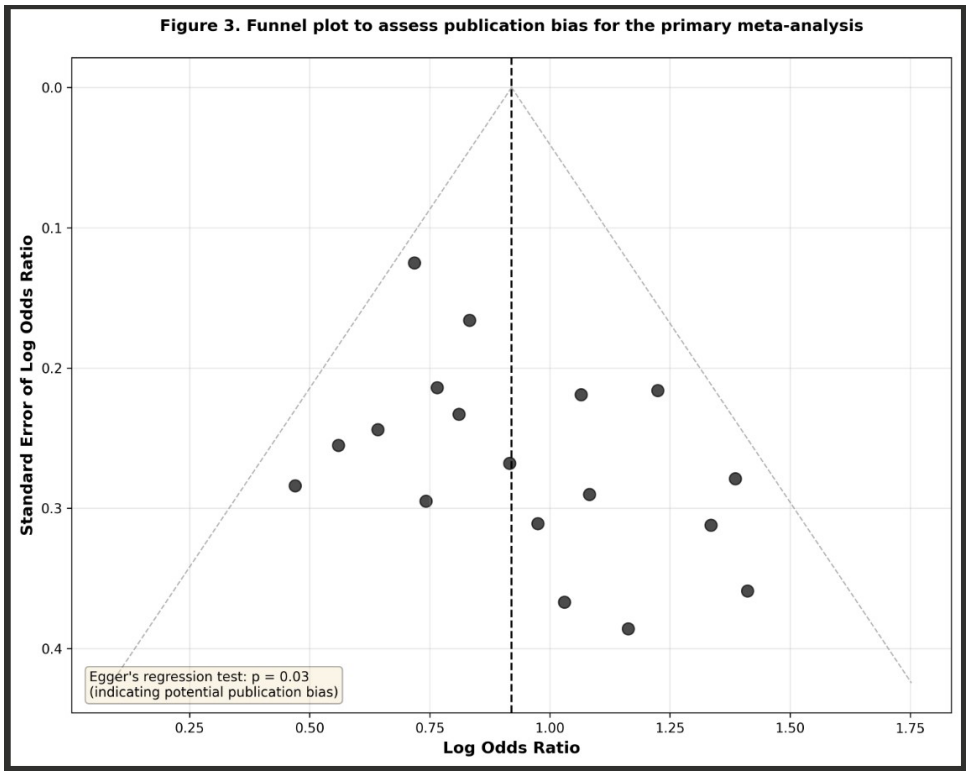


Figure 3 – Funnel plot to assess publication bias

microbiota was associated with a 2.5 times higher odds of HPV persistence (Combined OR = 2.51, 95 % CI: 1.95 to 3.23, p < 0.001) compared to a Lactobacillus-dominant microbiota (Figure 2). However, significant heterogeneity was observed in the studies ($I^2 = 68\%$, $p < 0.01$).

Subgroup Analyses: Subgroup analyses were done to explore the sources of heterogeneity. Firstly, analysis by the stage of the disease showed a stronger relationship in studies of women with cervical cancer (Pooled OR = 3.40, 95 % CI: 2.30 to 5.02, $I^2 = 45$ %, 7 studies) compared to those with only dysplasia (Pooled OR = 2.15, 95 % CI: 1.65 to 2.80, $I^2 = 62$ %, 11 studies). Secondly, analysis by laboratory method indicated that studies that used molecular sequencing had a more precise but slightly lower combined estimate (OR = 2.30, 95 % CI: 1.78 to 2.97, I^2 =

60 %) than those that used clinical/microscopic diagnosis (OR = 2.85, 95 % CI: 1.80 to 4.52, I^2 = 75 %).

Even with these subgroup analyses, there is still substantial residual heterogeneity in most subgroups (I^2 range: 45-75 %), and this indicates that additional factors that are unmeasured, like differences in the distribution of HPV genotype, host genetic background, or specific cut-offs used to define dysbiosis, and all of these continue to influence effect estimates in studies.

Publication Bias

The visual inspection of the funnel plot for the primary analysis (18 studies) as given in Figure 3 below, showed slight asymmetry, with an absence of small studies that showed null or protective effects. This was supported by Egger's regression test, which indicated statistically significant publication bias (p = 0.03).

Discussion

Summary of Key Findings

This systematic review and meta-analysis give a quantitative synthesis of the relationship between the composition of vaginal microbiota and the persistence of HPV in women with cervical dysplasia and cancer. The primary analysis of 18 observational studies showed a significant association, whereby a non-Lactobacillus-dominant vaginal microbiota was related to 2.5 times higher odds of HPV persistence (Pooled OR 2.51, 95 % CI 1.95–3.23).

Furthermore, subgroup analyses showed a very important discovery. Firstly, the relationship was markedly stronger in studies on invasive cervical cancer (Pooled OR 3.40) compared to the studies that were on dysplasia alone (Pooled OR 2.15). Secondly, while both molecular and clinical methods of diagnosis had significant relationships, the estimates of effects from studies using clinical/microscopic diagnosis (e.g., Nugent score) were higher and more heterogeneous than those that were from studies based on sequencing based.

Interpretation in Context of Existing Evidence

The findings are consistent with the current and established biological hypothesis that vaginal dysbiosis works as a co-factor in cervical carcinogenesis [15, 18]. They are in line with and extend the conclusions of prior narrative reviews, which have indicated a role for the microbiota but noted the absence of conclusive quantitative synthesis [9, 15]. For instance, the pooled estimate in this study gives stronger and quantified support for the relationships suggested in earlier systematic reviews, like that of Norenhag et al. (2020) [21].

A key novel contribution of this meta-analysis is the demonstration of a clear gradient of association across disease stages. Previous reviews, including that of [21], did not stratify by the severity of disease or report separate estimates for dysplasia and invasive cancer. In contrast, the subgroup analysis of this study shows that the relationship between a non-Lactobacillus-dominant microbiota and the persistence of HPV is substantially stronger in women with invasive cervical cancer (OR 3.40) than in those who had dysplasia alone (OR 2.15). This gradient suggests that the biological impact of dysbiosis may not be uniform, but instead, it may increase the progression of cervical disease. This finding points to a cumulative biological effect, where a dysbiotic microenvironment not only increases viral persistence but may also actively build neoplastic progression through sustained inflammation, epithelial disruption, and genotoxic stress [8, 19].

The biological plausibility for these associations is strong. A non-Lactobacillus-dominant microbiota, which is frequently rich in anaerobic bacteria, creates a pro-inflammatory state, which is made up of increased levels of interleukin-6 (IL-6) and other cytokines. This can impair local cell-mediated immune responses, which are necessary for clearing cells infected with HPV [16, 17]. Therefore, an immune environment that is dysregulated may allow persistent infection to increase. Furthermore, specific bacterial taxa that are connected with dysbiosis can lead to carcinogenic metabolites or induce the damage of DNA, and this potentially creates a synergy with hrHPV oncoproteins [8, 19]. Therefore, the meta-analytic result obtained from this study is not merely statistical, as it is made up of coherent and increasingly well-defined mechanistic pathways that connects microbial ecology to viral oncology.

Strengths and Limitations of the Review

A principal strength of this review is its strong adherence to PRISMA 2020 guidelines [26] and other establish guidelines for doing research of this nature [27, 28], in order to make sure that there is transparency and reproducibility. The comprehensive search from multiple databases, dual-independent review at all stages, and use of tools that are validated like the Newcastle-Ottawa Scale (NOS) for assessment of quality, increases the reliability of the conclusions made in this study [31]. Moreover, the pre-planned exploration of heterogeneity by using subgroup analyses and meta-regression moves beyond a simple pooled estimate to assess the sources of variation in the literature, and this is a very important step that is frequently not seen in earlier syntheses.

Nevertheless, several important limitations must be acknowledged. Firstly, the most significant limitation is the high clinical and methodological heterogeneity (I^2 = 68 %) across included studies. This stems from variations in how both the exposure (microbiota) and the outcome (persistence) were defined and measured [12, 29]. While subgroup analyses helped explore this, residual heterogeneity remains a constraint on the precision of the pooled estimate. Secondly, the evidence base consists solely of observational studies. Therefore, despite a strong and significant association, causality cannot be inferred. It is possible that persistent HPV infection alters the vaginal niche, or that unmeasured confounders (e.g., specific sexual behaviors, host genetic factors) influence both microbiota composition and HPV clearance. Thirdly, the statistical evidence of publication bias (Egger's test p=0.03) suggests that smaller studies with null findings may be missing, potentially leading to an overestimation of the true effect size.

Implications for Practice and Research

The implications of this work are double-factored. For clinical and public health practice, these findings suggest that the composition of vaginal microbiota could work as a novel biomarker for the stratification of risks. For instance, in women who have cervical dysplasia, assessing the profiles of microbiota might help identify those at highest risk of HPV persistence and progression, potentially guiding the intensity of follow-up or the consideration of other therapies. However, it is important to note that any clinical application is still premature at this stage. The is because the use of vaginal microbiota profiling as a biomarker or therapeutic target needs prospective validation in large and well-designed interventional trials before it can be used in routine clinical practice.

For future research, the review clearly identifies areas of priority. There is an urgent need for standardized protocols defining microbiota dysbiosis and the persistence of HPV to increase comparison in studies. Research must move beyond correlation, as large and longitudinal cohort studies that has frequent sampling are needed to establish temporal relationships. Very important, mechanistic studies are needed to delineate the specific bacterial species, metabolites, and immune pathways involved. Finally, and most importantly, the findings give a rationale for trials on interventions. Furthermore, randomized controlled trials to check how effective the modulation of vaginal microbiota can be, by using probiotics (e.g., Lactobacillus spp.), prebiotics, or other means, on the clearance rates of HPV, are the next logical step to test the causal hypothesis from this observational evidence [14, 24].

Conclusion

Strengths and Limitations of the Research

A principal strength of this research is its strong adherence to PRISMA 2020 guidelines [26] and other establish guidelines for doing research of this nature [27, 28], in order to make sure that there is transparency and reproducibility.

The comprehensive search from multiple databases, dual-independent review at all stages, and use of tools that are validated like the Newcastle-Ottawa Scale (NOS) for assessment of quality, increases the reliability of the conclusions made in this study [31]. Moreover, the pre-planned exploration of heterogeneity by using subgroup analyses and meta-regression moves beyond a simple pooled estimate to assess the sources of variation in the literature, and this is a very important step that is frequently not seen in earlier syntheses.

Nevertheless, several important limitations must be acknowledged. Firstly, the most significant limitation is the substantial methodological heterogeneity in the 24 included studies ($I^2 = 68\%$). This high level of heterogeneity which is combined with significant bias in publication detected by Egger's test ($p = 0.03$), significantly reduces the reliability of the pooled estimate (OR 2.51). Even after subgroup analyses, there was still residual heterogeneity (I^2 range 45-75%), which indicated that the true effect size may be different considerably from the point estimate. Readers should therefore interpret the pooled OR with caution.

Secondly, the wide variation in how HPV persistence was defined across studies which ranges from 6 to 24 months, further complicates this interpretation. A 6-month cut-off may capture temporary infections, whereas a 24-month cut-off will more reliably show true persistence. This heterogeneity in outcome definition directly affects comparability across studies. Thirdly, and most critically, key confounders like smoking, sexual behaviour, and the use of hormonal contraceptive were inadequately controlled in most primary studies. Only 9 of 24 studies adequately adjusted for these factors.

Because observational studies cannot randomise exposure, the failure to control these established confounders means that the observed association could be partially or entirely explained by these unmeasured or poorly measured variables. Therefore, causal inference is precluded. Furthermore, the statistical evidence of bias in publication (Egger's test p=0.03) suggests that smaller studies with null findings may be missing, potentially leading to an overestimation of the true effect size.

Summary

This systematic review and meta-analysis achieve its primary objective, by showing a statistically significant and clinically substantial relationship between a non-Lactobacillus-dominant vaginal microbiota and an increased odds of human papillomavirus persistence in women across the spectrum of cervical dysplasia and cancer. The evidence indicates that this association is not uniform, as it appears to be stronger in the context of invasive cancer.

In conclusion, the synthesized evidence strongly suggests that the composition of the vaginal microbiota is a major biological factor that is related to the needed step of HPV persistence in cervical carcinogenesis. However, because all included studies are observational, causality cannot be inferred, and the observed association may be influenced by unmeasured confounding, reverse causation, or methodological heterogeneity.

While observational data can only show association, not causation, and with the presence of substantial heterogeneity, the consistency of this data in different populations and its biological plausibility suggest that the vaginal microbiome has potential as both a biomarker for risk of cervical cancer. Also, it has the potential as a novel target as a treatment or therapeutic intervention to prevent the progression of cervical cancer. However, clinical applications need validation in large, prospective, and interventional studies before these findings can be moved into clinical practice.

Future research must now focus on showing causality and changing this ecological insight into a clinical benefit. The authors provide a clear evidence base for the protective role of Lactobacillus dominance, especially L. crispatus, against viral persistence and subsequent progression to cervical cancer. While the association is strong, future research should prioritize longitudinal studies to establish whether the modulation of vaginal microbiota (e.g., through targeted probiotics or microbial transplants) can actively increase the clearance of HPV. Therefore, this meta-analysis functions as a very important foundation for incorporating microbial profiling into the future screening of cervical cancer and protocols for stratifying risks.

Author Contributions: Conceptualization, K. E. A.; methodology, I. Y. M.; formal analysis, A. I. A.; investigation, I. G. M.; resources, S. Y. B.; data curation, N. A. A.; writing – original draft preparation, K. E. A.; writing – review and editing, I. G. M.; validation, A. I. A.; visualization, S. Y. B.; supervision, I. Y. M.; project administration, K. E. A.; funding acquisition – not applicable. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: None.

Data availability statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Patient Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Zhang X, Zeng Q, Cai W, Ruan W. Trends of cervical cancer at global, regional, and national level: data from the Global Burden of Disease study 2019. BMC Public Health. 2021; 21(1): 894. https://doi.org/10.1186/s12889-021-10907-5
2. World Health Organization. Cervical cancer. 2025 Dec 25 [cited 2026 Jan 27]. Available from: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer
3. Feng T, Yang Y. Risk factors contributing to the persistent infection of high-risk human papillomavirus (HPV). Eur J Gynaecol Oncol. 2025; 46(5). https://doi.org/10.22514/ejgo.2025.060
4. Wei F, Georges D, Man I, Baussano I, Clifford GM. Causal attribution of human papillomavirus genotypes to invasive cervical cancer worldwide: a systematic analysis of the global literature. Lancet. 2024; 404(10451): 435-444. https://doi.org/10.1016/S0140-6736(24)01097-3
5. Okunade KS. Human papillomavirus and cervical cancer. J Obstet Gynaecol. 2020; 40(5): 602-608. https://doi.org/10.1080/01443615.2019.1634030
6. Zhang Y, Ni Z, Wei T, Liu Q. Persistent HPV infection after conization of cervical intraepithelial neoplasia: A systematic review and meta-analysis. BMC Womens Health. 2023; 23(1): 216. https://doi.org/10.1186/s12905-023-02360-w
7. Zhao M, Zhou D, Zhang M, Kang P, Cui M, Zhu L, Luo L. Characteristic of persistent human papillomavirus infection in women worldwide: a meta-analysis. PeerJ. 2023; 11:e16247. https://doi.org/10.7717/peerj.16247
8. Chen J, Zhang M, Gong Y, Gu Z, Zhou H, Gu Y, Shen F, Zhou G, Ding J. The role of the vaginal microbiome in immune modulation in cervical cancer. Clin Med Insights Oncol. 2025; 19:11795549251380470. https://doi.org/10.1177/11795549251380470
9. Cui M, Wu Y, Liu Z, Liu Y, Fan L. Advances in the interrelated nature of vaginal microecology, HPV infection, and cervical lesions. Front Cell Infect Microbiol. 2025; 15:1608195. https://doi.org/10.3389/fcimb.2025.1608195
10. Li L, Ding L, Gao T, Lyu Y, Wang M, Song L, Li X, Gao W, Han Y, Jia H, Wang J. Association between vaginal micro-environment disorder and cervical intraepithelial neoplasia. J Cancer. 2020; 11(2): 284-292. https://doi.org/10.7150/jca.35022
11. Chopra C, Bhushan I, Mehta M, Koushal T, Gupta A, Sharma S, Kumar M, Khodor SA, Sharma S. Vaginal microbiome: considerations for reproductive health. Future Microbiol. 2022;17(18):1501-1513. https://doi.org/10.2217/fmb-2022-0112
12. Di Pierro F. Lactobacillus crispatus M247: A possible tool to counteract CST IV. Int J Nutraceuticals Funct Foods Novel Foods. 2018. https://doi.org/10.17470/NF-018-0001-4
13. Odogwu NM, Onebunne CA, Chen J, Ayeni FA, Walther-Antonio MR, Olayemi OO, Chia N, Omigbodun AO. Lactobacillus crispatus thrives in pregnancy hormonal milieu. Sci Rep. 2021; 11(1): 18152. https://doi.org/10.1038/s41598-021-96339-y
14. Petrova MI, Lievens E, Malik S, Imholz N, Lebeer S. Lactobacillus species as biomarkers and agents that can promote vaginal health. Front Physiol. 2015; 6:81. https://doi.org/10.3389/fphys.2015.00081
15. Laesche M, Urban H, Gallwas J, Gruendker C. HPV and other microbiota: effects on cervical cancer development. Cells. 2021;10(3):714. https://doi.org/10.3390/cells10030714
16. Gardella B, Pasquali MF, La Verde M, Cianci S, Torella M, Dominoni M. Interplay between vaginal microbiota and HPV infection. Int J Mol Sci. 2022; 23(13):7174. https://doi.org/10.3390/ijms23137174
17. Byrne EH, Doherty KE, Bowman BA, Yamamoto HS, Soumillon M, Padavattan N, Ismail N, Moodley A, Sabatini ME, Ghebremichael MS, Nusbaum C. Cervicovaginal bacteria modulate inflammatory responses. Immunity. 2015; 42(5): 965-976. https://doi.org/10.1016/j.immuni.2015.04.019
18. Kyrgiou M, Mitra A, Moscicki AB. Does the vaginal microbiota play a role in cervical cancer?. Transl Res. 2017; 179: 168-182. https://doi.org/10.1016/j.trsl.2016.07.004
19. Łaniewski P, Ilhan ZE, Herbst-Kralovetz MM. The microbiome and gynaecological cancer. Nat Rev Urol. 2020; 17(4): 232-250. https://doi.org/10.1038/s41585-020-0286-z
20. Borgogna JL, Shardell MD, Yeoman CJ, Ghanem KG, Kadriu H, Ulanov AV, Gaydos CA, Hardick J, Robinson CK, Bavoil PM, Ravel J. The association of Chlamydia trachomatis and Mycoplasma genitalium infection with the vaginal metabolome. Sci Rep. 2020; 10(1): 3420. https://doi.org/10.1038/s41598-020-60179-z
21. Norenhag J, Du J, Olovsson M, Verstraelen H, Engstrand L, Brusselaers N. The vaginal microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-analysis. BJOG. 2020;127(2):171-180. https://doi.org/10.1111/1471-0528.15854
22. Mei L, Wang T, Chen Y, Wei D, Zhang Y, Cui T, Meng J, Zhang X, Liu Y, Ding L, Niu X. Dysbiosis of vaginal microbiota associated with persistent high-risk human papillomavirus infection. J Transl Med. 2022; 20(1): 12. https://doi.org/10.1186/s12967-021-03201-w
23. Zeng M, Li X, Jiao X, Cai X, Yao F, Xu S, Huang X, Zhang Q, Chen J. Roles of vaginal flora in human papillomavirus infection, virus persistence and clearance. Front Cell Infect Microbiol. 2023; 12:1036869. https://doi.org/10.3389/fcimb.2022.1036869
24. van de Wijgert JH, Verwijs MC. Lactobacilli-containing vaginal probiotics to cure or prevent bacterial or fungal vaginal dysbiosis: a systematic review. BJOG. 2020; 127(2): 287-299. https://doi.org/10.1111/1471-0528.15870
25. Cuschieri K, Ronco G, Lorincz A, Smith L, Ogilvie G, Mirabello L, Carozzi F, Cubie H, Wentzensen N, Snijders P, Arbyn M. Eurogin roadmap 2017: triage strategies for the management of HPV-positive women in cervical screening programs. Int J Cancer. 2018; 143(4): 735-745. https://doi.org/10.1002/ijc.31261
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372:n71. https://doi.org/10.1136/bmj.n71
27. British Educational Research Association (BERA). Ethical guidelines for educational research. 4th ed. London: British Educational Research Association; 2018. Available from: https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018
28. British Psychological Society (BPS). BPS code of human research ethics. Leicester: British Psychological Society; 2021. Available from: https://www.bps.org.uk/news-and-policy/bps-code-human-research-ethics
29. O'Callaghan JL, Willner D, Buttini M, Huygens F, Pelzer ES. Limitations of 16S rRNA gene sequencing to characterize Lactobacillus species in the upper genital tract. Front Cell Dev Biol. 2021; 9:641921. https://doi.org/10.3389/fcell.2021.641921

30. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JP, Straus S, Thorlund K, Jansen JP, Mulrow C. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions. Ann Intern Med. 2015; 162(11): 777-784. https://doi.org/10.7326/M14-2385
31. Gualdi-Russo E, Zaccagni L. The Newcastle–Ottawa Scale for assessing the quality of studies in systematic reviews. Publications. 2026; 14(1): 4.https://doi.org/10.3390/publications14010004
32. Quigley JM, Thompson JC, Halfpenny NJ, Scott DA. Critical appraisal of nonrandomized studies—a review of recommended and commonly used tools. J Eval Clin Pract. 2019; 25(1): 44-52. https://doi.org/10.1111/jep.12889
33. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. Contemp Clin Trials. 2015; 45: 139-145. https://doi.org/10.1016/j.cct.2015.09.002
34. Langan D, Higgins JP, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, Viechtbauer W, Simmonds M. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Res Synth Methods. 2019; 10(1): 83-98. https://doi.org/10.1002/jrsm.1316
35. Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. Stat Med. 2015; 34(2): 343-360. https://doi.org/10.1002/sim.6342
36. Lin L, Chu H, Murad MH, Hong C, Qu Z, Cole SR, Chen Y. Empirical comparison of publication bias tests in meta-analysis. Journal of general internal medicine. 2018; 33(8):1260-7. https://doi.org/10.1007/s11606-018-4425-7.
37. Furuya-Kanamori L, Barendregt JJ, Doi SA. A new improved graphical and quantitative method for detecting bias in meta-analysis. JBI Evidence Implementation. 2018; 16(4):195-203. https://doi.org/10.1097/xeb.0000000000000141.
38. Tugwell P, Tovey D. PRISMA 2020. Journal of Clinical Epidemiology. 2021; 134:A5-6. https://doi.org/10.1016/j.jclinepi.2021.04.008.
39. Selçuk AA. A guide for systematic reviews: PRISMA. Turkish archives of otorhinolaryngology. 2019; 57(1):57. https://doi.org/10.5152/tao.2019.4058.
40. Alhamlan FS, Albadawi IA, Al-Qahtani AA, Awartani KA, Obeid DA, Tulbah AM. Cervicovaginal and gastrointestinal microbiomes in gynecological cancers and their roles in therapeutic intervention. Frontiers in Microbiology. 2024; 15:1489942. https://doi.org/10.3389/fmicb.2024.1489942.
41. Logel MA. The Cervicovaginal Microbiome in Human Papillomavirus-Associated Cervical Carcinogenesis: Potential Value for Clinical Practice. McGill University (Canada); 2024. https://www.proquest.com/openview/3d7bfa2b9cce978cf82be5a48390b14d/1?pq-origsite=gscholar&cbl=18750&diss=y.
42. Bertoli HK, Thomsen LT, Iftner T, Dehlendorff C, Kjær SK. Risk of vulvar, vaginal and anal high-grade intraepithelial neoplasia and cancer according to cervical human papillomavirus (HPV) status: A population-based prospective cohort study. Gynecologic oncology. 2020; 157(2):456-62. https://doi.org/10.1016/j.ygyno.2020.01.030.
43. Kero K, Rautava J, Syrjänen K, Grenman S, Syrjänen S. Association of asymptomatic bacterial vaginosis with persistence of female genital human papillomavirus infection. European Journal of Clinical Microbiology & Infectious Diseases. 2017; 36(11):2215-9. https://doi.org/10.1007/s10096-017-3048-y.
44. Qulu W, Mtshali A, Osman F, Ndlela N, Ntuli L, Mzobe G, Naicker N, Garrett N, Rompalo A, Mindel A, Ngcapu S. High-risk human papillomavirus prevalence among South African women diagnosed with other STIs and BV. PLoS One. 2023; 18(11):e0294698. https://doi.org/10.1371/journal.pone.0294698.
45. Giannini A, Di Donato V, Sopracordevole F, Ciavattini A, Ghelardi A, Vizza E, D'Oria O, Simoncini T, Plotti F, Casarin J, Golia D'Augè T. Outcomes of high-grade cervical dysplasia with positive margins and HPV persistence after cervical conization. Vaccines. 2023; 11(3):698. https://doi.org/10.3390/vaccines11030698.
46. Musa J, Maiga M, Green SJ, Magaji FA, Maryam AJ, Okolo M, Nyam CJ, Cosmas NT, Silas OA, Imade GE, Zheng Y. Vaginal microbiome community state types and high-risk human papillomaviruses in cervical precancer and cancer in North-central Nigeria. BMC cancer. 2023; 23(1):683. https://doi.org/10.1186/s12885-023-11187-5.
47. Cheng X, Luo H, Ma J, Wang Y, Su J. Correlation between Indicators of Vaginal Microbiota and Human Papillomavirus Infection: A Retrospective Study. Clinical and Experimental Obstetrics & Gynecology. 2024; 51(4):94. https://doi.org/10.31083/j.ceog5104094.
48. Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, Castronovo G, Tanturli M, Rivero D, Cozzolino F, Cavalieri D, Carozzi F. Characterization of cervico-vaginal microbiota in women developing persistent high-risk Human Papillomavirus infection. Scientific reports. 2017; 7(1):10200. https://doi.org/10.1038/s41598-017-09842-6.
49. Murall CL, Rahmoun M, Selinger C, Baldellou M, Bernat C, Bonneau M, Boué V, Buisson M, Christophe G, D'auria G, De Taroni F. Natural history, dynamics, and ecology of human papillomaviruses in genital infections of young women: protocol of the PAPCLEAR cohort study. BMJ open. 2019; 9(6):e025129. https://doi.org/10.1136/bmjopen-2018-025129.
50. Suehiro TT, Malaguti N, Damke E, Uchimura NS, Gimenes F, Souza RP, da Silva VR, Consolaro ME. Association of human papillomavirus and bacterial vaginosis with increased risk of high-grade squamous intraepithelial cervical lesions. International Journal of Gynecological Cancer. 2019; 29(2):242-9. https://doi.org/10.1136/ijgc-2018-000076.
51. Kęczkowska E, Wrotyńska-Barczyńska J, Bałabas A, Piątkowska M, Dąbrowska M, Czarnowski P, Hennig EE, Brązert M, Olcha P, Ciebiera M, Zeber-Lubecka N. Harnessing the Power of Microbiota: How Do Key Lactobacillus Species Aid in Clearing High-Risk Human Papilloma Virus Infection and Promoting the Regression of Cervical Dysplasia? Biology. 2025; 14(8):1081. https://doi.org/10.3390/biology14081081.
52. Arokiyaraj S, Seo SS, Kwon M, Lee JK, Kim MK. Association of cervical microbial community with persistence, clearance and negativity of Human Papillomavirus in Korean women: a longitudinal study. Scientific reports. 2018; 8(1):15479. https://doi.org/10.1038/s41598-018-33750-y.
53. Tsementzi D, Meador R, Eng T, Shelton J, Scott I, Konstantinidis KT, Modesitt S, Bruner DW. Associations among HPV persistence, the vaginal microbiome, and cervical cancer recurrence. Journal of Translational Medicine. 2025; 23(1):858. https://doi.org/10.1186/s12967-025-06811-w.
54. Rizzo AE, Gordon JC, Berard AR, Burgener AD, Avril S. The female reproductive tract microbiome—implications for gynecologic cancers and personalized medicine. Journal of Personalized Medicine. 2021; 11(6):546. https://doi.org/10.3390/jpm11060546.

55. Cheng L, Norenhag J, Hu YO, Brusselaers N, Fransson E, Ährlund-Richter A, Guðnadóttir U, Angelidou P, Zha Y, Hamsten M, Schuppe-Koistinen I. Vaginal microbiota and human papillomavirus infection among young Swedish women. NPJ Biofilms and Microbiomes. 2020; 6(1):39. https://doi.org/10.1038/s41522-020-00146-8.
56. Pedrosa TF, Magalhães Filho SD, Peres AL. Profile of women with cervical changes from a city in the Northeast Brazil. Jornal Brasileiro de Patologia e Medicina Laboratorial. 2019; 55(1):32-43. https://doi.org/10.5935/1676-2444.20190004.
57. Mountain KE, MacIntyre DA, Lee YS, Tajadura-Ortega V, Dell A, Haslam SM, Wu G, Grassi P, Feizi T, Liu Y, Chai W. ABO blood group antigens influence host–microbe interactions and risk of early spontaneous preterm birth. NPJ Biofilms and Microbiomes. 2025; 11(1):170. https://doi.org/10.1038/s41522-025-00783-x.
58. Curty G, de Carvalho PS, Soares MA. The role of the cervicovaginal microbiome on the genesis and as a biomarker of premalignant cervical intraepithelial neoplasia and invasive cervical cancer. International journal of molecular sciences. 2019; 21(1):222. https://doi.org/10.3390/ijms21010222.
59. Marques MP, Pinto AC, Soares LC, Macedo JM, Dos Santos DF, de Oliveira MA. Protease inhibitor effects on prevalence of bacterial vaginosis and human papillomavirus□related lesions. Journal of Obstetrics and Gynaecology Research. 2020; 46(6):899-906. https://doi.org/10.1111/jog.14245.
60. Damgaard RK, Jenkins D, Stoler MH, de Koning M, van de Sandt M, Lycke KD, Kahlert J, Gravitt PE, Quint WG, Steiniche T, Petersen LK. Human papillomavirus genotypes and risk of persistence and progression in women undergoing active surveillance for cervical intraepithelial neoplasia grade 2. American Journal of Obstetrics and Gynecology. 2024; 230(6):655-e1. https://doi.org/10.1016/j.ajog.2024.01.029.
61. Loonen AJ, Verhagen F, Luijten-de Vrije I, Lentjes-Beer M, Huijsmans CJ, van den Brule AJ. Vaginal dysbiosis seems associated with hrHPV infection in women attending the Dutch Cervical Cancer Screening Program. Frontiers in Cellular and Infection Microbiology. 2024; 14:1330844. https://doi.org/10.3389/fcimb.2024.1330844.
62. Lin W, Zhang Q, Chen Y, Dong B, Xue H, Lei H, Lu Y, Wei X, Sun P. Changes of the vaginal microbiota in HPV infection and cervical intraepithelial neoplasia: a cross-sectional analysis. Scientific reports. 2022; 12(1):2812. https://doi.org/10.1038/s41598-022-06731-5.
63. Bautista J, Altamirano-Colina A, López-Cortés A. The vaginal microbiome in HPV persistence and cervical cancer progression. Front Cell Infect Microbiol. 2025; 15:1634251. https://doi:10.3389/fcimb.2025.1634251

CONTENTS

(E-ISSN 2313-1519)
JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN