Discover Oncology

Analysis

Relationship between red cell distribution width and lung cancer: evidence from Mendelian randomization and National Health and Nutrition Examination Survey

Yongli Liu[1] · Jiajia Qu[1] · Chenyang Hu[1] · Wei Zhao[1] · Yuxin Zhang[1] · Yuchen Luo[1] · Yiqing Qu[1]

Received: 27 December 2024 / Accepted: 15 May 2025
Published online: 22 May 2025

Abstract

Background Lung cancer remains a primary contributor to cancer-related mortality globally. Red blood cell distribution width (RDW), a straightforward and cost-effective indicator, measures the variability in red blood cell size and is conventionally employed in hematological assessments for anemia differentiation. Nonetheless, limited research has explored the causal link between RDW levels and lung cancer incidence.

Methods Initially, Mendelian randomization (MR) was employed to explore the underlying causal connection between RDW and lung cancer. To ensure the robustness of the MR findings, sensitivity analyses were conducted. Following this, the National Health and Nutrition Examination Survey (NHANES) database was utilized to further substantiate the influence of RDW on the prognosis of lung cancer.

Results The MR analysis revealed a significant association between RDW and lung cancer risk in the European population (OR IVW 1.11, 95% CI 1.03–1.20, $p = 0.006$; OR Weighted-median 1.16, 95% CI 1.03–1.31, $p = 0.013$; OR MR-Egger 1.14, 95% CI 1.00–1.30, $p = 0.059$). Furthermore, findings from the NHANES database suggested that lower RDW values are associated with improved prognosis in lung cancer patients (HR 2, 95% CI 1.07–3.74, $p < 0.05$).

Conclusions Our study provides further evidence for the relationship between RDW levels and lung cancer, highlighting the potential significance of RDW as a biomarker for predicting lung cancer risk and prognosis.

1 Introduction

Lung cancer is the primary cause of cancer-related mortality worldwide. In 2020, it contributed to roughly 20% of the 1.9 million cancer-related deaths worldwide[1]. RDW is a common parameter in routine blood tests for most hospital patients. Although RDW is associated with various diseases and can predict adverse outcomes such as cardiovascular diseases and gastrointestinal tumors [2, 3], there is a lack of research investigating the causal relationship between RDW levels and lung cancer incidence.

MR analysis harnesses the intrinsic characteristics of common genetic variants and their correlation with environmental exposures, establishing itself as a robust approach to exploring potential causal links between modifiable environmental factors and disease outcomes [4, 5]. This method is particularly useful for assessing causality between specific exposures and outcomes.However, MR studies often necessitate large sample sizes to discern the effects of genetic variants, which can be constrained by data availability [6]. As a potent epidemiological tool, Mendelian randomization enables causal

✉ Yiqing Qu, quyiqing@sdu.edu.cn | [1]Department of Pulmonary and Critical Care Medicine, Qilu Hospital of Shandong University, Wenhuaxi Road 107#, Jinan 250012, China.

inference by mitigating confounding bias and circumventing reverse causation. While it has its limitations, the ongoing advancements in genomic technologies and the accumulation of data suggest that it will play an increasingly pivotal role in future medical research endeavors. Utilizing transcriptome data and summary statistics from previous genome-wide association studies (GWAS), we aim to conduct an MR analysis to explore the potential causal link between RDW values and lung cancer onset. Furthermore, we corroborate the predictive value of RDW in adverse survival outcomes among lung cancer patients using data from the NHANES database.

2 Objectives

Leveraging the abundant transcriptome data and summary statistics from previous GWAS, we aim to delve into the potential causal relationship between RDW values and lung cancer incidence through the advanced method of Mendelian randomization analysis. We anticipate that this study will offer new perspectives and evidence for the early prevention, risk assessment, and clinical treatment of lung cancer. Additionally, to further validate the predictive value of RDW in the context of lung cancer, we will utilize actual data from the NHANES database to analyze the specific performance of RDW in predicting adverse survival outcomes among lung cancer patients, thereby providing more comprehensive information to support the clinical management and prognostic evaluation of lung cancer patients.

3 Mendelian randomization analysis

3.1 Data sources

Our study examined the influence of single nucleotide polymorphisms (SNPs) linked to RDW on lung cancer by analyzing genome-wide association study (GWAS) data from individuals of European ancestry and lung cancer data from the Finn Gen consortium R11 (https://r11.finngen.fi/?.) All data are listed in Table 1. All participants were of European descent. The study reported 5820 cases of non-small cell lung cancer (NSCLC), 7497 cases of bronchial and lung cancer, and 1787 cases of squamous cell carcinoma, a subtype of NSCLC. This valuable dataset provides robust support for our research. This study utilized publicly accessible online databases, specifically MR and NHANES data. The National Center for Health Statistics (NCHS) conducts NHANES under the Centers for Disease Control and Prevention (CDC). This population-based cross-sectional survey integrates interviews and physical examinations to assess the health and nutritional status of adults and children in the United States. Lung cancer samples from NHANES surveys (1999–2018) were analyzed to investigate the association between RDW and poor prognosis in lung cancer patients.

Given that our research relies solely on these previously published and openly available summary statistics, ethical approval was not necessary.

4 Methods

4.1 Instrumental variable

The MR method relies on three fundamental assumptions for causal inference.

(1) It is necessary to identify genetic variants that are strongly correlated with the exposure variable of interest, which will serve as instrumental variables.
(2) These instrumental variables must satisfy the three basic assumptions of MR, including the relevance assumption (the instrumental variable is correlated with the exposure variable), the exclusion restriction (the instrumental variable is not correlated with any potential confounding factors), and the independence assumption (the instrumental variable can only influence the outcome variable through the exposure variable).

Table 1 Details of GWASs analyzed in the present MR analyses

Phenotype	Number of cases	Number of controls	Sample size	Number of variants	Ethnicity	Year	Trait ID in GWAS
Exposure							
Red cell distribution	NA	NA	350,473	13,586,288	European	2018	ukb-d-30070_irnt
Outcomes							
Non-small cell lung cancer	5820	345,118	350,938	20,092,484	European	2024	finngen_R11_C3_LUNG_NONSMALL_EXALLC
Malignant neoplasm of bronchus and lung	7497	345,118	352,615	20,092,541	European	2024	finngen_R11_C3_BRONCHUS_LUNG_EXALLC
Non-small lung cancer,squamous	1787	345,118	346,905	20,092,358	European	2024	finngen_R11_C3_NSCLC_SQUAM_EXALLC
Smoking initiation	311,629	321,173	607,291	11,802,365	European	2019	ieu-b-4877

(3) Once valid instrumental variables are identified, they can be used to estimate the causal effect of the exposure variable on the outcome variable. This typically involves statistical models and methods such as two-stage least squares, inverse variance weighting, weighted median, model selection and other approaches.

We chose suitable instrumental variables from the GWAS results for MR. We initially selected SNPs with p-values below the genome-wide significance threshold of $5 \times 10\text{^}8$. Subsequently, to control for the effects of linkage disequilibrium, we set thresholds of r^2 greater than 0.001 and a distance within 10000 kb. We harmonized the exposure and outcome data by excluding palindromic SNPs with intermediate allele frequencies to maintain data accuracy. We evaluated the selected instruments'strength using the F-statistic, considering an F-value below 10 as indicative of insufficient instrument strength. The F-statistic of 147.15 in our study indicates the strong efficacy of the selected instruments.

4.2 MR estimates

We utilized various techniques to measure the causal relationship between the exposure and the outcome.We utilized Inverse-Variance Weighted (IVW) regression, the Weighted Median method, and MR-Egger regression to evaluate the causal effect of the exposure factor on the outcome.These models rigorously addressed potential confounders, including smoking to maintain the objectivity of the effect estimates. To ensure the robustness of our findings across various statistical models and assumptions, we performed sensitivity analyses, such as Leave-one-out analysis and the MR-intercept method (Fig. 1). These analyses not only helped us identify potential weaknesses in the MR analysis but also revealed any potential biases or uncertainties.

4.3 Relationship between RDW and lung cancer from NHANES

To assess the ability of RDW to anticipate all-cause mortality of the patients with lung cancer, we employed KM analysis to estimate survival probabilities and visualize the data. Additionally, the Cox proportional hazards regression model was used to identify potential risk factors associated with mortality. The subgroup analyses were conducted to investigate the impact across different populations, aiming to fully understand the significance of this association in disease prevention and treatment.

The RDW were divided into binary variable based on RDW values. Classification variables were expressed as frequencies (percentages) and sequential ones as mean ± standard deviation (SD). The Kruskal–Wallis test (for non-normal distributions), one-way ANOVA (for normal distributions), and chi-square test (for categorical variables) were used to compare different RDW groups. Cox regression analysis was employed to assess the independent effect of RDW levels on mortality. To enhance the accuracy of the analysis, we adjusted for factors such as age, gender, race, hemoglobin, hematocrit, platelet count, red blood cell distribution width, eosinophils number,WBC count,segmented neutrophils number.

Three models were developed based on these influencing factors:

Model I: unadjusted,

Model II: adjusted for race, age, and gender,

Model III: further adjusted for the variables in Model II plus additional factors identified through univariate and multivariate regression ($P < 0.05$), including RDW, gender, WBC count, military status, segmented neutrophils number, monocyte number.

Finally, subgroup analyses were conducted according to Models III, examining the interactions of RDW levels within different subgroups categorized by gender, military Status, Emphysema,Chronic bronchitis, Congestive heart failure, Coronary heart disease, Liver condition, Education Level.

4.4 Statistical methods

MR analyses utilized the'Two Sample MR'package. We computed odds ratios (ORs), hazard ratios (HRs), and their 95% confidence intervals (CIs). A p-value below 0.05 was deemed statistically significant.

R 4.3.0 statistical analysis software was used in this study, and the rank-sum test for two independent samples was used; categorical data were expressed as n (%), and the chi-square test or Fisher's exact probability method was used, and the difference was considered statistically significant at $P < 0.05$ for comparisons between groups.

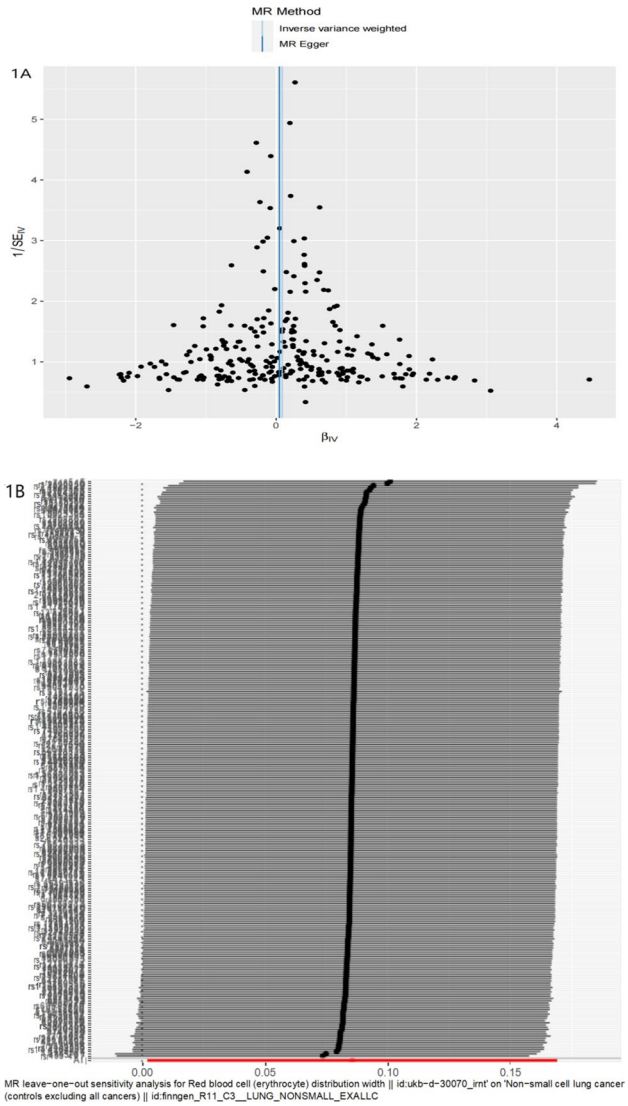


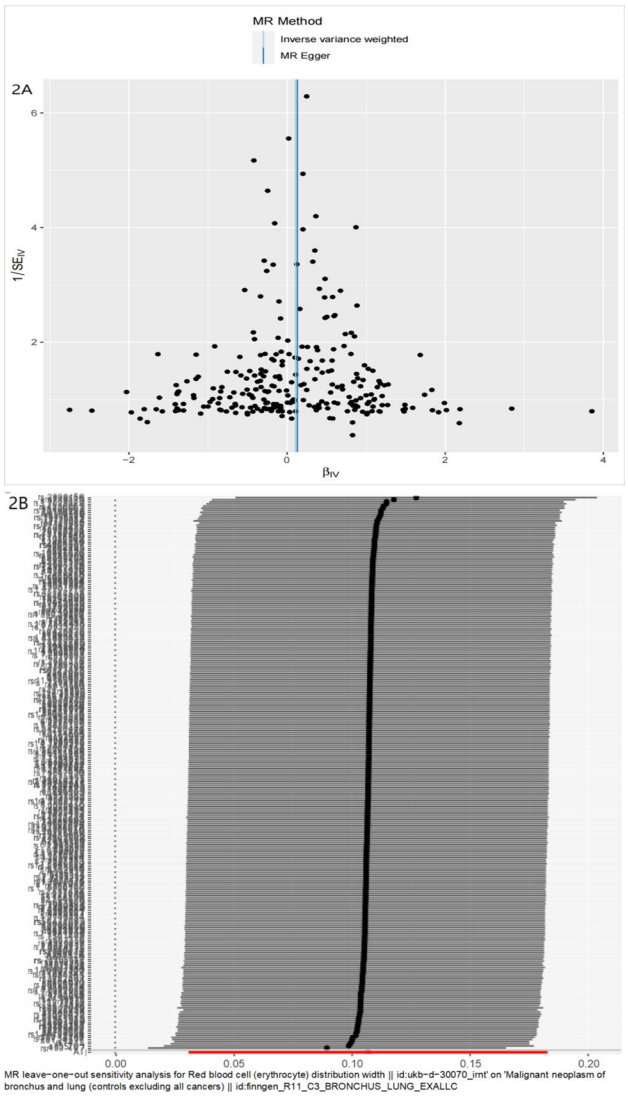
Fig. 1 MR sensitivity analysis **A**: The funnel plot of Mendelian randomization analysis; **B**: The forest plot of leave-one-out analysis. 1-finngen_R11_C3_LUNG_NONSMALL_EXALLC 2-finngen_R11_C3_BRONCHUS_LUNG_EXALLC. 3-finngen_R11_C3_NSCLC_SQUAM_EXALLC

5 Results

Our MR analysis reveals that elevated RDW is linked to a higher lung cancer risk, supported by odds ratios: IVW (OR 1.11, 95% CI 1.03–1.20, p = 0.006), Weighted-median (OR 1.16, 95% CI 1.03–1.31, p = 0.013), and MR-Egger (OR 1.14, 95% CI 1.00–1.30, p = 0.06). We identified a heightened risk of squamous cell lung cancer, indicated by odds ratios: OR IVW 1.19 (95% CI 1.02–1.40, p = 0.029), OR Weighted-median 1.05 (95% CI 0.82–1.34, p = 0.69), and OR MR-Egger 1.08 (95% CI 0.81–1.43, p = 0.60). For non-small cell lung cancer, significant differences were found between two methods: OR IVW 1.09 (95% CI 1.00–1.18, p = 0.04), OR Weighted-median 1.17 (95% CI 1.01–1.34, p = 0.03) and OR MR-Egger 1.05 (95% CI 0.90–1.21, p = 0.54). The leave-one-out analysis (Fig. 1) indicates that the MR results are stable and not influenced by any single SNP. The forest plot and scatter plot also suggest that there is a stable relationship between RDW and lung cancer (Fig. 2).

This study further corroborates the findings of the initial MR analysis through sensitivity analysis and ensures their alignment with the three fundamental MR assumptions outlined in the "4" section. The results of the sensitivity analysis are presented in Table 2. During the study, we first screened 261 SNPs significantly associated with RDW and lung cancer, using a p-value threshold of less than 5.0 × 10^8. Detailed data are provided in Table 1. Using the Leave-one-out validation method, we found that none of the 261 SNPs significantly influenced the association between RDW and lung cancer, including squamous cell carcinoma and non-small cell lung cancer(Fig. 1). Furthermore, the

Fig. 1 (continued)

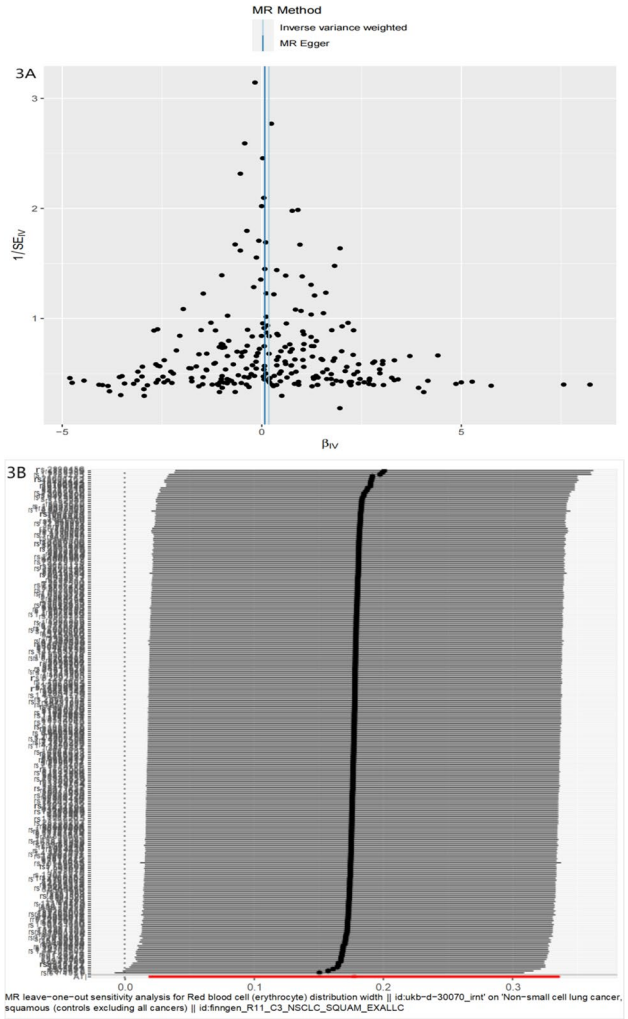


results of the MR-Egger intercept regression analysis also suggested the absence of notable horizontal pleiotropy in the study, thereby confirming our adherence to the second MR assumption. Our data analysis supports the third assumption, showing no significant causal link between RDW and smoking (OR = 1.00, 95% CI 0.98–1.02, p = 0.88). In summary, through sensitivity analysis, we successfully validated the effectiveness of the genetic tool comprising 261 SNPs related to RDW. Our study shows that the relationship between the exposure factor and the outcome is not influenced by other potential risk factors.

Ultimately, the findings from the NHANES database further corroborated the results of the MR study. From 1999 to 2018, NHANES collected data on 138 lung cancer patients. After excluding 31 adult patients with missing RDW values, data from 107 patients were analyzed(Fig. 3). Survival analysis indicated that elevated RDW values correlate with a poor prognosis in lung cancer patients (HR = 1.74, 95% CI 1.03–2.93, p = 0.037) (Fig. 4). The majority of patients (64.49%) were Non-Hispanic White. The mean age of the patients was 71 (64.50, 78.50) years and 59.81% were male. Overall mean RDW was 13.70 (12.90, 14.60). Patients with high RDW values were with increased hematocrit, WBC count, hemoglobin, Segmented neutrophils number. No statistically significant relationship was seen between RDW values and social status and educational attainment. There was a borderline statistical difference in Military Status (Table 3).

To identify the independent factor influencing RDW in relation to all-cause mortality in patients with lung cancer, three distinct Cox regression models were meticulously constructed, the results of which are detailed in Table 4. As the RDW decreased, the HR values across all three models exhibited an upward trend. In Model I, a lower RDW was found to correlate with an increased incidence of all-cause mortality among patients with lung cancer. Similarly, in Model II, a lower RDW was linked to all-cause mortality in patients with lung cancer, following adjustments for age, sex, and race.

Fig. 1 (continued)



In the study, we conducted univariate and multivariate Cox regression analyses on patients with lung cancer to identify risk factors associated with mortality. We selected variables with a p-value of less than 0.05 in the univariate analysis and incorporated them into the multivariate Cox regression analysis to determine independent risk factors. In model III, the RDW emerged as an independent predictor for mortality in this population, even after adjusting for additional confounding variables (adjusted hazard ratios and 95% CI: 2.00(1.07 ~ 3.74), with p-values less than 0.05). These findings imply that levels of RDW in patients with lung cancer are independently associated with all-cause mortality, making it a significant risk factor(Table 4).Subgroup analyses were performed to assess the impact of gender,coexisting disease,military status and educational level, with comprehensive results available in Fig. 5. The analyses showed no significant association between these factors and survival rates ($p > 0.05$).

6 Discussions

This study utilized a two-sample MR method to examine the causal link between elevated RDW and heightened risk of lung cancer, including squamous cell carcinoma and non-small cell lung cancer. Previous research has indicated that RDW can predict poor prognosis in lung cancer [7, 8], yet there remains a dearth of exploration into the causal relationship between RDW and lung cancer. Our study explores the causal relationship between RDW and lung cancer, providing a new perspective and screening tool for identifying high-risk individuals. Utilizing comprehensive NHANES data, we have validated RDW's significant role in lung cancer prognosis, offering new insights for future clinical practice advancements.

Our research has expanded upon existing foundations, not only by incorporating key confounding variables such as smoking habits into the comprehensive analysis, but also by delving deeply into the complex association between

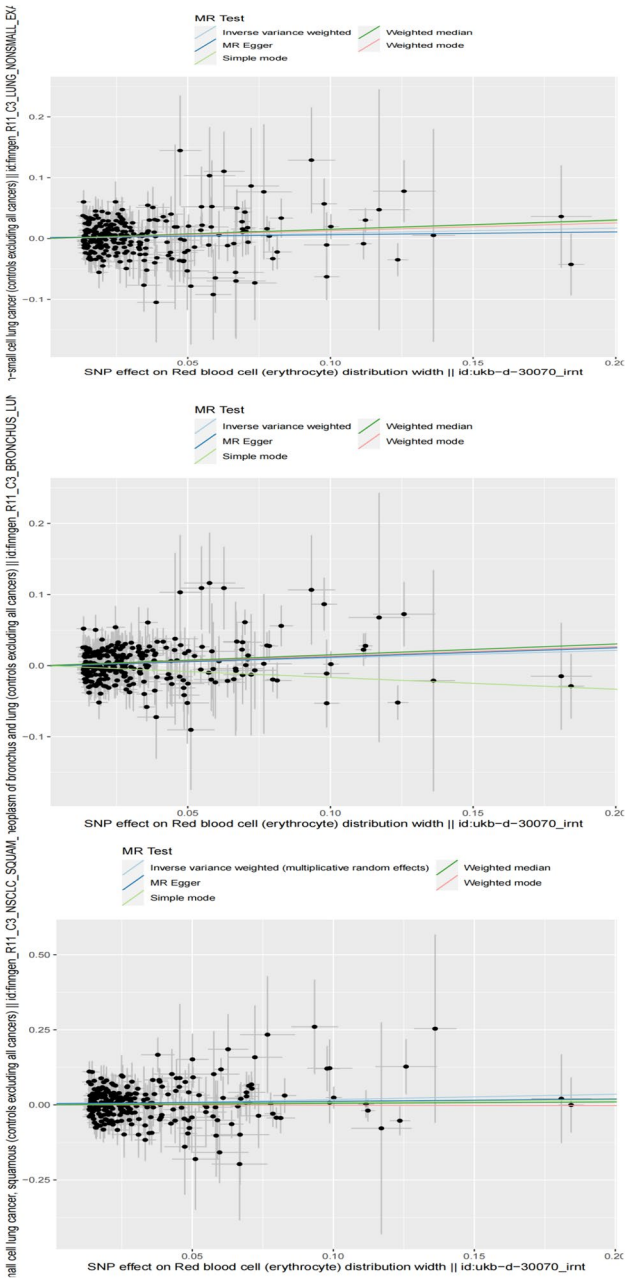


Fig. 2 Two-sample Mendelian randomization (MR) Each point represents the SNP effects on RDW and Lung cancer. Colored lines represent inverse-variance-weighted (red), weighted median (green), and MR-Egger (blue) estimates of the association between a 1-SD increase in RDW and risk of Lung cancer. A-finngen_R11_C3_LUNG_NONSMALL_EXALLC. B-finngen_R11_C3_BRONCHUS_LUNG_EXALLC. C-finngen_R11_C3_NSCLC_SQUAM_EXALLC

RDW levels and lung cancer susceptibility. The findings significantly reveal the potential value of RDW in lung cancer prevention strategies and early diagnosis, further highlighting the central role of this biomarker in the management and intervention measures for lung cancer.

RDW is correlated with multiple types of cancer. Elevated RDW levels are linked to a heightened risk of several cancers, including colon, breast, lung and pancreatic cancers.Studies on RDW and cancer: Tumor stage and invasiveness: Several studies have found that elevated RDW is associated with advanced cancer (e.g., colorectal cancer, lung cancer) stage, lymph node metastasis, and distant metastasis (possibly reflecting increased tumor load). In lung cancer: elevated RDW is independently associated with shorter overall survival (OS). elevated RDW can be seen in a variety of diseases such as anemia, infection, heart failure, etc., and confounding factors need to be ruled out in conjunction with other indicators. Blood transfusion, iron supplementation or anti-inflammatory treatment can change RDW in the short term, and dynamic monitoring is needed.The prognostic value of RDW may be more significant in hematologic tumors (e.g., lymphoma) and less so in solid tumors (e.g., thyroid cancer).RDW, as a low-cost and easily accessible indicator, has demonstrated its usefulness in prognostic stratification and efficacy monitoring of cancers, but its application should be monitored

Table 2 Mendelian randomization analysis of the probabilistic association between RDW levels and lung cancer

Outcome	Ethnicity	nSNP	IVW method		Weighted-median		MR-Egger	
			OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value
Malignant neoplasm of bronchus and lung	European	261	1.11(1.03–1.20)	0.006	1.16(1.03–1.31)	0.013	1.14(1.00–1.30)	0.060
Non-small cell lung cancer	European	261	1.09(1.00–1.18)	0.040	1.17(1.01–1.34)	0.030	1.05(0.90–1.21)	0.540
Non-small cell lung cancer, squamous	European	261	1.19(1.02–1.40)	0.029	1.05(0.82–1.34)	0.690	1.08(0.81–1.43)	0.600
Smoking initiation	European	251	1.00(0.98–1.02)	0.884	1.00(0.98–1.03)	0.612	0.99(0.96–1.03)	0.766

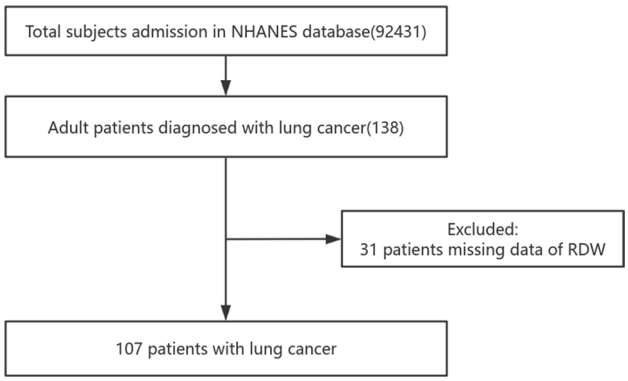


Fig. 3 Flow chart of the population included in the study from the NHANES. NHANES-National Health and Nutrition Examination Survey; RDW-Red cell Distribution Width

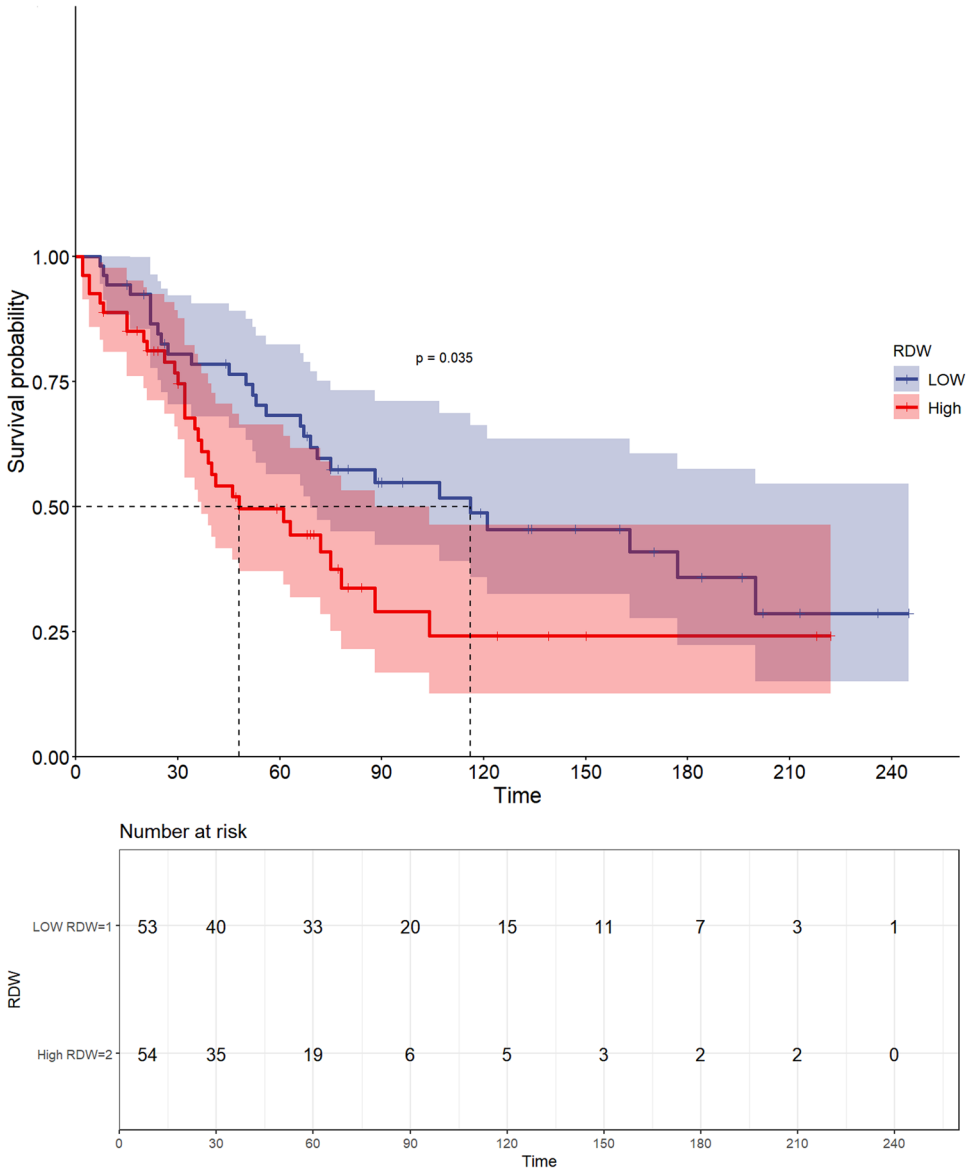


Fig. 4 Kaplan–Meier survival curves showing the association between RDW and all-cause mortality of the lung cancer

dynamically: the significance of a single RDW test is limited, and the trend of its changes should be tracked in order to reflect the progression of the disease or the response to treatment.Translated with www.DeepL.com/Translator (free version)[2, 7, 9–11]. The tumor microenvironment is often accompanied by chronic inflammation (elevated IL-6, TNF-α), and

Table 3 Baseline characteristics of 107 subjects

Characteristics	All patients	RDW		P
		Low RDW (< 13.7)	High RDW (13.7–19.8)	
Number	Total (n = 107)	1 (n = 53)	2 (n = 54)	
RDW (%)	13.70 (12.90, 14.60)	12.90 (12.40, 13.30)	14.55 (14.03, 15.28)	< 0.001
age	71.00 (64.50, 78.50)	70.00 (64.00, 76.00)	74.00 (65.00, 80.00)	0.466
Gender, n (%)				0.906
Male	64 (59.81)	32 (60.38)	32 (59.26)	
Female	43 (40.19)	21 (39.62)	22 (40.74)	
Race, n (%)				0.738
1	1 (0.93)	0 (0.00)	1 (1.85)	
2	3 (2.80)	2 (3.77)	1 (1.85)	
3	69 (64.49)	33 (62.26)	36 (66.67)	
4	23 (21.50)	11 (20.75)	12 (22.22)	
5	11 (10.28)	7 (13.21)	4 (7.41)	
Laboratory parameters				
Hematocrit (%)	40.80 (38.80, 42.65)	41.30 (39.30, 43.50)	40.00 (37.40, 42.60)	0.047
Mean cell volume (fL)	90.60 (86.60, 93.65)	91.10 (88.70, 93.70)	89.10 (84.95, 93.57)	0.084
Mean cell hemoglobin (pg)	30.30 (29.15, 31.75)	30.60 (29.60, 32.10)	29.95 (28.33, 31.30)	0.068
Platelet count (%) SI	248.00 (191.00, 300.00)	253.00 (200.00, 303.00)	232.50 (186.25, 296.25)	0.704
Hemoglobin (g/dL)	13.60 (12.85, 14.50)	13.90 (13.30, 14.90)	13.40 (12.43, 14.38)	0.014
Eosinophils number, 10^9/L	0.20 (0.10, 0.30)	0.20 (0.10, 0.30)	0.20 (0.10, 0.20)	0.895
WBC count, 10^9/L	7.30 (5.85, 9.15)	6.30 (5.40, 8.40)	7.65 (6.80, 9.60)	0.016
Lymphocyte number, 10^9/L	1.60 (1.20, 2.40)	1.50 (1.20, 2.20)	1.75 (1.20, 2.40)	0.426
Monocyte number, 10^9/L	0.60 (0.50, 0.75)	0.60 (0.50, 0.70)	0.60 (0.50, 0.80)	0.344
Segmented neutrophils number, 10^9/L	4.60 (3.60, 5.65)	4.10 (3.40, 5.50)	5.00 (4.17, 5.68)	0.011
Eosinophils number, 10^9/L	0.20 (0.10, 0.30)	0.20 (0.10, 0.30)	0.20 (0.10, 0.20)	0.895
Comorbidities, n (%)				
Emphysema, n (%)				0.292
Yes	32 (29.91)	13 (24.53)	19 (35.19)	
No	74 (69.16)	39 (73.58)	35 (64.81)	
Unknown	1 (0.93)	1 (1.89)	0 (0.00)	
Chronic bronchitis, n (%)				0.544
Yes	21 (19.63)	12 (22.64)	9 (16.67)	
NO	83 (77.57)	39 (73.58)	44 (81.48)	
Unknown	3 (2.80)	2 (3.77)	1 (1.85)	
Congestive heart failure, n (%)				0.346
Yes	9 (8.41)	3 (5.66)	6 (11.11)	
No	84 (78.50)	47 (88.68)	37 (68.52)	
Unknown	14(13.08)	3(5.66)	11(20.37)	
Coronary heart disease, n (%)				0.643
Yes	12 (12.90)	6 (12.00)	6 (13.95)	
NO	80 (86.02)	44 (88.00)	36 (83.72)	
Unknown	1 (1.08)	0 (0.00)	1 (2.33)	
Liver condition, n (%)				0.857
Yes	8 (7.48)	4 (7.55)	4 (7.41)	
NO	98 (91.59)	48 (90.57)	50 (92.59)	
Unknown	1 (0.93)	1 (1.89)	0 (0.00)	
Personal situation				
Military status, n (%)				0.060
Yes	39 (36.45)	24 (45.28)	15 (27.78)	

Table 3 (continued)

Characteristics	All patients	RDW		P
		Low RDW (< 13.7)	High RDW (13.7–19.8)	
Number	Total (n = 107)	1 (n = 53)	2 (n = 54)	
NO	68 (63.55)	29 (54.72)	39 (72.22)	
Education Level, n (%)				0.412
Less than high school education	34 (31.78)	17 (32.08)	17 (31.48)	
High school grad/GED or equivalent	30 (28.04)	12 (22.64)	18 (33.33)	
College graduate or above	43 (40.19)	24 (45.28)	19 (35.19)	

Table 4 HRs (95% CIs) for all-cause mortality across groups of RDW in the NHANES database

	Model I(not adjusted)		Model II		Model III	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI))	P
all-cause mortality						
Quartiles		0.018		0.170		0.029
Q2	1 (Reference)		1 (Reference)		1 (Reference)	
Q1	4.76 (1.31 ~ 17.27)		1.67 (0.80 ~ 3.50)		2.00 (1.07 ~ 3.74)	

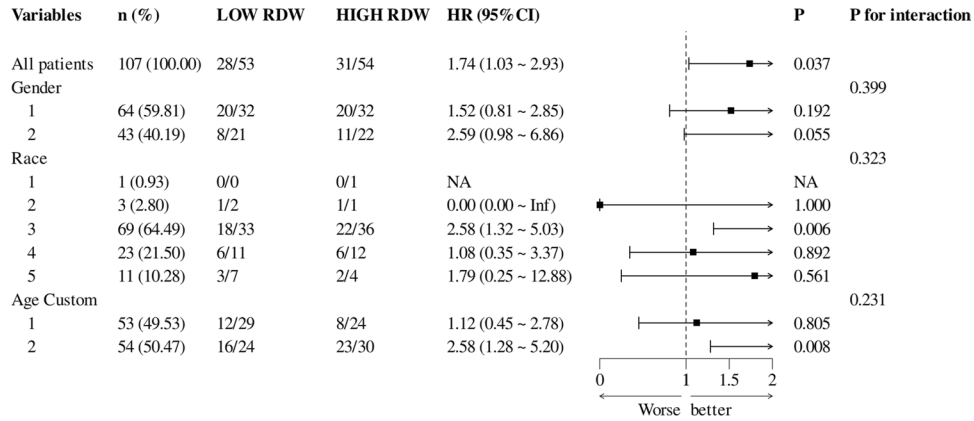


Fig. 5 Subgroup analysis of the associations between all-cause mortality of the lung cancer and the RDW

inflammatory factors inhibit erythropoiesis and destabilize erythrocytes, leading to increased erythrocyte size heterogeneity (elevated RDW). Meanwhile, the inflammatory response promotes tumor proliferation and metastasis by activating pathways such as NF-κB, and RDW indirectly reflects this vicious cycle. Elevated levels of reactive oxygen species (ROS) in cancer patients damage erythrocyte membranes and accelerate erythrocyte destruction, leading to elevated RDW. Oxidative stress can also promote carcinogenesis by inducing DNA damage and gene mutation, and RDW becomes an indirect marker of the degree of oxidative stress.Cancer patients typically exhibit elevated RDW values compared to the normal population. Cancer patients often experience inflammation and malnutrition due to the disease itself or side effects during treatment [12, 13]. These conditions may lead to abnormalities in erythrocyte production, subsequently affecting RDW values.

RDW is an affordable and straightforward measure indicating the variability in erythrocyte volume, known as anisocytosis. Under normal circumstances, the size and shape of erythrocytes are relatively consistent, but when RDW values increase, it indicates greater variation in erythrocyte size, which may be due to abnormalities in erythrocyte production, increased destruction, or certain disease states [14, 15]. RDW is significantly associated with malnutrition, especially deficiencies in iron, vitamin B12, and folate, which can diminish cancer patients'treatment responsiveness and lead to poor prognosis. Furthermore, in the advanced stages of malignant tumors, iron absorption may be severely affected due to impaired digestive system function, leading to decreased iron transport in the blood. This series of physiological changes can ultimately contribute to an increase in RDW levels [16–18]. Our research findings can serve as an auxiliary diagnostic tool for early screening of lung cancer in the future. As an indicator obtained from routine hospital admission

tests, RDW is simple and economical to measure. By monitoring this novel biomarker, we can predict the onset and prognosis of lung cancer [19].

7 limitations

The study has its limitations. Firstly, there is a correlation between cancer and inflammation. Chronic inflammation can lead to persistent tissue damage and repair, which increases the likelihood of cell variation and mutation. Certain molecules in the inflammatory response, such as cytokines and free radicals, promote angiogenesis and metastasis, providing a favorable environment for cancer cell growth. Inflammation can also affect the expression of related genes, increasing the probability of activating oncogenic mutations [20, 21]. The interaction between the two complicates the task of clarifying their precise relationship. Secondly, the study population for MR was limited to European individuals. Consequently, care must be taken when applying the study's findings to different racial or ethnic groups. This is because there may be genetic, environmental, lifestyle and other differences among different races or ethnic groups, which can all affect the applicability of the research results. Future studies should aim to increase sample size to encompass diverse races and ethnicities, while thoroughly accounting for confounding variables to enhance the precision and dependability of the findings.

8 Conclusion

Our study demonstrates a causal link between RDW and lung cancer, highlighting RDW's potential as a prognostic indicator for poor outcomes in lung cancer. However, further studies are needed to elucidate the exact mechanisms underlying the relationship between RDW and lung cancer.

Author contributions Contributions: (I) Conception and design: Yongli Liu; (II) Administrative support: Yiqing Qu; (III) Provision of study materials or patients: Yongli Liu; (IV) Collection and assembly of data: Yongli Liu; (V) Data analysis and interpretation: Yongli Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Funding We thank all the members involved in this study. The work described in this paper was supported by the Natural Science Foundation of Shandong Province (Grant No. ZR2023QH397), Major Project of National Natural Science Foundation of China (Grant No. 72293582) and ECCM Program of Clinical Research Center of Shandong University (Grant No.2021SDUCRCB001).

Data availability This study was conducted based on genome-wide association study (GWAS) data from individuals of European ancestry and lung cancer data from the Finn Gen consortium R11 (https://r11.finngen.fi/?).The second data from the National Health and Nutrition Examination Surveys (https://www.cdc.gov/nchs/nhanes/index.htm).

Declarations

Competing interests The authors declare no competing interests.

References

1. Dyba T, Randi G, Bray F, et al. The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. Eur J Cancer. 2021;157:308–47. https://doi.org/10.1016/j.ejca.2021.07.039.
2. Lu X, Huang X, Xue M, et al. Prognostic significance of increased preoperative red cell distribution width (RDW) and changes in RDW for colorectal cancer. Cancer Med. 2023;12(12):13361–73. https://doi.org/10.1002/cam4.6036.

3. Liang L, Huang L, Zhao X, et al. Prognostic value of RDW alone and in combination with NT-proBNP in patients with heart failure. Clin Cardiol. 2022;45(7):802–13. https://doi.org/10.1002/clc.23850.
4. Smith GD, Ebrahim S. "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1–22. https://doi.org/10.1093/ije/dyg070.
5. Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. J Am Soc Nephrol. 2016;27(11):3253–65. https://doi.org/10.1681/ASN.2016010098.
6. Wang MH, Cordell HJ, Van Steen K. Statistical methods for genome-wide association studies. Semin Cancer Biol. 2019;55:53–60. https://doi.org/10.1016/j.semcancer.2018.04.008.
7. Lv X, Xu B, Zou Q, Han S, Feng Y. Clinical application of common inflammatory and nutritional indicators before treatment in prognosis evaluation of non-small cell lung cancer: a retrospective real-world study. Front Med. 2023;10:1183886. https://doi.org/10.3389/fmed.2023.1183886.
8. Song B, Shi P, Xiao J, et al. Utility of red cell distribution width as a diagnostic and prognostic marker in non-small cell lung cancer. Sci Rep. 2020;10(1):15717. https://doi.org/10.1038/s41598-020-72585-4.
9. Dang C, Wang M, Qin T, Qin R. Clinical importance of preoperative red-cell volume distribution width as a prognostic marker in patients undergoing radical surgery for pancreatic cancer. Surg Today. 2022;52(3):465–74. https://doi.org/10.1007/s00595-021-02374-7.
10. Kos M, Hocazade C, Kos FT, et al. Evaluation of the effects of red blood cell distribution width on survival in lung cancer patients. Contemp Oncol. 2016;20(2):153–7. https://doi.org/10.5114/wo.2016.60072.
11. Yoo YC, Park S, Kim HJ, et al. Preoperative routine laboratory markers for predicting postoperative recurrence and death in patients with breast cancer. J Clin Med. 2021;10(12):2610. https://doi.org/10.3390/jcm10122610.
12. Lippi G, Targher G, Montagnana M, et al. Relation between red blood cell distribution width and inflammatory biomarkers in a large cohort of unselected outpatients published correction appears in Arch Pathol Lab Med. Arch Pathol Lab Med. 2009;133(4):628–32. https://doi.org/10.5858/133.4.628.
13. Oñate-Ocaña LF, Aiello-Crocifoglio V, Gallardo-Rincón D, et al. Serum albumin as a significant prognostic factor for patients with gastric carcinoma. Ann Surg Oncol. 2007;14(2):381–9. https://doi.org/10.1245/s10434-006-9093-x.
14. Salvagno GL, Sanchis-Gomar F, Picanza A, Lippi G. Red blood cell distribution width: a simple parameter with multiple clinical applications. Crit Rev Clin Lab Sci. 2015;52(2):86–105. https://doi.org/10.3109/10408363.2014.992064.
15. Lichtman MA. Red cell distribution width as a bellwether of prognosis. Blood Cells Mol Dis. 2024;109: 102884. https://doi.org/10.1016/j.bcmd.2024.102884.
16. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12(1):31–46. https://doi.org/10.1158/2159-8290.CD-21-1059.
17. Banaszkiewicz M, Małyszko J, Batko K, et al. The key role of hepcidin-25 in anemia in multiple myeloma patients with renal impairment. Medicina. 2022;58(3):417. https://doi.org/10.3390/medicina58030417.
18. Wang Y, Liao S, Pan Z, et al. Hydrogen sulfide alleviates particulate matter-induced emphysema and airway inflammation by suppressing ferroptosis. Free Radic Biol Med. 2022;186:1–16. https://doi.org/10.1016/j.freeradbiomed.2022.04.014.
19. Chen JL, Wu JN, Lv XD, Yang QC, Chen JR, Zhang DM. The value of red blood cell distribution width, neutrophil-to-lymphocyte ratio, and hemoglobin-to-red blood cell distribution width ratio in the progression of non-small cell lung cancer. PLoS One. 2020;15(8):e0237947. https://doi.org/10.1371/journal.pone.0237947.
20. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008;454(7203):436–44. https://doi.org/10.1038/nature07205.
21. Patel R, English L, Liu WK, et al. Red cell differential width (RDW) as a predictor of survival outcomes with palliative and adjuvant chemotherapy for metastatic penile cancer. Int Urol Nephrol. 2020;52(12):2301–6. https://doi.org/10.1007/s11255-020-02565-0.