BMC Rheumatology

RESEARCH

Open Access

# Association between red blood cell distribution width and genetic risk in patients with rheumatoid arthritis: a prospective cohort study and Mendelian randomization analysis

Mingyang Chen[1†], Jing Lei[2,3†], Zhenqiu Liu[1,3], Renjia Zhao[1,3], Yanfeng Jiang[1,3], Kelin Xu[2,3], Tiejun Zhang[2,3], Chen Suo[2,3,4*] and Xingdong Chen[1,3,5,6*]

## Abstract

**Objective** Elevated red blood cell distribution width (RDW) is associated with increased risk of rheumatoid arthritis (RA), but the potential interactions of RDW with genetic risk of incident RA remain unclear. This study aimed to investigate the associations between RDW, genetics, and the risk of developing RA.

**Methods** We analysed data from 145,025 healthy participants at baseline in the UK Biobank. The endpoint was diagnosed rheumatoid arthritis (ICD-10 codes M05 and M06). Using previously reported results, we constructed a polygenic risk score for RA to evaluate the joint effects of RDW and RA-related genetic risk. Two-sample mendelian randomization and bayesian colocalization were used to infer the causal relation between them.

**Results** A total of 675 patients with RA were enrolled and had a median followed up of 5.1 years, with an incidence rate of 0.57/1000 person-years. The hazard ratio of RA was 1.89 (95% CI: 1.45, 2.47) in highest RDW quartile group compared with the lowest RDW quartile group. Individuals within the top quintile of PRS showed a significantly high risk of RA. Moreover, Participants with high genetic risk and those in highest RDW group exhibited a significantly elevated hazard ratio (7.67, 95% CI: 3.98, 14.81), as opposed to participants with low genetic risk and those in lowest RDW group. Interactions between PRS and RDW on the multiplicative and additive scale were observed. Mendelian randomization provided suggestive evidence of a bi-directional causal relationship between RDW and RA. Loci near IL6R, IL1RN, FADS1/FADS2, UBE2L3 and HELZ2 showed colocalization.

†Mingyang Chen and Jing Lei contributed equally to this work.

*Correspondence:
Chen Suo
iamsuochen@gmail.com
Xingdong Chen
xingdongchen@fudan.edu.cn

Full list of author information is available at the end of the article

**Conclusion** Increased RDW is associated with elevated risk of incident RA especially in the high genetic risk populations, but only suggestive evidence supports a causal relationship between them.

**Keywords** Rheumatoid arthritis, Red blood cell distribution width, Genetic risk, Mendelian randomization, Colocalization

## Introduction

Rheumatoid arthritis (RA) is one of the most common chronic inflammatory diseases in the world [1]. As a systemic autoimmune disease, it is characterized by persistent inflammation in the synovial joints that result in joint destruction, joint deformity, and disorders, and contributes to the irreversible stage of the disease [2–4]. The high disability rate and harmfulness of RA, have necessitated the identification of novel potential RA-associated biomarkers for early detection and prevention of disease.

Red blood cell distribution width (RDW) is one of the parameters measured in normal complete blood counts and is used to reflect the heterogeneity of the red blood cell volume. RDW is commonly used in conjunction with mean red blood cell volume for differential diagnosis of different types of anemia, but a high negative predictive value has been discovered for its association with variety of underlying metabolic abnormalities such as oxidative stress, inflammation, dyslipidemia, hypertension [5]. Recently several recent studies have reported that RDW is associated with some autoimmune diseases like rheumatoid arthritis [6–8].

RA is also known to be a highly hereditary disease and twin studies have shown that genetic factors account for over half of the risk of developing RA [9]. Previous large-scale genome-wide association studies (GWAS) and meta-analyses have identified common genetic loci associated with RA risk in the population, some of which have shown across different races [10]. However, the influence of individual single-nucleotide polymorphisms (SNPs) on the genetic susceptibility of RA is relatively small or moderate. Therefore, the literature dose not clarify whether RDW and multiple genetic variants have a joint effect on the pathogenesis of RA.

In previous studies conducted in a healthy population through a follow-up cohort, the study population was not representative but the results were more generalizable [11, 12]. Therefore, in this study, we investigated the relationship between RDW and the incidence of RA in a healthy population. We also assessed the effect of RDW on the development of RA in participants with different levels of genetic risk, which could provide novel insights for early prevention of RA. A further bi-directional mendelian randomization (MR) were performed to check the genetic causal relationship.

## Methods

### Study population

The UK Biobank is a population-based prospective cohort study, that has been described in detail previously. In brief, 22 assessment centers in the United Kingdom recruited over 500,000 participants by postal invitation to individuals within a 25-mile radius between 2006 and 2010 [13]. The baseline summary characteristics of the cohort are available on the UK Biobank website (http://www.ukbiobank.ac.uk). The North West Multicenter Research Ethics Committee approved the study and all participants submitted informed consent forms. In this study, we excluded participants without RDW data and those who were not self-reported white British people. To avoid lagging records, participants who were diagnosed with any disease or withdrew before baseline and within three months during follow-up were excluded, leaving a total of 161,536 participants. Subsequently, we performed a genetic quality-checking before generating the polygenic risk score (PRS). Participants without genotyped data, showing a gender mismatch, high relatedness using kinship coefficients calculated from the PLINK (>0.125), or a high genotype missing rate (>5%) were excluded. These exclusions yielded a final sample size of 145,025.

### Outcomes

The main outcome of this study was the incidence of RA, which was defined in accordance with the International Classification Diseases, 10th edition (ICD-10 Diagnosis codes: M05, M06). The follow-up period for the participants began at enrollment and lasted until they are diagnosed as showing RA or censored (March 31, 2017). Censoring was defined as death, withdrawal from the study, or the end of the follow-up period, whichever came first.

### Measurements of variables

RDW was measured using four Beckman Coulter LH750 instruments within 24 h of the blood draw, with extensive quality control performed by UK Biobank. RDW is a continuous, highly skewed trait, so we divided the participants into four categories based on quartiles (≤12.90%, 12.91% ~ 13.33%, 13.34% ~ 13.85%, and ≥13.86%).

Other variates and covariates for the primary analysis included age(continuous), sex (male, female), body mass index (BMI; kg/m$^2$, continuous), C-reactive protein (CRP, mg/L, continuous), hemoglobin concentration

(Hb, g/dL, continuous), smoking status (current, former, never), drinking status (current, former, never), total physical activity (TPA, low, moderate, high), healthy diet score (continuous), prevalent hypertension (yes/no) and Townsend deprivation index (TDI, continuous). The evaluations of healthy diet score and total physical activity categories were presented in Supplementary Table S1. Rheumatoid factor was not included in the study due to the low coverage of measurement (<10%).

#### GWAS datasets for RDW and RA

Information regarding the genome-wide association summary data used was provided in the Supplementary Table S2. Details about the cohorts and quality control were explained in the original publications [14, 15]. The standardization of GWAS summary data via the R package MungeSumstats included (1) aligning to hg19 human reference genomes; (2) filtering single-nucleotide polymorphisms (SNPs) without rsID or with duplicated rsID; (3) removing dbSNP 144 based non-biallelic and strand-ambiguous SNPs; (4) keeping SNPs with a minor allele frequency > 0.01 [16]. SNPs with consistent base pair positions, reference and alternate alleles between two datasets were kept for further analyses.

#### Polygenic risk score

The PRS was developed from a meta-analysis of genome-wide association studies (GWAS) in populations of European ancestry, which benefitted from a substantial sample size despite including a few UK Biobank samples [15]. SNPs in the individual genotyped data showing a low minor allele frequency (<1%), deviation from the Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$), or high variant call missing rate (<99%) were removed before the analysis. SNPs were recorded and complemented if necessary and those with discordant alleles between RA GWAS summary data and genotyped data were discarded. Individual PRS were generated at nine *P*-value thresholds using PLINK 1.9. The odds ratio (OR) to evaluate the relationship between PRS and RA, adjusted for sex, age, and the first five genetic principal components, was calculated using logistic regression, results as shown in Supplementary Table S3. In this study, the threshold for SNP inclusion was a significance level of $p < 1 \times 10^{-5}$, which showed the highest OR among the models and included 1413 SNPs after clumping ($R^2 = 0.1$, windows = 250 kb) [17]. The major histocompatibility complex(MHC) region was also included considering its important role in the heredity of RA despite its long linkage disequilibrium.

#### Statistical analysis

Data analysis was performed using R version 4.2.2. All *P*-values for the tests were two-sided. *P*-values < 0.05 were considered statistically significant. The baseline participant characteristics were categorized in terms of the incidence of RA and summarized as percentage values for categorical variables and means with standard deviation (SD) for continuous variables showing a normal distribution. Continuous variables not showing normal distribution were summarized as median with quartiles. *P* values were calculated by one-way analysis of variance for continuous variables following normal distribution and by Kruskal-Wallis test for those not following. Chi-square test was used for categorical variables.

Cox proportional-hazard regression models were used to estimate the hazard ratios (HRs) and 95% confidence intervals (CIs). The models with age as the time scale were adjusted for sex, BMI, TPA, TDI, and healthy diet score. A possible nonlinear effect was tested using restricted cubic spline analysis. The number of knots, from three to five, was selected by the lowest value of Akaike information criterion (AIC) and finally we fitted the model with 5 knots [18]. RDW, other biochemical markers (CRP and, Hb concentrations), and several traditional risk factors (smoking, drinking, and prevalent hypertension) were considered in the study. A multivariable model was built with the factors which showed significant associations with RA. Two sensitivity analyses were performed: (1) to reduce the influence of outliers, we excluded the participants whose RDW level was not in the range of 2.5% to 97.5%; (2) to avoid the potential reverse causality, we excluded the participants diagnosed with RA within three years.

To estimate the HR for the association of PRS, cox proportional hazard regression models were adjusted for the covariates mentioned above and the first five genetic principal components. The scores were then categorized as low (<20%), intermediate (20 ~ 80%), or high (>80%). A stratification analysis was conducted for the combination of PRS and RDW (with low genetic risk and RDW Q1 as the reference). Kaplan–Meier incidence curves of PRS categories according to different RDW groups in relation to RA and the log-rank tests were calculated. Moreover, the interactions between PRS and RDW were tested at both multiplicative and additive scale by the function epi.interaction in the R package epiR.

To further evaluate the causal relationship between RDW and RA from the perspective of genetic structure, bi-directional two-sample MR methods were performed. We first applied cross-trait linkage disequilibrium score regression (LDSC) to check the possible inflation of population stratification and estimate the genetic correlations of RDW and RA [19, 20]. Besides the inverse variance weighted (IVW) method, other robust MR methods were utilized for sensitivity analyses, including consensus methods (weighted median, weighted mode-base), outlier-robust method (MR-PRESSO) and modelling methods (MR-Egger, contamination mixture) [21]. For these

six MR methods, instruments variants (IV) threshold was $p \leq 5 \times 10^{-8}$ and slected by SNP clumping (r2 < 0.001 within 10000 kb windows). For MHC region (chr6:25Mb-35Mb), only the most significant IV was retained. The method CAUSE (Causal Analysis Using Summary Effect Estimates) was used as well, which accounted for both correlated and uncorrelated pleiotropic effects [22]. A further genome-wide colocalization was performed at the region of the ±100 Kb of all the IVs under a single causal variant assumption between the two GWAS summary data, using the default setting of R package coloc [23]. Regions with the posterior probability of colocalization (PP.H4) over 0.80 were viewed as having a shared signal. It is important to note that colocalization cannot distinguish between vertical and horizontal pleiotropy.

## Results

### Baseline characteristics of the study participants in the UK Biobank

Among the 145,025 study participants, 675 cases of RA were recorded during a median follow-up period of 5.1 years. The incidence of RA was 0.57/1000 person-years (0.40/1000 person-years for men and 0.74/1000 person-years for women). The baseline characteristics of the participants stratified by incident RA are provided in Table 1. Participants with incident RA had a higher proportion of RDW in the group Q4 (42.5%) than those without (24.4%).

**Table 1** Baseline characteristics of the study participants in the UK Biobank

| Characteristics | Normal (n = 144,350) | Incident RA (n = 675) | *P* value |
|---|---|---|---|
| Age, mean (SD) | 56.47 (7.83) | 59.65 (7.16) | <0.001 |
| Sex, n (%) | | | <0.001 |
| Female | 72,844 (50.5) | 444 (65.8) | |
| Male | 71,506 (49.5) | 231 (34.2) | |
| BMI, mean (SD) | 26.88 (4.40) | 27.48 (5.08) | <0.001 |
| Missing, n (%) | 281 (0.2) | 2 (0.3) | |
| TDI, mean (SD) | −1.79 (2.81) | −1.51 (2.78) | 0.011 |
| Missing, n (%) | 172 (0.1) | 0 (0.0) | |
| Healthy diet score, median [IQR] | 3.00 [2.00, 4.00] | 3.00 [2.00, 4.00] | 0.622 |
| TPA group, n (%) | | | 0.109 |
| Low | 6,680 (4.6) | 40 (5.9) | |
| Moderate | 14,725 (10.2) | 63 (9.3) | |
| High | 98,487 (68.2) | 417 (61.8) | |
| Missing | 24,458 (16.9) | 155(23.0) | |
| RDW group, n (%) | | | <0.001 |
| Q1 (10.78–12.89) | 36,013 (24.9) | 105 (15.6) | |
| Q2 (12.89–13.30) | 35,481 (24.6) | 131 (19.4) | |
| Q3 (13.30–13.80) | 37,704 (26.1) | 152 (22.5) | |
| Q4 (>13.80) | 35,152 (24.4) | 287 (42.5) | |
| Alcohol status, n (%) | | | 0.001 |
| Never | 3,872 (2.7) | 27 (4.0) | |
| Previous | 3,323 (2.3) | 27 (4.0) | |
| Current | 137,075 (95.0) | 621 (92.0) | |
| Missing | 80 (0.0) | 0 (0.0) | |
| Smoking status, n (%) | | | <0.001 |
| Never | 84,127 (58.3) | 305 (45.2) | |
| Previous | 46,928 (32.5) | 262 (38.8) | |
| Current | 12,908 (8.9) | 102 (15.1) | |
| Missing | 387 (0.0) | 6 (0.9) | |
| Prevalent hypertension, n (%) | | | <0.001 |
| No | 112,341 (77.8) | 478 (70.8) | |
| Yes | 32,009 (22.2) | 197 (29.2) | |
| Hb, median [IQR] | 14.29 [13.44, 15.12] | 13.73 [12.96, 14.59] | <0.001 |
| Missing, n (%) | 1 (0.0) | 0 (0.0) | |
| CRP, median [IQR] | 1.17 [0.60, 2.35] | 2.37 [1.18, 5.43] | <0.001 |
| Missing, n (%) | 6061 (4.2) | 26 (3.9) | |
| Genetic risk category, n (%) | | | <0.001 |
| Low | 28,929 (20.0) | 73 (10.8) | |
| Intermediate | 86,650 (60.0) | 368 (54.5) | |
| High | 28,771 (19.9) | 234 (34.7) | |

*BMI* body mass index, *CRP* C-reactive protein, *IQR* interquartile range, *Hb* hemoglobin, *RA* rheumatoid arthritis, *RDW* red blood cell distribution width, *SD* standard deviation, *TDI* Townsend deprivation index, *TPA* total physical activity

### Observed association between RDW and incidence RA

The results of the univariable and multivariable Cox proportional risk regression models are shown in Fig. 1. In comparison with the lowest RDW group (Q1), only the highest RDW group showed a significantly increased RA risk (Q2: HR = 1.18, 95% CI = 0.88, 1.58; Q3: HR = 1.23, 95% CI = 0.93, 1.64; Q4: HR = 2.33, 95% CI = 1.81, 3.01). This sample showed no significant association between RA and drinking status and prevalent hypertension after adjustment. After adjusting for other variables, a higher RDW level was still a risk factor in the multivariable model, with the HR estimate only decreasing slightly. We performed test for linear trend based on the median value of each quantile of RDW ($P_{trend} < 0.001$) [24]. Time-varying effect was not observed in the proportional-hazard test, indicating the absence of a significant interaction between RDW and follow-up time. A "J-shaped" nonlinear correlation was observed in the model, which was consistent with our category status ($P_{non\text{-}linear} < 0.001$, Supplementary Fig. S1). Sensitivity analyses yielded similar results as before (Supplementary Tables S4, S5).

### Interaction effect between RDW and PRS for incident RA

The distribution of scaled PRS, which derived from 1413 SNPs, was shown in the Supplementary Fig. S2S. The RA risk increased monotonically per SD increase in PRS (HR = 1.51, 95% CI = 1.40, 1.63). Thus, participants with a higher genetic risk showed a higher risk of incident RA (intermediate: HR = 1.81, 95% CI = 1.35, 2.42; high: HR = 3.48, 95% CI = 2.56, 4.72, Supplementary Fig. S3).

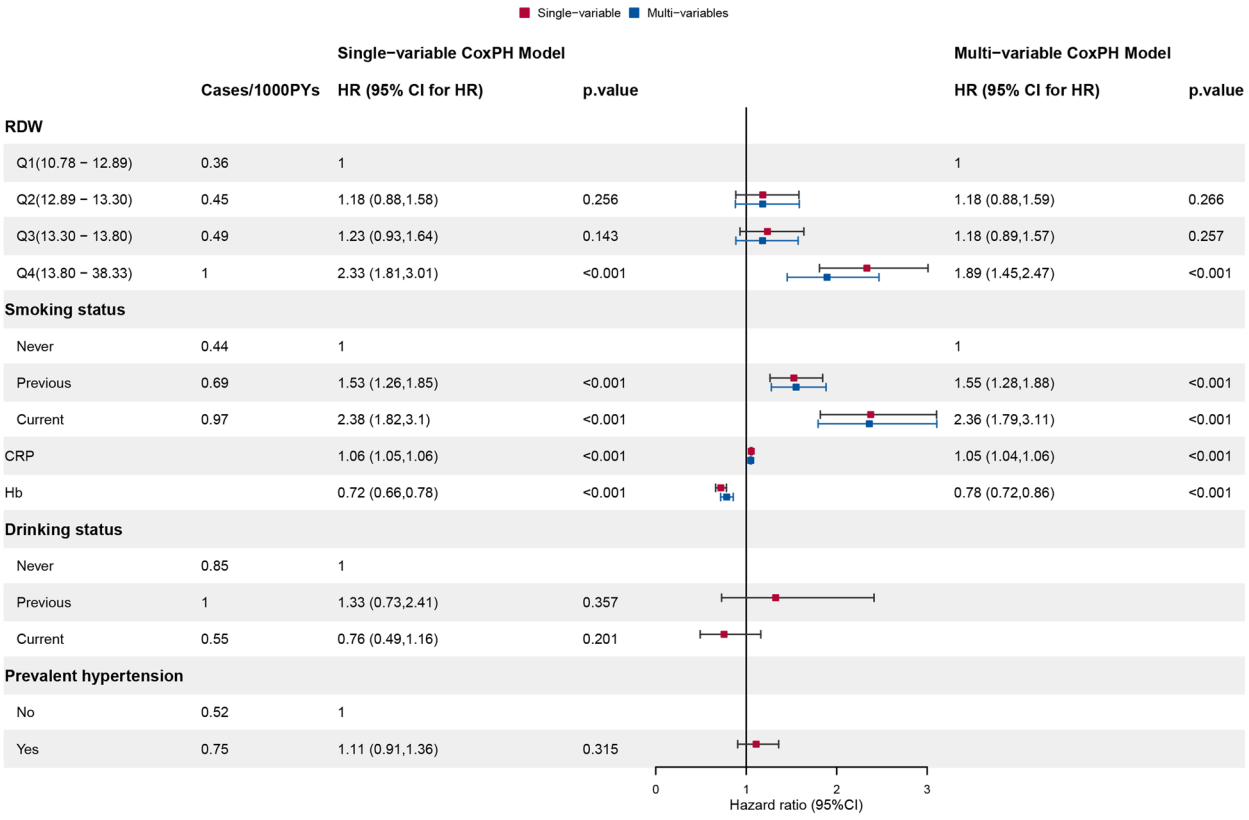


**Fig. 1** The association between incident RA and its risk factors for univariable and multivariable models. The association of incident RA and its potential risk factors was estimated by Cox Proportional Hazards models with age as the time scale. Analyses excluded participants who were diagnosed with any disease or withdrew before baseline and within three months during follow-up, and were adjusted for sex, BMI, total physical activity (TPA), Townsend deprivation index, and healthy diet score. Factors significant in the univariable model were included in the multivariable model. The error bars denote 95% confidence interval of hazard ratio. *CRP* C-reactive protein, *Hb* hemoglobin

Within each defined category of genetic risk, higher RDW was associated with a heightened relative risk for RA (Fig. 2). In comparison with participants showing low genetic risk and those in RDW group Q1, participants with high genetic risk and those in RDW group Q4 exhibited a substantially higher HR (HR=7.67, 95% CI=3.98, 14.81). Among participants in the intermediate and high genetic risk categories, the risk of incident RA escalated by 111% and 229% respectively for those in RDW group Q4 compared with those in RDW group Q1, as shown in Supplementary Table S6. Additionally, Substantial differences in relation to the varying degrees of genetic risks were also found in the RDW groups Q2, Q3, and Q4 with the significant log-rank test of the cumulative incidence curves (Supplementary Fig. S4).

A significant interaction effect was observed between high PRS group and RDW group Q4 as presented in Table 2. The estimated synergy index values for the interaction were 4.32 (95% CI=1.53, 12.19), with RDW group Q1 and low genetic risk serving as the reference, suggesting that the combined effect of RDW levels and genetic risk factors on an additive scale exceeds the individual effects of either factor alone. A synergistic interaction was evident on the multiplicative scale as well. These results suggested that RDW could distinguish the risk of incident RA, especially in conjunction with genetic risk, and that it could potentially optimize the definition of sub-populations at high risk to facilitate individualized RA prevention.

### Genetic association of RDW with RA

The observed impact of RDW in populations of diverse RA genetic risk highlighted the imperative for deeper investigation into the underlying genetic associations between them. The GWAS of RDW yielded an LDSC intercept of 1.19 (SE=0.07), whereas the GWAS of RA presented an LDSC intercept of 1.06 (SE=0.01), indicating that RDW might suffer from a more pronounced inflation in the mean chi-square statistic due to population stratification [19]. The estimated genetic correlation without constrained intercept was not significant ($r_g$=0.05, SE=0.03, $P$=0.088).

We first explored the genetic causal relationship from RDW to RA. The IVW method showed genetically predicted RDW had a positive causal effect on the incidence of RA (OR=1.11, 95% CI=1.03, 1.20, $P$=0.008,

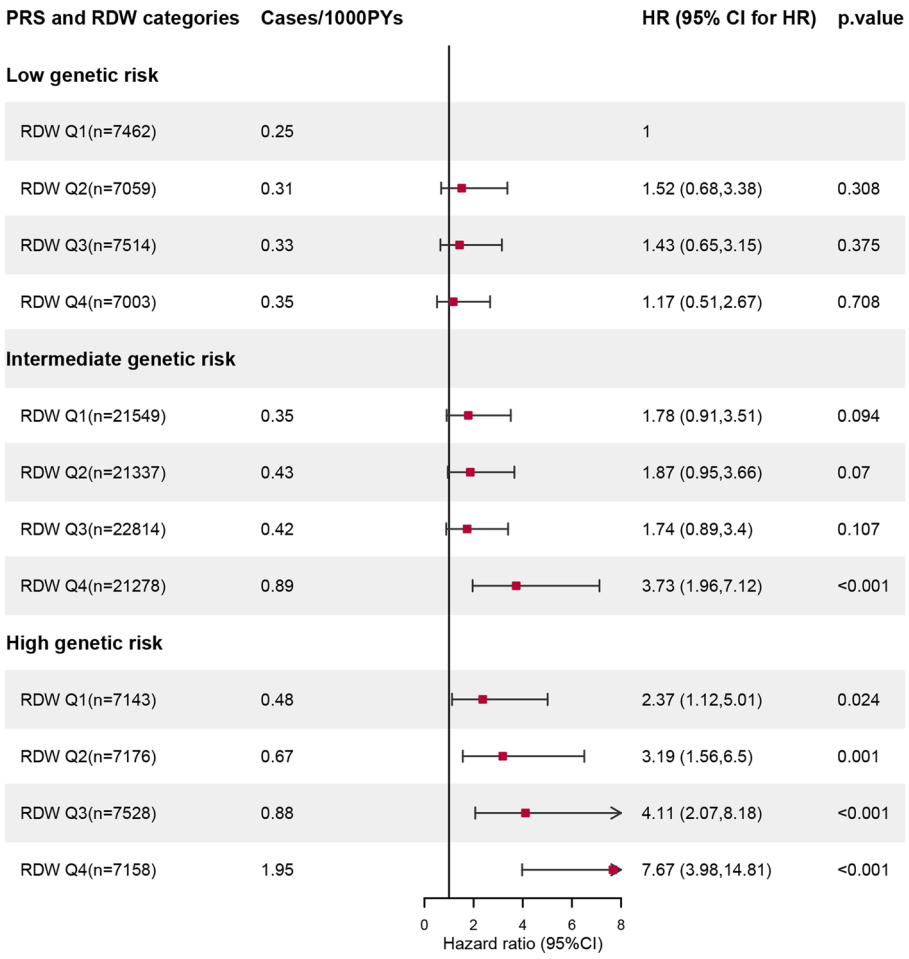


**Fig. 2** The association between incident RA and a combination of PRS and RDW. The groups of RDW were defined by the quartiles. The genetic risk categories were defined according to PRS as low (<20%), intermediate (20–80%), high (>80%). Joint association between combined genetic risk categories and RDW groups and RA was estimated by Cox Proportional Hazards models with age as the time scale, adjusted for sex, BMI, total physical activity (TPA), Townsend deprivation index, healthy diet score, and the first five genetic principal components. The error bars represent 95% confidence interval of hazard ratio. *PRS* polygenic risk score

**Table 2** Interaction effect between RDW and PRS with incident RA on the multiplicative and additive scale

| RDW | PRS | RERI | AP | S | Multiplicative effect | $P_{int}$ |
|---|---|---|---|---|---|---|
| Q1 | Low | 0 | 0 | 1 | 1 | - |
| Q2 | Intermediate | −0.43 (−1.78, 0.92) | −0.23 (−0.92, 0.45) | 0.66 (0.25, 1.71) | 0.69 (0.28, 1.67) | 0.411 |
| Q2 | High | 0.30 (−1.38, 1.98) | 0.10 (−0.45, 0.64) | 1.17 (0.45, 3.04) | 0.89 (0.34,2.33) | 0.806 |
| Q3 | Intermediate | −0.49 (−1.81, 0.84) | −0.29 (−0.99, 0.42) | 0.59 (0.22, 1.55) | 0.68 (0.28, 1.64) | 0.390 |
| Q3 | High | 1.24 (−0.46, 2.93) | 0.31 (−0.08, 0.71) | 1.71 (0.67, 4.37) | 1.21 (0.47, 3.11) | 0.688 |
| Q4 | Intermediate | 1.77 (0.63, 2.92) | 0.48 (0.19, 0.77) | 2.88 (0.71, 11.70) | 1.79 (0.74, 4.34) | 0.199 |
| Q4 | High | 5.38 (1.90, 8.86) | 0.67 (0.49, 0.86) | 4.32 (1.53, 12.19) | 2.76 (1.07, 7.14) | 0.036 |

*RERI* relative excess risk due to interaction, *AP* attributable proportion due to interaction, *S* synergy index

Fig. 3a, Supplementary Table S7). Consensus methods gave the insignificant results but in the same direction. High heterogeneity and possibly horizontal pleiotropy were detected in the IVW method (Cochran's Q=696.11, $P<0.001$). Therefore outlier-test in the MR-PRESSO was carried out and the causal relationship was still discovered (OR=1.08, 95% CI=1.01, 1.16, $P=0.021$). The MR-Egger intercept was compatible with no unbalanced pleiotropy ($P=0.570$). The contamination mixture also suggested incidence RA was affected by genetic instrumented RDW (OR=1.11, 95% CI=1.03, 1.19, $P=0.026$). Considering the possible inflation in the mean chi-square statistic, we repeated the analysis with a more restrict $p$ value threshold from $5\times10^{-8}$ to $1\times10^{-9}$. A

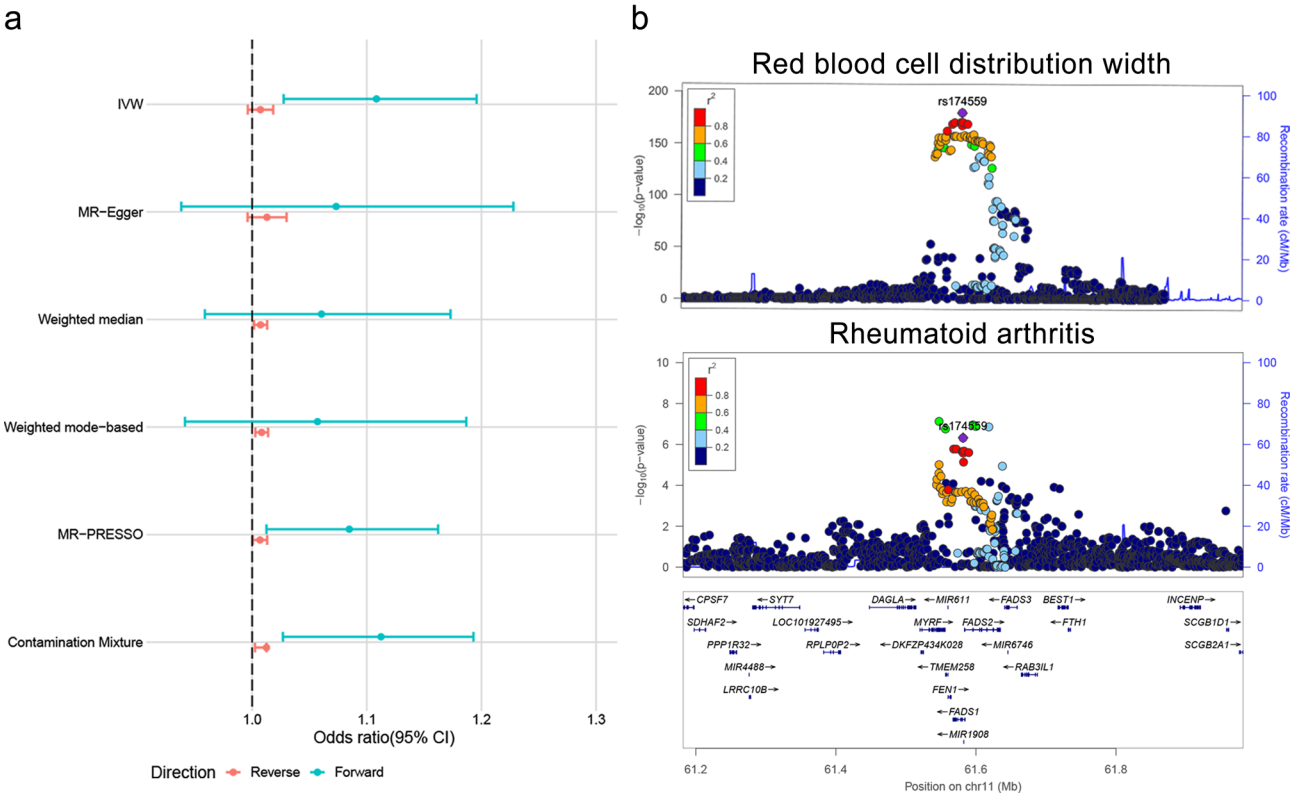


**Fig. 3** Genetic relationship between incident RA and RDW. **a** Bi-directional two-sample Mendelian randomization (MR) analysis was performed with different methods. Forward represented from RDW to RA and reverse represented from RA to RDW. **b** The LocusZoom plot at the region around rs174559

comparable result was obtained from the sensitivity analysis (Supplementary Table S7). For the reverse situation, despite the IVW method results being non-significant (OR=1.01, 95% CI=1.00–1.02, $P$=0.199), four out of the five sensitivity analysis methods yielded significant results (Fig. 3a). Nonetheless, the result of CAUSE, the only approach capable of distinguishing between correlated and uncorrelated pleiotropy, aligned with the null hypothesis, thereby suggesting an absence of a causal relationship in either direction (Supplementary Table S8). Consequently, both directions showed suggestive evidence of causal relationship between RDW and RA, though lacking robustness.

The results of MR analysis motivated us to understand the genetic relationship within one locus using colocalization. Genome-wide colocalization was performed around the IVs revealed five regions near as being associated with both RDW and RA (Fig. S5, Supplementary Table S9). Two loci near IL6R,IL1RN have previously been shown to be associated with RA and been identified as drug targets [1, 25, 26]. The region around SNP rs174559 was considered colocalization with a posterior probability of 97% (Fig. 3b). The rs174559 variant, with a wald ratio of 2.94 (95% CI=1.93, 4.47, $P$<0.001), is an intron variant of FADS1 and is located at the upstream of FADS2. FADS1/FADS2 are important enzymes producing long-chain polyunsaturated fatty acids and are associated with many cardiometabolic traits. FADS1 has been reported impacting metabolic disease by balancing proinflammatory and pro-resolving lipid mediators [27]. Another two regions near UBE2L3 and HELZ2 also showed evidence of colocalization. UBE2L3 encodes a member of the E2 ubiquitin-conjugating enzyme family and has been found to be associated with many autoimmune diseases in GWAS studies [28].

## Discussion

In this prospective cohort study based on a healthy population, we observed that a higher RDW was associated with an increased risk of incident RA. When the combined effect of RDW and genetic risk on RA risk was tested, the relative increase in risk was shown to be the greatest among those with a high RDW level and high genetic risk, and a positive interaction was observed on both the multiplicative and additive scale. Moreover, the MR analysis revealed suggestive evidence indicating a bidirectional causal relationship between RDW and RA.

A high RDW has been shown to be positively associated with the risk of RA in clinical case-control studies in Iraq [29], Turkey [30] and China [31]. In the research of China, RDW is elevated in patients with RA but not ankylosing spondylitis or osteoarthritis. It is also

reported that RDW further correlates with pain and Disease Activity Score 28 (DAS28) in RA patients from the study of Turkey. All the observational studies imply the necessity to explore the relationship between RDW and RA in a cohort study. To our knowledge, this study, which used data from the UK Biobank, is a relative large-scale longitudinal study in European adults to examine the relationship between baseline RDW and the incidence of RA with a 5.1-year median follow-up period and to test whether the interaction between RDW and genetic risk can increase the incidence of RA.

Given the strong association between RA and genetic factors, we also evaluated the combined effect and interaction of RDW and genetic factors on RA. Intriguingly, the finding revealed an interaction between RDW and the genetic risk of RA, suggesting high genetic risk and high RDW synergistically increased the risk of RA. This interaction pointed to a potential biological interconnection among these risk factors [12]. Consequently, more attention should be paid to the individuals within the general population who, despite appearing healthy, carry a high genetic risk coupled with increased RDW levels.

The current study on RA was relatively large in terms of the number of new cases included and therefore had sufficient statistical power to detect existing associations. However, it is necessary to distinguish whether the elevation of RDW is completely secondary to the development of RA. To eliminate the interference of confounding factors, we further performed bi-directional MR analyses to test the genetic causal association between RDW and RA. The MR results, which should be interpreted with caution, were not consistent in the sensitivity analyses for horizontal pleiotropy and confounding. Five regions showed colocalization and provided evidence for the role of RDW in RA although colocalization alone does not allow for the differentiation between horizontal and vertical pleiotropy. Our findings suggest that the elevation of RDW may be due to the genetic factors associated with RA, with more notable increases observed in individuals from high genetic risk populations, which highlight the potential of RDW as a promising and effective biomarker, deserving further exploration and validation in future studies.

Several underlying mechanisms may mediate the association between RDW and RA. This association may be related to the potential proinflammatory state and increased oxidative stress, both of which are related to impaired erythrocyte maturation [32, 33]. The characteristic structure of red blood cells maintained with the help of a phospholipid bilayer and several transmembrane proteins. However, some of these proteins, such as calpastatin, may stimulate the inflammatory activity of other cells, becoming antigens in physiological autoimmune reactions in patients with RA [34]. The result of colocalization also indicated the possible role of lipid metabolism between RDW and RA. The protein encoded by HELZ2 is a nuclear transcriptional co-activator for peroxisome proliferator activated receptor α (PPARα), which is the the master regulator of lipid metabolism in the liver [35, 36]. Through the PPARα, FADS1 can influence hepatic lipid homeostasis [37]. The dysregulation of lipid metabolism may further lead to inflammatory responses and trigger autoimmune diseases, creating a detrimental cycle. Although the pathophysiology of RA is completely clear, the exact latent mechanisms underlying the relationship of RDW with RA is complicated and needs to be explored further.

Several potential limitations of this study require consideration. First, our findings lacked validation from other larger-sample epidemiological studies. Second, despite statistical adjustment for covariates by various methods in this study, unmeasured or undetected influences such as drug use may have led to some unknown confounding. Third, information on the changes in RDW during the follow-up period was not available. The severity of diagnosed RA was also obscure for the use of electronic medical record, which made it challenging to eliminate potential reverse causality effects. Fourth, we did not subdivide analysis on RA subtypes. Finally, additional functional studies are warranted to shed light on the mechanisms underlying the effects of the interactions of genetic risk and RDW on the risk of RA.

## Conclusion

Elevated levels of RDW have been found to be associated with an increased risk of incident RA, particularly in individuals with a high genetic risk. However, it is important to exercise caution when interpreting the genetic causal relationship between RDW and RA. Further studies in other populations and experimental studies are needed to confirm our findings and to expound the basic mechanisms underlying these interactions.

### Abbreviations

| | |
|---|---|
| AIC | Akaike information criterion |
| BMI | Body mass index |
| CAUSE | Causal Analysis Using Summary Effect Estimates |
| CI | Confidence interval |
| CRP | C-reactive protein |
| DAS28 | Disease activity score 28 |
| GWAS | Genome-Wide Association Studies |
| Hb | Hemoglobin |
| HR | Hazard ratio |
| ICD | International Classification Diseases |
| IQR | Interquartile range |
| IVW | Inverse variance-weighted |
| LDSC | Linkage Disequilibrium Score Regression |
| MHC | Major Histocompatibility Complex |
| MR | Mendelian randomization |
| OR | Odds ratio |
| PRS | Polygenic risk score |
| RA | Rheumatoid arthritis |
| RDW | Red blood cell distribution width |

| | |
|---|---|
| SD | Standard deviation |
| SE | Standard Error |
| SNP | Single-nucleotide polymorphisms |
| TDI | Townsend deprivation index |
| TPA | Total physical activity |

## Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s41927-024-00451-1.

Additional file 1 Supplementary Figures

Additional file 2 Supplementary Tables

**Acknowledgements**
We sincerely appreciate the great effort of the UK Biobank working groups.

**Author contributions**
MC and JL conceived the idea for the study. JL and RZ obtained the data. MC and JL performed the data analyses. MC, JL, and CS interpreted the results of the data analyses. All authors wrote the manuscript. All authors read and approved the final manuscript.

**Funding**
This work was supported by the National Natural Science Foundation of China (grant numbers: 82073637, 82122060, 82473700), Shanghai Municipal Science and Technology Major Project (grant number: ZD2021CY001), National Key Research and Development Program of China (grant number: 2023YFC2508001, 2019YFC1315804), and the Shanghai Municipal Science and Technology Major Project (grant number: 2023SHZDZX02).

**Data availability**
The UK Biobank data are available from the UK Biobank on request. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Ethics approval and consent to participate**
The UK Biobank study was conducted according to the Declaration of Helsinki and ethical approval was granted by the North West Multi-Centre Research Ethics Committee (reference number 06/MRE08/65). At recruitment, all participants gave informed consent to participate and be followed up through data linkage. Details of the study protocol have been published elsewhere.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]State Key Laboratory of Genetic Engineering, Human Phenome Institute, Zhangjiang Fudan International Innovation Center, and School of Life Science, Fudan University, Shanghai 200120, China
[2]Department of Epidemiology, School of Public Health, Fudan University, Shanghai 200032, China
[3]Fudan University Taizhou Institute of Health Sciences, Taizhou, Jiangsu, China
[4]Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, China
[5]Yiwu Research Institute of Fudan University, Yiwu, Zhejiang, China
[6]National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai, China

Received: 18 November 2023 / Accepted: 17 December 2024
Published online: 02 January 2025

**References**

1. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet. 2016;388(10055):2023–38.
2. Smolen JS, Aletaha D. Rheumatoid arthritis therapy reappraisal: strategies, opportunities and challenges. Nat Rev Rheumatol. 2015;11(5):276–89.
3. Goswami RP, Basu K, Das S, Mondal S, Ghosh P, Ghosh A. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search update. Ann Rheum Dis. 2016;75(7):e35.
4. Silman AJ, Pearson JE. Epidemiology and genetics of rheumatoid arthritis. Arthritis Res Ther. 2002;4(3):S265.
5. 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.
6. Szczepanek-Parulska E, Hernik A, Ruchała M. Anemia in thyroid diseases. Pol Arch Intern Med. 2017;127(5):352–60.
7. Silva Litao MK, Kamat D. Back to basics: red blood cell distribution width: clinical use beyond hematology. Pediatr Rev. 2018;39(4):204–09.
8. Yousefi B, Sanaie S, Ghamari AA, Soleimanpour H, Karimian A, Mahmoodpoor A. Red cell distribution width as a novel prognostic marker in multiple clinical studies. Indian J Crit Care Med. 2020;24(1):49–54.
9. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43:30–37.
10. Dedmon LE. The genetics of rheumatoid arthritis. Rheumatology (Oxford). 2020;59(10):2661–70.
11. Pilling LC, Atkins JL, Kuchel GA, Ferrucci L, Melzer D. Red cell distribution width and common disease onsets in 240,477 healthy volunteers followed for up to 9 years. PLoS One. 2018;13(9):e0203504.
12. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am J Epidemiol 2017;186:1026–34.
13. Trehearne A. Genetics, lifestyle and environment. UK Biobank is an open access resource following the lives of 500,000 participants to improve the health of future generations. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2016;59(3):361–67.
14. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The polygenic and monogenic basis of blood traits and diseases. Cell 2020;182:1214–31.e11.
15. Ishigaki K, Sakaue S, Terao C, Luo Y, Sonehara K, Yamaguchi K, et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. Nat Genet 2022;54:1640–51.
16. Murphy AE, Schilder BM, Skene NG. MungeSumstats: a bioconductor package for the standardization and quality control of many GWAS summary statistics. Bioinformatics. 2021;37(23):4593–96.
17. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020;15(9):2759–72.
18. Zhu M, Ma Z, Zhang X, Hang D, Yin R, Feng J, et al. C-reactive protein and cancer risk: a pan-cancer study of prospective cohort and Mendelian randomization analysis. BMC Med 2022;20:301.
19. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291–95.
20. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47:1236–41.
21. Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using summary data. Genet Epidemiol. 2020;44(4):313–29.
22. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet. 2020;52(7):740–47.
23. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383.
24. Park SY, Freedman ND, Haiman CA, Le Marchand L, Wilkens LR, Setiawan VW. Association of coffee consumption with total and cause-specific mortality among nonwhite populations. Ann Intern Med. 2017;167(4):228–35.
25. Ramírez-Pérez S, De la Cruz-mosso U, Hernández-Bello J, Martínez-Bonilla GE, Ramírez-Dueñas MG, Pereira-Suárez AL, et al. High expression of interleukine-1 receptor antagonist in rheumatoid arthritis: association with IL1RN*2/2 genotype. Autoimmunity 2017;50:468–75.
26. Ramírez-Pérez S, Salazar-Páramo M, Pineda-Monjarás S, De la Cruz-mosso U, Hernández-Bello J, Martínez-Bonilla GE, et al. Association of 86 bp variable

number of tandem repeat (VNTR) polymorphism of interleukin-1 receptor antagonist (IL1RN) with susceptibility and clinical activity in rheumatoid arthritis. Clin Rheumatol 2017;36:1247–52.
27. Gromovsky AD, Schugar RC, Brown AL, Helsley RN, Burrows AC, Ferguson D, et al. Δ-5 fatty acid desaturase FADS1 impacts metabolic disease by balancing proinflammatory and proresolving lipid mediators. Arterioscler Thromb Vasc Biol 2018;38:218–31.
28. Zhang X, Huo C, Liu Y, Su R, Zhao Y, Li Y. Mechanism and disease association with a ubiquitin conjugating E2 enzyme: UBE2L3. Front Immunol. 2022;13:793610.
29. Al-Rawi ZS, Gorial FI, Al-Bayati AA. Red cell distribution width in rheumatoid arthritis. Mediterr J Rheumatol. 2018;29(1):38–42.
30. Tecer D, Sezgin M, Kanık A, İncel NA, Çimen ÖB, Biçer A, et al. Can mean platelet volume and red blood cell distribution width show disease activity in rheumatoid arthritis? Biomarker Med 2016;10:967–74.
31. Lin F, Wang X, Liang Y, Liu D, Zhang Y, Zhong R, et al. Red blood cell distribution width in rheumatoid arthritis, ankylosing spondylitis and osteoarthritis: true inflammatory index or effect of anemia? Ann Clin Lab Sci 2018;48:301–07.
32. Yunchun L, Yue W, Jun FZ, Qizhu S, Liumei D. Clinical significance of red blood cell distribution width and inflammatory factors for the disease activity in rheumatoid arthritis. Clin Lab. 2016;62(12):2327–31.
33. Ramos-González EJ, Bitzer-Quintero OK, Ortiz G, Hernández-Cruz JJ, Ramírez-Jirano LJ. Relationship between inflammation and oxidative stress and its effect on multiple sclerosis. Neurologia (Engl Ed). 2024 Apr;39(3):292–301.
34. Lackner KJ, Schlosser U, Lang B, Schmitz G. Autoantibodies against human calpastatin in rheumatoid arthritis: epitope mapping and analysis of patient sera. Rheumatology. 1998;37(11):1164–71.
35. Hashimoto T, Cook WS, Qi C, Yeldandi AV, Reddy JK, Rao MS. Defect in peroxisome proliferator-activated receptor alpha-inducible fatty acid oxidation determines the severity of hepatic steatosis in response to fasting. J Biol Chem. 2000;275(37):28918–28.
36. Kersten S, Seydoux J, Peters JM, Gonzalez FJ, Desvergne B, Wahli W. Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting. J Clin Invest. 1999;103(11):1489–98.
37. Athinarayanan S, Fan YY, Wang X, Callaway E, Cai D, Chalasani N, et al. Fatty acid Desaturase 1 influences hepatic lipid homeostasis by modulating the PPARα-FGF21 axis. Hepatol Commun 2021;5:461–77.

## Publisher's Note