

Unbiased Phenome-Wide Association Studies of Red Cell Distribution Width Identifies Key Associations with Pulmonary Hypertension

Timothy E. Thayer¹, Shi Huang², Rebecca T. Levinson¹, Eric Farber-Eger¹, Tufik R. Assad¹, Jessica H. Huston¹, Jonathan D. Mosley¹, Quinn S. Wells¹, and Evan L. Brittain¹

¹Department of Medicine and ²Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee

Abstract

Rationale: Red cell distribution width (RDW) is a prognostic factor in many diseases; however, its clinical utility remains limited because the relative value of RDW as a biomarker across disease states has not been established.

Objectives: To establish an unbiased RDW disease hierarchy to guide the clinical use of RDW and to assess its relationship to cardiovascular hemodynamic and structural parameters.

Methods: We performed phenome-wide association studies for RDW in discovery and replication cohorts derived from a deidentified electronic health record in nonanemic individuals. RDW values obtained within 30 days of echocardiogram or right heart catheterization were tested for association with structural and hemodynamic variables.

Results: RDW was associated with 263 phenotypes in both men and women in the discovery cohort ($n = 121,530$), 48 of which replicated in an independent cohort ($n = 2,039$). The strongest associations were observed with pulmonary arterial hypertension (odds ratio [OR], 2.1; 95% confidence interval [CI], 1.9–2.3), chronic

pulmonary heart disease (OR, 2.0; 95% CI, 1.9–2.2), and congestive heart failure (OR, 1.9; 95% CI, 1.8–2.0); $P < 1 \times 10^{-74}$ for all.

By echocardiography, RDW was higher in the setting of right ventricular dysfunction than left ventricular dysfunction ($P < 0.001$). Measured invasively, mean pulmonary arterial pressure was associated with RDW (21 vs. 33 mm Hg at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$) and remained strongly significant even when controlling for mean pulmonary capillary wedge pressure (21 vs. 29 mm Hg at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$).

Conclusions: Among 1,364 coded medical conditions, increased RDW was strongly associated with pulmonary hypertension and heart failure. Hemodynamic and echocardiographic phenotyping confirmed these associations and underscored that the most clinically relevant phenotype associated with RDW was pulmonary hypertension. These hypothesis-generating findings highlight the potential shared pathophysiology of pulmonary hypertension and elevated RDW. Elevated RDW in the absence of anemia should alert clinicians to the potential for underlying cardiopulmonary disease.

Keywords: erythrocyte indices; epidemiology; heart failure; biomarker

(Received in original form September 5, 2018; accepted in final form January 3, 2019)

Supported by American Heart Association Fellow to Faculty Grant 13FTF16070002, Pulmonary Hypertension Association Proof-of-Concept Award, Actelion Entelligence Young Investigator Award, and Gilead PAH Scholars Award Program (E.L.B.).

Author Contributions: T.E.T., Q.S.W., and E.L.B. participated in design of data analyses. T.E.T., R.T.L., E.F.-E., T.R.A., J.H.H., J.D.M., Q.S.W., and E.L.B. participated in acquisition and curating of data used for analyses. S.H. provided statistical planning and analyses. T.E.T. and E.L.B. prepared the manuscript.

Correspondence and requests for reprints should be addressed to Timothy E. Thayer, M.D., Division of Cardiovascular Medicine, Vanderbilt University Medical Center, 383 Preston Research Building, 2220 Pierce Avenue, Nashville, TN 37232. E-mail: timothy.thayer@vumc.org.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Ann Am Thorac Soc Vol 16, No 5, pp 589–598, May 2019

Copyright © 2019 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201809-594OC

Internet address: www.atsjournals.org

A query of PubMed with “red cell distribution width” (RDW) and “prognosis” yields more than 1,300 publications, highlighting the prognostic ability of RDW in a broad assortment of diseases. RDW is calculated from the width of the red blood cell (RBC) volume histogram and thus is a

measure of RBC volume variability. It has long been used in the diagnosis of anemias, but within the last decade RDW was discovered to be a potent predictor of mortality in both hospitalized and community-dwelling populations (1, 2). The mortality signal in those studies was not

explained by hematologic conditions, thus extending the role of RDW beyond hematology. Among nonhematologic conditions, the relationship between RDW and cardiovascular disease has been the most extensively studied, linking elevated RDW to mortality in patients with

congestive heart failure, coronary artery disease, and atrial fibrillation (3, 4).

Despite its prognostic value, the clinical utility of RDW is limited because it is unclear which clinical conditions are most strongly associated with RDW. In contrast, other seemingly nonspecific biomarkers, such as C-reactive protein and uric acid, are clinically useful, because diseases with which they are most strongly associated have been established. We hypothesized that an unbiased analysis of disease associations with RDW would establish a clinically useful hierarchy of RDW–disease associations.

The phenome-wide association study (PheWAS) is a powerful agnostic approach for identifying disease–exposure relationships (5). The PheWAS approach uses clinical phenotypes identified from approximately 18,000 *International Classification of Disease, Ninth Revision* (ICD-9) billing codes and condenses them into 1,815 hierarchical phenotypes, which is representative of the medical phenome. The PheWAS approach was first leveraged to test genotype–phenotype associations, but it has since been adapted to examine phenotypic associations with other variables, including laboratory values (6). We adapted this technique to test the association of RDW values in nonanemic subjects with ICD-coded diseases. Left and right heart failure were the strongest associated clinical phenotypes, and pulmonary arterial hypertension had the greatest magnitude of effect among significant phenotypes. We confirmed these associations in using raw hemodynamic and echocardiographic data.

Methods

PheWAS Study Populations

Individuals in the Synthetic Derivative, a deidentified mirror image of Vanderbilt's electronic health record, were divided into discovery and replication cohorts on the basis of the availability of genetic data as described below. The rationale for preserving a replication cohort with genetic data is that principal components could be used to better control for race, which is known to influence RDW (7).

Discovery cohort: All included individuals were at least 18 years old at date of last interaction with the electronic health record. At the time these data were extracted, the Synthetic Derivative included more than 2.8 million unique patient

records. To exclude individuals with an inadequate amount of clinical data, we required included individuals to have at least two RDW measurements between the years 2000 and 2018 and documented sex and race. Sex was extracted from administrative data that are collected at the time care is initiated. Because acute or chronic changes in red cell populations will increase RDW, we excluded individuals with anemia (defined as presence of any hemoglobin measurement less than 12 g/dl for women and less than 13.5 g/dl for men) (8).

Replication cohort: Vanderbilt has a DNA biorepository (BioVU) linked to the Synthetic Derivative (9). At the time of the study 13,076 individuals in the Synthetic Derivative also had genotype information on the Illumina Multi-Ethnic Genotyping Array (MEGA) chip, which had been obtained as part of other research initiatives. Individuals who had previously been genotyped on the MEGA chip were excluded from the discovery cohort. A PheWAS analysis was performed on this replication cohort after removing individuals who did not meet the same criteria as described above: at least two RDW measurements and no hemoglobin measurement less than 12 g/dl for women and less than 13.5 g/dl for men (8).

As no Health Insurance Portability and Accountability Act identifiers are available in the Synthetic Derivative database, this study met criteria for nonhuman subjects research.

Red Blood Cell Distribution Trait Measurement

Over the time period of this study, it was standard practice for our clinical chemistry laboratory to measure complete blood counts using tubes containing EDTA (ethylenediaminetetraacetic acid). Clinical measurements were typically performed within 1 hour of blood draw. All blood samples were run on automated hematology analyzers (Sysmex, models XN, XE, and XT).

PheWAS of Selected RDW Values

The PheWAS approach is a well-established method to identify attribute–phenotype associations in clinical cohorts (10–13). To limit the possibility that an RDW value was obtained far before a coded condition, the median RDW spanning the preceding 10 years from date of last record for each individual was calculated. PheWAS R software package was used to test the association of RDW across the breadth of the medical phenome as ascertained by curated aggregations of ICD-9 codes (ICD-9 to phecode mapping available here: <https://pewascatalog.org/phecodes>) (12–14). Covariates used are listed in the statistical section below. Phenotypes with fewer than 100 total cases were excluded from analyses in the discovery cohort to avoid false-positive associations with low-frequency diseases. Phenotypes with at least 100 cases that met Bonferroni corrected *P* value for association with RDW were then ordered by odds ratio (OR) to identify the most

Table 1. Demographics and comorbidities of discovery and replication phenome-wide association study populations

	Discovery Cohort (<i>n</i> = 121,530)	Replication Cohort (<i>n</i> = 2,039)
Age, mean (SD)	49.52 (16.70)	64.46 (11.06)
Race, white	104,468 (86)	2,039 (100)
Male	48,344 (40)	937 (46)
RDW, median (IQR)	13.1 (12.7–13.6)	13.3 (12.8–13.8)
Hemoglobin, mean (SD)	14.32 (1.13)	14.52 (1.08)
Atrial fibrillation	4,381 (4)	465 (23)
Chronic kidney disease	1,481 (1)	76 (4)
Chronic obstructive pulmonary disease	4,664 (4)	184 (9)
Chronic pulmonary heart disease	965 (1)	128 (6)
Coronary artery disease	8,212 (7)	705 (35)
Diabetes mellitus	11,726 (10)	342 (17)
Congestive heart failure	3,530 (3)	349 (17)
Hypertension	36,298 (30)	1,388 (68)
Obesity	13,737 (11)	352 (17)
Pulmonary arterial hypertension	365 (0.3)	46 (2)

Definition of abbreviations: IQR = interquartile range; RDW = red cell distribution width; SD = standard deviation.

Data presented as *n* (%) unless otherwise noted.

clinically significant associations. In the smaller replication cohort, phenotypes with fewer than 10 total cases were excluded from analysis to limit risk of false-positive associations. Results were organized by magnitude of effect to identify clinically significant association. Ordering by magnitude of effect had the added benefit of mitigating the effect of more prevalent phenotypes being more powered to identify RDW–phenotype associations.

Echocardiogram Measurements

The population used for RDW and echocardiography association analyses consisted of patients referred for an echocardiogram as part of clinical practice, as described previously (15). Echocardiograms were interpreted at a single laboratory by expert echocardiographers. We excluded individuals who did not have an RDW measurement within 30 days of the echocardiogram. If a subject was referred for multiple echocardiograms, data from the first study were used for analysis. Semiquantitative (i.e., normal, mild, moderate, etc.) assessments of chamber size were extracted from echocardiogram reports as interpreted by expert readers and condensed to a 0 to 5 scale, with increasing numbers representing larger chambers (0: normal size, 1: mildly dilated, 2: mild to moderately dilated, 3: moderately dilated, 4: moderate to severely dilated; 5: severely dilated). Semiquantitative assessments of ventricular function were similarly extracted, and degree of dysfunction was condensed into normal, mild, moderate, or severe for each individual. All continuous echocardiographic values (i.e., left atrial diameter, right ventricular systolic pressure) were extracted directly from the report in the medical record. Nonphysiologic outliers were removed as previously described (15).

Right Heart Catheterization Data

The development of the right heart catheterization (RHC) cohort used for these analyses was performed as previously described (16–18). Briefly, the Synthetic Derivative was queried to identify all individuals who underwent initial RHC. We manually reviewed charts of all patients meeting the consensus hemodynamic definition of pulmonary arterial hypertension to confirm this diagnosis (19). Comorbidities were determined by the

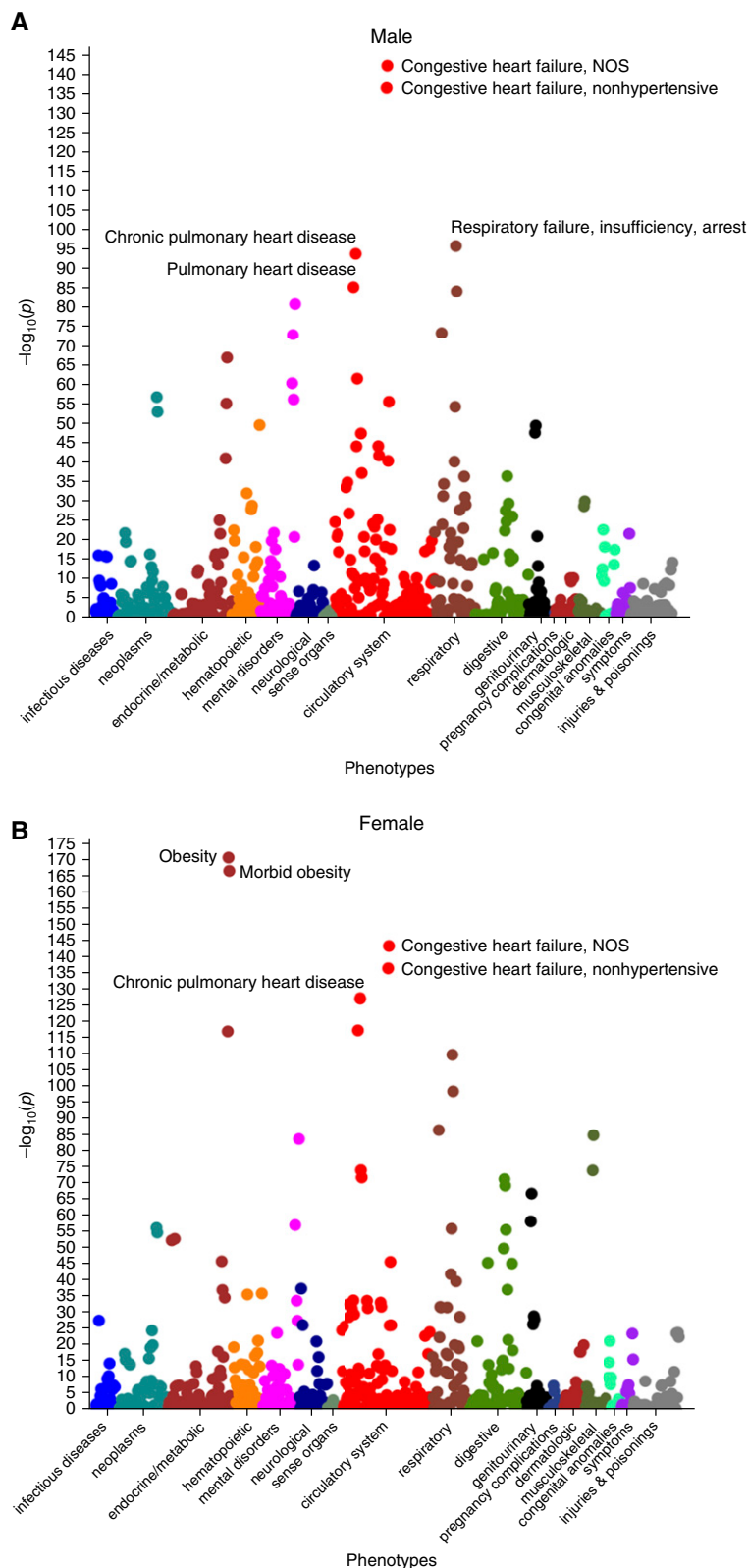


Figure 1. Association plots of red cell distribution width (RDW)–phenotype relationships that are associated with RDW in both (A) men and (B) women. Colors represent distinct phenotype categories, which are organized by organ system. NOS = not otherwise specified.

presence of at least two instances of ICD-9 codes for a given diagnosis in an individual's medical record. Mean hemoglobin measurements were calculated using the available value closest to the date of catheterization for each individual.

Association studies of RHC variables and RDW values were limited to individuals who had RDW values measured within 30 days of the procedure. If there were multiple RDW values available for an individual, the RDW value closest to date of the procedure was used.

In follow-up analyses to compare the mean RDW value for right versus left heart failure, the RHC cohort was divided into the subgroups on the basis of catheterization profile, as previously described (16). The first analyses compared the following groups: individuals without left heart failure or pulmonary hypertension (mean pulmonary artery pressure [mPAP] < 25 and pulmonary capillary wedge pressure [PCWP] < 15), left heart failure without pulmonary hypertension (mPAP < 25 and PCWP ≥ 15), left heart failure with isolated postcapillary pulmonary hypertension (mPAP ≥ 25, PCWP ≥ 15, diastolic pressure gradient [DPG]) < 7 mm Hg, and pulmonary vascular resistance [PVR] ≤ 3 Wood units), and precapillary pulmonary hypertension without left heart failure (mPAP ≥ 25, PCWP < 15, and PVR > 3 Wood units). Next, individuals with left heart failure and pulmonary hypertension were divided into those with combined pre- and postcapillary pulmonary hypertension (mPAP ≥ 25, PCWP ≥ 15, DPG ≥ 7 mm Hg, and PVR > 3 Wood units) versus isolated postcapillary pulmonary hypertension (mPAP ≥ 25, PCWP ≥ 15, DPG < 7 mm Hg, and PVR ≤ 3 Wood units). These cutoffs were determined using the 2015 European Society of Cardiology and the European Respiratory Society guidelines for diagnosis of pulmonary hypertension (20).

Informed Consent

All studies were approved by the Vanderbilt Institutional Review Board. Informed consent was waived because all data are deidentified.

Statistical Analyses

Our PheWAS analyses tested for the association of RDW (as a continuous variable) and each clinical phenotype by logistic regression and adjusted for the

covariates age at last record and race (self-declared race for the discovery cohort and race by principal components 1 through 3 in the replication cohort). PheWAS analyses were stratified by sex, because many clinical phenotypes are sex specific. The Bonferroni corrected *P* value for the PheWAS of each sex was calculated for number of phenotypes with at least 100 cases for the discovery cohort (1,179 phenotypes for women: Bonferroni *P* value = 4.2×10^{-5} ; 913 phenotypes for men: Bonferroni *P* value = 5.5×10^{-5}) and at least 10 cases for the replication cohort (649 phenotypes for women:

Bonferroni *P* value = 7.7×10^{-5} ; 494 phenotypes for men: Bonferroni *P* value = 1×10^{-4}).

Because the distributions of the majority of the echocardiogram and RHC parameters were highly skewed, we conducted ordinal regressions (with logit link) to examine the relationships between RDW (as a continuous variable) and each echocardiogram or RHC parameter. Compared with ordinary least squares regression, ordinal regression has no assumption for the distribution of outcome, and parameter estimates are robust to extreme outcome values. Semiparametric ordinal regression is highly robust and

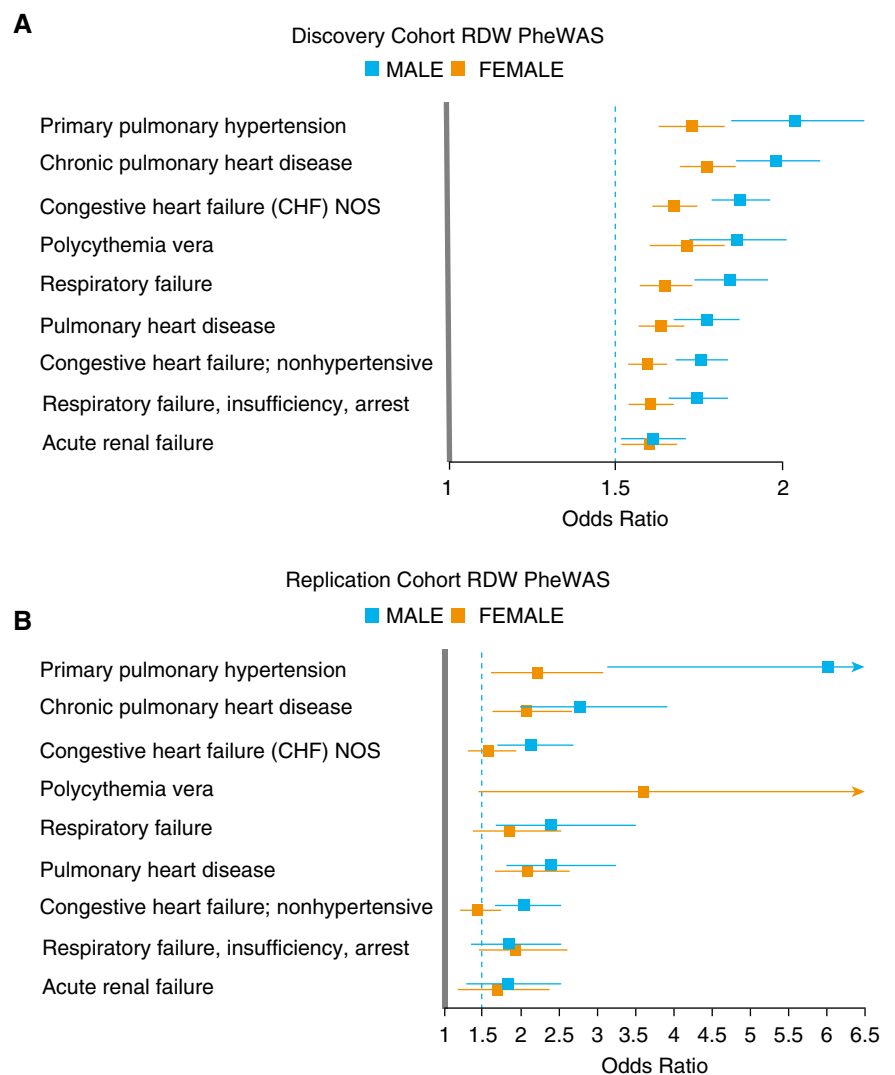


Figure 2. Top phenotypes associated with red cell distribution width (RDW) by phenome-wide association study (PheWAS) analyses. (A) PheWAS results for RDW in the discovery cohort stratified by sex. All displayed phenotypes from the discovery cohort had a lower 95% confidence interval above 1.5. (B) PheWAS results for RDW from the replication cohort, limited to the phenotypes identified in the discovery cohort. Mean odds ratios and 95% confidence intervals are shown. NOS = not otherwise specified.

efficient and has a special advantage over the other methods, as it is invariant to transformations of the dependent variable and particularly robust to outliers (21). Covariates were age at echocardiogram, sex, and self-declared race for the analysis. The ORs (OR of comparing the 75th percentile vs. the 25th percentile of RDW) with their 95% confidence intervals (CIs) were presented for these analyses. A significant OR greater than 1 represents larger size for ventricular measurements. Because we examined multiple outcomes, the *P* values for RDW in each model were adjusted by Bonferroni method. The Bonferroni corrected *P* value for both the echocardiogram analyses (15 variables) and RHC hemodynamic analyses (23 variables) were 0.003 and 0.002, respectfully. RDW values were compared between subgroups of the RHC cohort using one-way analysis of variance with uncorrected Dunn's *post hoc* multiple comparisons testing. The comparison of combined pre- and postcapillary versus postcapillary pulmonary hypertension was performed using the Mann-Whitney test. In follow-up RHC analyses, the pulmonary hypertension group was tested for interaction with RDW–mPAP and RDW–PVR associations using linear regression analysis adjusted for age, sex, and self-declared race.

Results

PheWAS of RDW

The discovery cohort included 121,530 individuals (60% women, 49 ± 17 yr). Demographics and comorbidities are shown in Table 1. PheWAS analyses of both cohorts were stratified by sex, given the presence of multiple sex-specific phenotypes. RDW was associated with 263 phenotypes in both men ($P < 5.5 \times 10^{-5}$, 913 phenotypes with at least 100 cases) and women ($P < 4.2 \times 10^{-5}$, 1,179 phenotypes with at least 100 cases; Figure 1). The strongest associations per 1% increase in RDW for men were congestive heart failure (OR, 1.9; 95% CI, 1.82–2.01; $P = 6 \times 10^{-143}$), respiratory failure (OR, 1.8; 95% CI, 1.67–1.86, $P = 1 \times 10^{-96}$), and chronic pulmonary heart disease (OR, 2.0; 95% CI, 1.88–2.15; $P = 1 \times 10^{-94}$). The strongest associations for women were obesity (OR, 1.4; 95% CI, 1.37–1.45; $P = 4 \times 10^{-171}$), congestive heart failure (OR, 1.7; 95% CI, 1.63–1.77; $P = 4 \times 10^{-144}$), and chronic

pulmonary heart disease (OR, 1.8; 95% CI, 1.73–1.91; $P = 9 \times 10^{-128}$).

After establishing statistical significance adjusted for multiple comparisons, we prioritized associations on the basis of effect size to identify the most clinically relevant associations. Among significantly associated phenotypes, pulmonary arterial hypertension (men: OR, 2.1; 95% CI, 1.9–2.3; $P = 2 \times 10^{-44}$; women: OR, 1.77; 95% CI, 1.66–1.88; $P = 2 \times 10^{-74}$) and chronic pulmonary heart disease (men: OR, 2.0; 95% CI, 1.88–2.15; $P = 1 \times 10^{-94}$; women: OR, 1.8; 95% CI, 1.73–1.91; $P = 9 \times 10^{-128}$) had the largest magnitude of effect per 1% increase in RDW in both men and women. Figure 2, top, displays the phenotypes associated with RDW that have an OR of at least 1.5 per 1% increase in RDW and a significant adjusted *P* value. Phenotypes previously established to be positively associated with RDW were recapitulated in both cohorts by PheWAS, including chronic kidney disease, cirrhosis, atrial fibrillation, and obesity ($P < 1 \times 10^{-18}$ for all) (22–25). Phenotypes that were negatively associated with RDW were strongly enriched with relatively benign conditions, such as gingivitis, myopia, and benign neoplasm of the skin ($P < 1 \times 10^{-10}$ for all; see Figure E1 in the online supplement). Full tabular results for men and women are provided in Tables E1 and E2.

The replication cohort comprised 2,049 individuals previously genotyped on the MEGA chip. Demographics and

comorbidities are shown in Table 1. PheWAS analyses of the replication cohort were performed identically as for the discovery cohort, except for the replacement of self-identified race for race by principal components (full tabular results for men and women are shown in Tables E3 and E4). The phenotype with the highest OR per 1% increase in RDW was pulmonary arterial hypertension in both sexes (men: OR, 6.0; 95% CI, 3.1–13.0; $P = 5 \times 10^{-7}$; women: OR, 2.2; 95% CI, 1.6–3.1; $P = 8 \times 10^{-7}$). Forty-eight phenotypes replicated in the replication cohort ($P < 0.05$, Table E5) and 12 phenotypes met Bonferroni correction in both sexes and both cohorts (Table 2). Analyses of phenotypes in the replication cohort identified in the discovery cohort with an OR lower 95% CI of at least 1.5 are displayed in Figure 2, bottom.

RDW Was Associated with Echocardiographic Measurements of Right Heart Dysfunction

We identified 4,880 individuals with no history of anemia who had RDW values obtained within 30 days of a clinically indicated echocardiogram (median, -2.9 d; interquartile range [IQR], -7 to 0 d). The cohort had a mean age of 56 years; 61% were women, 5% had pulmonary arterial hypertension, and 20% had congestive heart failure (Table 3). RDW values were tested for association with available echocardiographic measurements in an unbiased analysis. Increasing RDW was most strongly associated with right

Table 2. Phenotypes associated with red cell distribution width by Bonferroni correction for multiple comparisons in both sexes in both cohorts

Phenotype	Weighted Mean OR and 95% CI per 1% RDW	<i>P</i> Value
Primary pulmonary hypertension	2.00 (1.75–2.40)	2×10^{-74}
Chronic pulmonary heart disease	1.95 (1.79–2.13)	9×10^{-128}
Congestive heart failure NOS	1.81 (1.70–1.93)	7×10^{-144}
Respiratory failure	1.79 (1.68–1.92)	4×10^{-99}
Pulmonary heart disease	1.76 (1.64–1.90)	6×10^{-118}
Respiratory failure, insufficiency, arrest	1.71 (1.62–1.81)	2×10^{-110}
Congestive heart failure; nonhypertensive	1.70 (1.61–1.80)	4×10^{-137}
Cardiomegaly	1.56 (1.46–1.68)	4×10^{-72}
Morbid obesity	1.54 (1.48–1.61)	5×10^{-167}
Chronic airway obstruction	1.47 (1.40–1.55)	6×10^{-87}
Obesity	1.40 (1.35–1.44)	4×10^{-171}
Overweight, obesity, and other hyperalimentation	1.30 (1.26–1.34)	1×10^{-117}

Definition of abbreviations: CI = confidence interval; NOS = not otherwise specified; OR = odds ratio; RDW = red cell distribution width. Displayed *P* values are lowest found in either cohort and sex.

Table 3. Demographics and comorbidities of nonanemic patients referred for echocardiography or right heart catheterization

	TTE cohort (n = 4,880)	RHC cohort (n = 705)
Age, mean (SD)	56 (18)	55 (15)
Race, white	3,941 (81)	618 (88)
Male	1,878 (39)	308 (44)
RDW, mean (SD)	14.32 (1.17)	14.54 (1.31)
Hemoglobin, mean (SD)	13.55 (1.22)	15.86 (12.39)
Days from RDW to TTE/RHC, median (IQR)	4.1 (1–5)	2.9 (7–0)
Atrial fibrillation	910 (19)	140 (20)
Chronic kidney disease	111 (2)	19 (3)
Chronic obstructive pulmonary disease	414 (9)	58 (8)
Chronic pulmonary heart disease	474 (10)	382 (54)
Coronary artery disease	1,099 (23)	281 (40)
Diabetes mellitus	588 (12)	94 (13)
Congestive heart failure	952 (20)	325 (46)
Hypertension	2,451 (50)	348 (49)
Obesity	498 (10)	115 (16)
Pulmonary arterial hypertension	248 (5)	217 (31)

Definition of abbreviations: IQR = interquartile range; RDW = red cell distribution width; RHC = right heart catheterization; SD = standard deviation; TTE = transthoracic echocardiogram. Data presented as n (%) unless otherwise noted.

ventricular systolic pressure (RVSP) (predicted median RVSP was 27.5 and 33.1 mm Hg at 25th and 75th percentile of RDW, respectively) and right atrial pressure (predicted median right atrial pressure was 4.6 and 7.3 mm Hg at 25th and 75th percentile of RDW, respectively; $P < 1 \times 10^{-10}$ for both; Figure 3, left panel). Although RDW was associated

with both right and left ventricular dysfunction as assessed by expert sonographers ($P < 2 \times 10^{-5}$ by ordinal regression), RDW was higher in the setting of right ventricular failure than left ventricular failure (Figure 2, bottom; $P < 0.001$). Full tabular results for RDW–echocardiographic associations are shown in Table E6.

RDW Was Associated with Measurements of Right Heart Afterload by Invasive Hemodynamics

We identified 705 individuals referred for RHC who met our prespecified criteria for RDW values (median interval between RHC and RDW value, –4.1 d; IQR, –5 to –1 d). The mean age of the RHC cohort was 55 years; 56% were women, 31% had pulmonary arterial hypertension, and 46% had congestive heart failure (Table 3).

RDW was most strongly positively associated with pulmonary artery mean pressure (21 mm Hg vs. 33 mm Hg, at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$) and negatively associated with pulmonary artery saturation (73% vs. 67%, at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$) in addition to other parameters obtained at RHC (Figure 4). Multiple hemodynamic surrogates of pulmonary vascular remodeling were positively associated with rising RDW, including transpulmonary pressure gradient (10 vs. 17 mm Hg, at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$), pulmonary vascular resistance (1.9 vs. 2.7 Wood units, at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$), and diastolic pressure gradient (1.9 mm Hg vs. 6 mm Hg, at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$). RDW remained strongly associated with mPAP even when controlling for pulmonary capillary wedge pressure

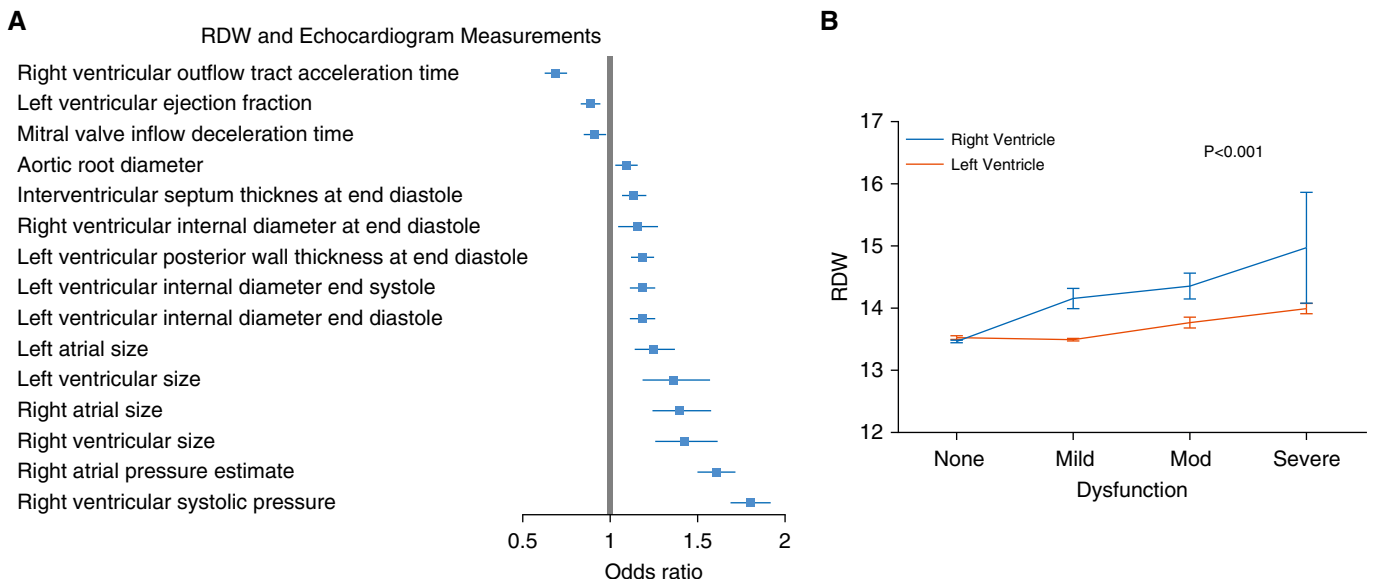


Figure 3. The association of red cell distribution width (RDW) with multiple parameters of cardiac structure and function by echocardiography. (A) For each variable, the boxes represent the odds ratio comparing 25th versus 75th percentile of RDW. Horizontal lines represent 95% confidence intervals. Analyses performed by ordinal regression adjusting for age at echocardiogram, sex, and self-declared race. (B) Mean RDW by degree of ventricular dysfunction. P value calculated by two-way analysis of variance. Vertical bars represent standard error of the mean.

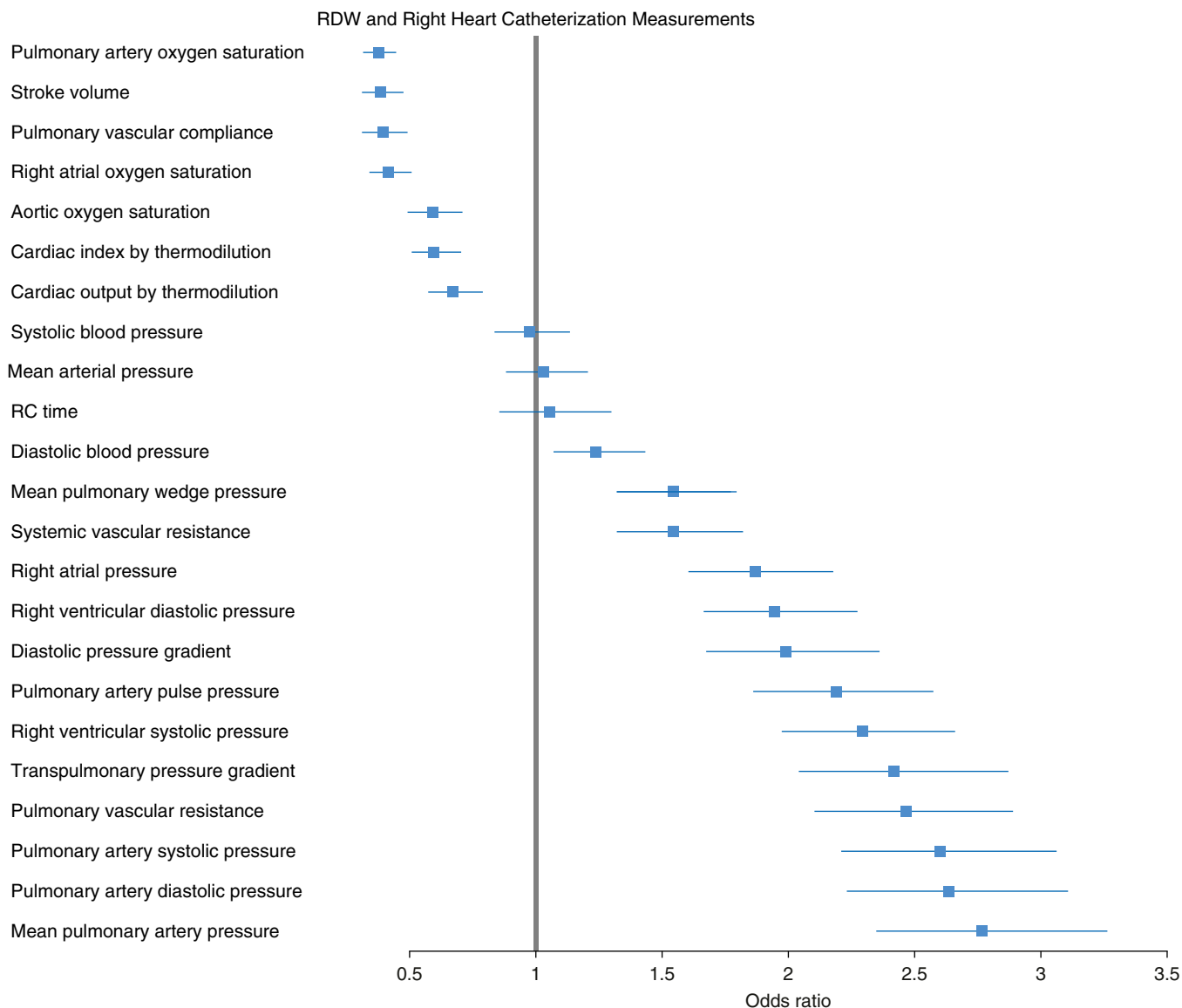


Figure 4. The association of red cell distribution width (RDW) with hemodynamic measurements by right heart catheterization. For each variable, the boxes represent the odds ratio comparing 25th versus 75th percentile of RDW. Horizontal lines represent 95% confidence intervals. Analyses performed by ordinal regression adjusting for age at echocardiogram, sex, and self-declared race. RC = resistance-compliance.

(21 vs. 29 mm Hg at 25th vs. 75th percentile RDW; $P < 1 \times 10^{-7}$). In addition, there was no interaction of pulmonary hypertension group and RDW–mPAP association (P value for interaction, 0.14).

To assess relative mean RDW by heart failure and pulmonary hypertension status, individuals were categorized into four clinical groups by mean pulmonary artery pressure (<25 mm Hg or not) and pulmonary capillary wedge pressure (<15 mm Hg or not). Mean RDW was higher in heart failure without pulmonary hypertension compared with individuals without either heart failure or pulmonary

hypertension and higher still in pulmonary hypertension (pre- or postcapillary) compared with either heart failure or those without either heart failure or pulmonary hypertension (Figure 5, left panel). We next stratified patients with postcapillary pulmonary hypertension to those with isolated postcapillary pulmonary hypertension and those with combined pre- and postcapillary pulmonary hypertension (on the basis of diastolic pressure gradient ≥ 7); mean RDW values were 14.3 ± 1.4 and 15.2 ± 2 , respectively ($P = 0.02$; Figure 5, right panel). These data support the conclusion that RDW is more

strongly associated with hemodynamic evidence of pulmonary vascular disease compared with either postcapillary pulmonary hypertension or left heart failure without pulmonary hypertension. Full tabular results for RDW–catheterization associations are listed in Table E7.

Discussion

In unbiased analyses of codable medical diagnoses in two independent populations, RDW was most strongly associated with

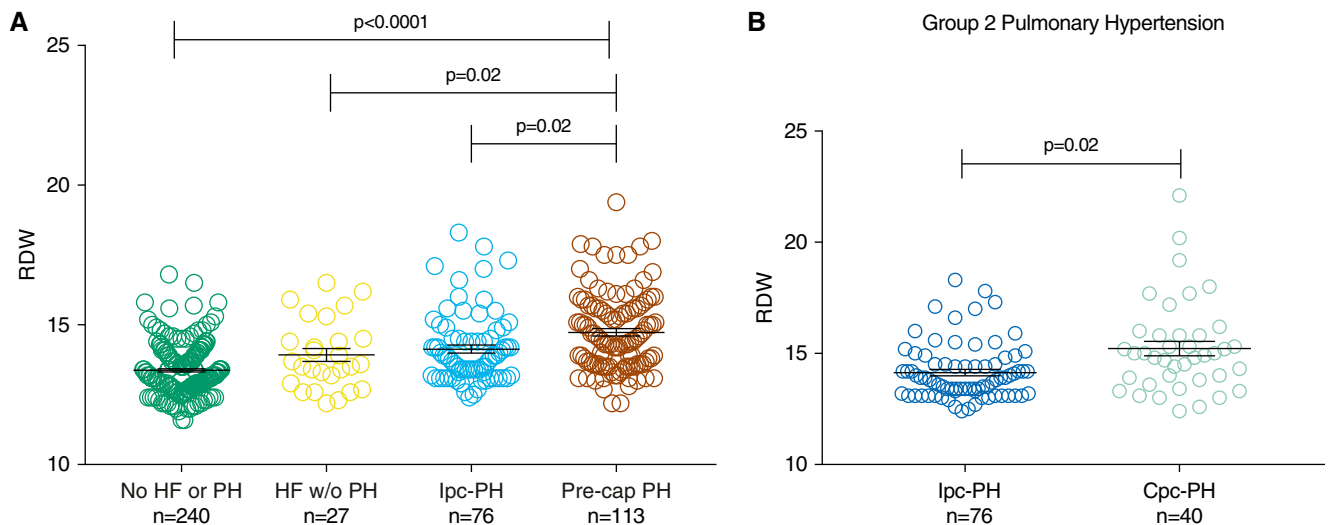


Figure 5. Red cell distribution width (RDW) values by clinical phenotype as determined by right heart catheterization. (A) RDW values in individuals without heart failure (HF, defined by mean pulmonary capillary wedge pressure [mPWP] \geq 15 mm Hg) or pulmonary hypertension (PH, defined by mean pulmonary artery pressure [mPAP] \geq 25 mm Hg) compared with HF without PH, isolated postcapillary PH (lpc-PH: mPWP \geq 15 mm Hg, mPAP \geq 25 mm Hg, diastolic pressure gradient [DPG] $<$ 7 mm Hg, and pulmonary vascular resistance [PVR] \leq 3 Wood units), and precapillary PH (Pre-cap: mPWP $<$ 15 mm Hg, mPAP \geq 25 mm Hg, and PVR \geq 3 Wood units; i.e., group 1 PH). (B) Right: Individuals with postcapillary PH stratified by isolated postcapillary pulmonary hypertension versus individuals with combined pre- and postcapillary PH (Cpc-PH: PVR \leq 3 or $>$ 3 Wood units, respectively, and DPG $<$ 7 or \geq 7 mm Hg, respectively). Left panel statistical analyses were executed using Kruskal-Wallis test ($P < 0.001$) with Dunn's *post hoc* multiple comparisons testing. Right panel analyses performed using Mann-Whitney test. Error bars represent standard error of the mean.

multiple cardiopulmonary conditions, including pulmonary arterial hypertension, congestive heart failure, and chronic pulmonary heart disease. Previously identified RDW–phenotype associations, including obesity, coronary artery disease, chronic kidney disease, and cirrhosis, were replicated in this study, which provides internal validity to our study sample and discovery approach (22, 23, 25, 26). Ordering significant RDW–phenotype associations by magnitude of effect revealed that pulmonary arterial hypertension and chronic pulmonary heart disease were the most clinically relevant associated phenotypes with high RDW. In a cohort of individuals referred for echocardiography, RDW was strongly associated with sequela of long-standing pulmonary hypertension, including elevated RVSP and reduced RV function. Analyses in patients referred for RHC revealed that RDW was associated with higher mPAP and hemodynamic markers of pulmonary vascular remodeling. Thus, increased RDW is associated with clinical, echocardiographic, and hemodynamic assessments of right heart failure and pulmonary vascular remodeling.

The most common cause of right heart failure is pulmonary hypertension and, in our studies, RDW was strongly associated

with both pulmonary arterial hypertension and chronic pulmonary heart disease. Two previous studies identified the prognostic value of RDW in subjects with pulmonary arterial hypertension (27, 28). Rhodes and colleagues systematically studied the prognostic value of five biomarkers (RDW, growth differentiation factor 15, interleukin-6, creatinine, and N-terminal pro-brain natriuretic peptide) among 139 subjects with idiopathic pulmonary arterial hypertension compared with matched healthy control subjects (29). RDW emerged as the biomarker that best predicted survival by receiver operating characteristic analyses in this cohort. No associations between RDW and hemodynamics were identified in that study, which did not include subjects without pulmonary hypertension or subjects with pulmonary hypertension of other etiologies. By contrast, we examined all nonanemic subjects referred for invasive catheterization without respect to pulmonary hypertension status or etiology, which allowed for increased power to identify RDW–hemodynamic relationships irrespective of pulmonary hypertension status. In a cohort of individuals with chronic thromboembolic pulmonary hypertension, RDW was positively associated with pulmonary vascular resistance and negatively associated with

cardiac index in parallel to our findings (29). By extending RDW–hemodynamic analyses to subjects with and without pulmonary hypertension, we identified new RDW–hemodynamic associations with mPAP, pulmonary vascular resistance, and pulmonary artery saturation. Moreover, our echocardiographic analyses detected a strong association with right ventricular dysfunction, which highlights the ability of RDW to report on elevated pulmonary artery pressures and its consequences.

The mechanism by which RDW carries a strong prognostic value across many diseases is unclear. Focusing on the strongest RDW–phenotype association in our study, right heart failure, may provide mechanistic clues as to how RDW is capturing valuable physiologic information. The clinical sequelae of right heart failure include both systemic hypoxia and total body fluid overload. Chronic intermittent hypoxemia is associated with elevated RDW, possibly via intermittent surges in erythropoietin (30). Although fluid overload has not been directly linked to elevated RDW, fluid shifts have been shown to affect RDW values, such as those during the chloride shift that occurs as RBCs pass through the pulmonary vasculature and exchange bicarbonate for chloride (31). In a genome-wide association study of RDW in a

Hispanic population, the strongest associated single-nucleotide polymorphism was in Na-K-Cl cotransporter 1, which is a ubiquitously expressed solute cotransporter responsible for individual cell volume regulation (32). Because RDW can capture such microshifts in volume, it is possible that macroshifts that occur during fluid overload conditions also affect RDW. Previous studies have found an association of RDW with brain natriuretic peptide level (a sensitive biomarker for intravascular volume overload), which we recapitulated in our cohort (data not shown) (33, 34). Additional mechanisms that may be increasing RDW values and pulmonary hypertension severity include subclinical iron deficiency and bone marrow abnormalities, both of which have previously been linked to pulmonary hypertension pathophysiology (35–37). Given our cohort is a clinical population, ascertainment bias in the availability of iron studies precluded controlling for iron status in our analyses. Further mechanistic studies of RDW in the pulmonary hypertension population are warranted.

There are multiple limitations of our work. The data used for our analyses were extracted from an electronic health record

from a single tertiary care hospital network and carry the limitations that come with a single-center study. Also, PheWAS analyses determine phenotypes based on ICD codes, which can misclassify diagnoses and are influenced by the prevalence of disease. In addition, relatively benign conditions will be undercoded in patients with serious diagnoses that will justify inpatient and outpatient billing more robustly. We attempted to mitigate the limitation of ICD coding–based phenotyping and add specificity to our cardiopulmonary findings by testing associations between RDW and echocardiographic and hemodynamic measurements. We would have liked to analyze our echocardiographic cohort by pulmonary hypertension subgroup; however, the specificity of ICD coding for pulmonary hypertension subgroups is low (65% in one study), which precluded this analysis (38). A further limitation of our work is that the populations studied were predominantly of European descent, which may limit the generalizability of our findings to populations not of European descent.

Our unbiased association studies of RDW with phenotypic, echocardiographic, and hemodynamic data in nonanemic individuals provide a framework for clinicians and researchers alike to interpret elevated RDW. Clinically, RDW remains ignored outside of the context of anemia, and these studies demonstrate that clinicians should consider that an elevated RDW value in the absence of anemia may signal the presence of cardiopulmonary dysfunction. These findings are hypothesis generating, as they highlight the potential shared pathophysiology of pulmonary hypertension and elevated RDW. Further work is needed to understand the mechanisms that drive the association between pulmonary vascular remodeling and elevated RDW. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank the Vanderbilt Institute for Clinical and Translational research and its maintenance of Vanderbilt's deidentified electronic health record, the Synthetic Derivative, which contains more than 2.3 million electronic records, with no defined exclusions.

References

- 1 Perlstein TS, Weuve J, Pfeffer MA, Beckman JA. Red blood cell distribution width and mortality risk in a community-based prospective cohort. *Arch Intern Med* 2009;169:588–594.
- 2 Shteinshnaider M, Barchel D, Almozino-Sarafian D, Tzur I, Tsatsanashvili N, Swarka M, et al. Prognostic significance of changes in red cell distribution width in an internal medicine ward. *Eur J Intern Med* 2015;26:616–622.
- 3 Guimarães PO, Sun JL, Kragholm K, Shah SH, Pieper KS, Kraus WE, et al.; MURDOCK Horizon 1 Cardiovascular Study Investigators. Association of standard clinical and laboratory variables with red blood cell distribution width. *Am Heart J* 2016;174:22–28.
- 4 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;52:86–105.
- 5 Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 2010;26:1205–1210.
- 6 Liao KP, Kurreeman F, Li G, Duclos G, Murphy S, Guzman R, et al. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non-rheumatoid arthritis controls. *Arthritis Rheum* 2013;65:571–581.
- 7 Zalawadiya SK, Veeranna V, Panaich SS, Afonso L, Ghali JK. Gender and ethnic differences in red cell distribution width and its association with mortality among low risk healthy United State adults. *Am J Cardiol* 2012;109:1664–1670.
- 8 Conrad ME. Anemia. In: Walker HK, Hall WD, Hurst JW, editors. *Clinical methods: the history, physical, and laboratory examinations*. Boston: Butterworth Publishers; 1990. Chapter 147.
- 9 Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008;84:362–369.
- 10 Mitchell SL, Hall JB, Goodloe RJ, Boston J, Farber-Eger E, Pendergrass SA, et al. Investigating the relationship between mitochondrial genetic variation and cardiovascular-related traits to develop a framework for mitochondrial phenome-wide association studies. *BioData Min* 2014;7:6.
- 11 Verma A, Verma SS, Pendergrass SA, Crawford DC, Crosslin DR, Kuivaniemi H, et al. eMERGE Phenome-Wide Association Study (PheWAS) identifies clinical associations and pleiotropy for stop-gain variants. *BMC Med Genomics* 2016;9:32.
- 12 Choi L, Carroll RJ, Beck C, Mosley JD, Roden DM, Denny JC, et al. Evaluating statistical approaches to leverage large clinical datasets for uncovering therapeutic and adverse medication effects. *Bioinformatics* 2018;34:2988–2996.
- 13 Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet* 2016;17:129–145.
- 14 Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* 2014;30:2375–2376.
- 15 Wells QS, Farber-Eger E, Crawford DC. Extraction of echocardiographic data from the electronic medical record is a rapid and efficient method for study of cardiac structure and function. *J Clin Bioinforma* 2014;4:12.
- 16 Assad TR, Hemnes AR, Larkin EK, Glazer AM, Xu M, Wells QS, et al. Clinical and biological insights into combined post- and pre-capillary pulmonary hypertension. *J Am Coll Cardiol* 2016;68:2525–2536.
- 17 Assad TR, Maron BA, Robbins IM, Xu M, Huang S, Harrell FE, et al. Prognostic effect and longitudinal hemodynamic assessment of

- borderline pulmonary hypertension. *JAMA Cardiol* 2017;2:1361–1368.
- 18 Brittain EL, Pugh ME, Wheeler LA, Robbins IM, Loyd JE, Newman JH, *et al*. Shorter survival in familial versus idiopathic pulmonary arterial hypertension is associated with hemodynamic markers of impaired right ventricular function. *Pulm Circ* 2013;3:589–598.
 - 19 Hoeper MM, Bogaard HJ, Condliffe R, Frantz R, Khanna D, Kurzyna M, *et al*. Definitions and diagnosis of pulmonary hypertension. *J Am Coll Cardiol* 2013;62:D42–D50.
 - 20 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS). Nazzareno Galie, Marc Humbert, Jean-Luc Vachiery, Simon Gibbs, Irene Lang, Adam Torbicki, Gerald Simonneau, Andrew Peacock, Anton Vonk Noordegraaf, Maurice Beghetti, Ardeschir Ghofrani, Miguel Angel Gomez Sanchez, Georg Hansmann, Walter Klepetko, Patrizio Lancellotti, Marco Matucci, Theresa McDonagh, Luc A. Pierard, Pedro T. Trindade, Maurizio Zompatori and Marius Hoeper. *Eur Respir J* 2015;46:903–975.
 - 21 Liu Q, Shepherd BE, Li C, Harrell FE Jr. Modeling continuous response variables using ordinal regression. *Stat Med* 2017;36:4316–4335.
 - 22 Lippi G, Targher G, Montagnana M, Salvagno GL, Zoppini G, Guidi GC. Relationship between red blood cell distribution width and kidney function tests in a large cohort of unselected outpatients. *Scand J Clin Lab Invest* 2008;68:745–748.
 - 23 Hu Z, Sun Y, Wang Q, Han Z, Huang Y, Liu X, *et al*. Red blood cell distribution width is a potential prognostic index for liver disease. *Clin Chem Lab Med* 2013;51:1403–1408.
 - 24 Adamsson Eryd S, Borné Y, Melander O, Persson M, Smith JG, Hedblad B, *et al*. Red blood cell distribution width is associated with incidence of atrial fibrillation. *J Intern Med* 2014;275:84–92.
 - 25 Laufer Perl M, Havakuk O, Finkelstein A, Halkin A, Revivo M, Elbaz M, *et al*. High red blood cell distribution width is associated with the metabolic syndrome. *Clin Hemorheol Microcirc* 2015;63:35–43.
 - 26 Bujak K, Wasilewski J, Osadnik T, Jonczyk S, Kołodziejaska A, Gierlotka M, *et al*. The prognostic role of red blood cell distribution width in coronary artery disease: a review of the pathophysiology. *Dis Markers* 2015;2015:824624.
 - 27 Rhodes CJ, Wharton J, Howard LS, Gibbs JS, Wilkins MR. Red cell distribution width outperforms other potential circulating biomarkers in predicting survival in idiopathic pulmonary arterial hypertension. *Heart* 2011;97:1054–1060.
 - 28 Hampole CV, Mehrotra AK, Thenappan T, Gomberg-Maitland M, Shah SJ. Usefulness of red cell distribution width as a prognostic marker in pulmonary hypertension. *Am J Cardiol* 2009;104:868–872.
 - 29 Wang W, Liu J, Yang YH, Zhai ZG, Wang C, Wang J. Red cell distribution width is increased in chronic thromboembolic pulmonary hypertension. *Clin Respir J* 2016;10:54–60.
 - 30 Yčas JW, Horrow JC, Horne BD. Persistent increase in red cell size distribution width after acute diseases: a biomarker of hypoxemia? *Clin Chim Acta* 2015;448:107–117.
 - 31 Aliberti G, Proietta M, Pulignano I, Di Giovanni C, Tritapepe L, Vercillo G. Respiratory changes in human red cells. *Clin Lab Haematol* 2001;23:361–363.
 - 32 Hodonsky CJ, Jain D, Schick UM, Morrison JV, Brown L, McHugh CP, *et al*. Genome-wide association study of red blood cell traits in Hispanics/Latinos: the Hispanic Community Health Study/Study of Latinos. *PLoS Genet* 2017;13:e1006760.
 - 33 Subhashree AR. Red cell distribution width and serum BNP level correlation in diabetic patients with cardiac failure: a cross-sectional study. *J Clin Diagn Res* 2014;8:FC01–FC03.
 - 34 Fukuta H, Ohte N, Mukai S, Saeki T, Asada K, Wakami K, *et al*. Elevated plasma levels of B-type natriuretic peptide but not C-reactive protein are associated with higher red cell distribution width in patients with coronary artery disease. *Int Heart J* 2009;50:301–312.
 - 35 Yan L, Chen X, Talati M, Nunley BW, Gladson S, Blackwell T, *et al*. Bone marrow-derived cells contribute to the pathogenesis of pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2016;193:898–909.
 - 36 Farha S, Asosingh K, Xu W, Sharp J, George D, Comhair S, *et al*. Hypoxia-inducible factors in human pulmonary arterial hypertension: a link to the intrinsic myeloid abnormalities. *Blood* 2011;117:3485–3493.
 - 37 Soon E, Treacy CM, Toshner MR, MacKenzie-Ross R, Manglam V, Busbridge M, *et al*. Unexplained iron deficiency in idiopathic and heritable pulmonary arterial hypertension. *Thorax* 2011;66:326–332.
 - 38 Wijeratne DT, Lajkosz K, Brogly SB, Loughheed MD, Jiang L, Housin A, *et al*. Increasing incidence and prevalence of World Health Organization groups 1 to 4 pulmonary hypertension: a population-based cohort study in Ontario, Canada. *Circ Cardiovasc Qual Outcomes* 2018;11:e003973.