



# Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension

Anna Ulrich<sup>1</sup>, John Wharton<sup>1</sup>, Timothy E. Thayer<sup>2</sup>, Emilia M. Swietlik<sup>3,4</sup>, Tufik R. Assad<sup>5</sup>, Ankit A. Desai<sup>6</sup>, Stefan Gräf<sup>3,7,8</sup>, Lars Harbaum<sup>1</sup>, Marc Humbert<sup>9,10,11</sup>, Nicholas W. Morrell<sup>3,7</sup>, William C. Nichols<sup>12</sup>, Florent Soubrier<sup>13</sup>, Laura Southgate<sup>14</sup>, David-Alexandre Tréguët<sup>15</sup>, Richard C. Trembath<sup>16</sup>, Evan L. Brittain<sup>2,17</sup>, Martin R. Wilkins<sup>1</sup>, Inga Prokopenko<sup>18,19,21</sup> and Christopher J. Rhodes<sup>1,21</sup>, on behalf of The NIHR BioResource – Rare Diseases Consortium<sup>20</sup>, the UK PAH Cohort Study Consortium<sup>20</sup> and the US PAH Biobank Consortium<sup>20</sup>

@ERSpublications  
**Mendelian randomisation using genetic data from the largest-to-date PAH cohort does not support red cell distribution width or iron deficiency as a cause of PAH, which is important when interpreting iron replacement trials in this condition** <http://bit.ly/2PPaa88>

**Cite this article as:** Ulrich A, Wharton J, Thayer TE, *et al.* Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension. *Eur Respir J* 2020; 55: 1901486 [<https://doi.org/10.1183/13993003.01486-2019>].

**ABSTRACT** Pulmonary arterial hypertension (PAH) is a rare disease that leads to premature death from right heart failure. It is strongly associated with elevated red cell distribution width (RDW), a correlate of several iron status biomarkers. High RDW values can signal early-stage iron deficiency or iron deficiency anaemia. This study investigated whether elevated RDW is causally associated with PAH.

A two-sample Mendelian randomisation (MR) approach was applied to investigate whether genetic predisposition to higher levels of RDW increases the odds of developing PAH. Primary and secondary MR analyses were performed using all available genome-wide significant RDW variants (n=179) and five genome-wide significant RDW variants that act *via* systemic iron status, respectively.

We confirmed the observed association between RDW and PAH (OR 1.90, 95% CI 1.80–2.01) in a multicentre case–control study (cases n=642, disease controls n=15889). The primary MR analysis was adequately powered to detect a causal effect (odds ratio) between 1.25 and 1.52 or greater based on estimates reported in the RDW genome-wide association study or from our own data. There was no evidence for a causal association between RDW and PAH in either the primary (OR<sub>causal</sub> 1.07, 95% CI 0.92–1.24) or the secondary (OR<sub>causal</sub> 1.09, 95% CI 0.77–1.54) MR analysis.

The results suggest that at least some of the observed association of RDW with PAH is secondary to disease progression. Results of iron therapeutic trials in PAH should be interpreted with caution, as any improvements observed may not be mechanistically linked to the development of PAH.

This article has supplementary material available from [erj.ersjournals.com](http://erj.ersjournals.com)

Received: 25 July 2019 | Accepted after revision: 29 Oct 2019

Copyright ©ERS 2020. This version is distributed under the terms of the Creative Commons Attribution Licence 4.0.

## Introduction

Pulmonary arterial hypertension (PAH) is a rare disease with an estimated prevalence of 7–26 cases per million in the developed world [1]. It is characterised by increased pulmonary vascular resistance due to vasoconstriction and structural remodelling of pulmonary arterioles, leading to right ventricular hypertrophy and end-stage right heart failure [2]. Despite increased awareness and new therapeutic options, annual mortality remains ~10% [1]. Approximately 70–80% of heritable PAH and 10–20% of idiopathic PAH patients are known to harbour mutations in the bone morphogenetic protein type II receptor (*BMPR2*) gene [3]. A recent large study of >1000 PAH patients confirmed the prevalence of causal mutations in *BMPR2*, as well as in five other established genes (*TBX4*, *ACVRL1*, *ENG*, *SMAD9* and *KCNK3*), and identified PAH-associated mutations in four new genes (*ATP13A3*, *SOX17*, *AQP1* and *GDF2*), altogether accounting for 23.5% of the cases studied [4]. The rare mutations in all these genes are inherited in an autosomal dominant manner and exhibit reduced penetrance, indicating that other genetic, epigenetic and/or environmental factors influence the development of PAH.

We and others have demonstrated that one factor strongly correlated with survival in PAH is red cell distribution width (RDW) [5, 6]. A recent hypothesis-free phenome-wide analysis indicated PAH, among several disease descriptors, as the most strongly associated with RDW (OR 2.0, 95% CI 1.75–2.4 per % increase in RDW) in a hospital population [7]. RDW (a measure of red blood cell (RBC) size variability in an individual) is part of the full blood count in standard hospital practice and readily available as a biomarker. RDW correlates with iron status biomarkers and high values can signal early stage iron deficiency or iron deficiency anaemia [8]. RDW increases with decreasing iron as available body iron stores fail to meet the iron demand of RBC synthesis, resulting in RBCs of varied size.

Iron deficiency is commonly observed in PAH patients and is under investigation as a therapeutic target [9–14]. An observed correlation between two traits does not necessarily imply that interventions on one trait will change the other and there are numerous examples where false-positive associations have led to unsuccessful randomised controlled trials [15, 16]. Clinical trials are expensive and time-consuming, and recruitment can be challenging, especially in rare diseases. With the growing availability of genetic data in large disease-focused and population-based studies, testing causal relationships between traits of interest has become possible by harnessing the naturally occurring genetic variation in the population. The collection of methods used to test causal relationships using genetic variants is called Mendelian randomisation (MR) [17, 18]. MR has been used successfully to help prioritise intervention and drug targets and to identify causal factors for several diseases [19–23]. In general, candidate drugs with genetic evidence for effectiveness are more successful in drug trials compared to those without such genetic support [24].

It remains to be investigated whether elevated RDW is largely a consequence of PAH or plays a causal role in the condition. The common genetic variation determining RDW levels has been defined in very large population studies [25], with more power than equivalent studies of iron status. This makes RDW a strong candidate for MR analysis and our primary aim was to dissect the epidemiological association between RDW and PAH by assessing causality using MR. We refined the estimate of association between RDW and PAH using biomarker data in 642 PAH cases and >15000 controls with common diseases from the Imperial College Pulmonary Hypertension Biobank, the UK PAH cohort and the Vanderbilt University

---

**Affiliations:** <sup>1</sup>National Heart and Lung Institute, Hammersmith Campus, Imperial College London, London, UK. <sup>2</sup>Vanderbilt University Medical Center, Division of Cardiovascular Medicine, Nashville, TN, USA. <sup>3</sup>Dept of Medicine, University of Cambridge, Cambridge, UK. <sup>4</sup>Pulmonary Vascular Disease Unit, Royal Papworth Hospital NHS Foundation Trust, Cambridge, UK. <sup>5</sup>Williamson Country Medical Center, Franklin, TN, USA. <sup>6</sup>Dept of Medicine, Indiana University, Indianapolis, IN, USA. <sup>7</sup>NIHR BioResource – Rare Diseases, Cambridge, UK. <sup>8</sup>Dept of Haematology, University of Cambridge, Cambridge, UK. <sup>9</sup>Université Paris-Sud, Faculté de Médecine, Université Paris-Saclay, Paris, France. <sup>10</sup>AP-HP, Service de Pneumologie, Centre de référence de l'hypertension pulmonaire, Hôpital Bicêtre, Le Kremlin-Bicêtre, France. <sup>11</sup>INSERM UMR\_S 999, Hôpital Marie Lannelongue, Le Plessis Robinson, France. <sup>12</sup>Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Dept of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. <sup>13</sup>Sorbonne Universités, UPMC Univ. Paris 06, Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR\_S) 1166, Paris, France. <sup>14</sup>Molecular and Clinical Sciences Research Institute, St George's University of London, London, UK. <sup>15</sup>INSERM UMR\_S 1219, Bordeaux Population Health Research Center, University of Bordeaux, Bordeaux, France. <sup>16</sup>Division of Genetics and Molecular Medicine, King's College London, London, UK. <sup>17</sup>Vanderbilt Translational and Clinical Cardiovascular Research Center, Nashville, TN, USA. <sup>18</sup>Dept of Clinical and Experimental Medicine, University of Surrey, Guildford, UK. <sup>19</sup>Dept of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. <sup>20</sup>Members listed in the supplementary material. <sup>21</sup>These authors contributed equally.

**Correspondence:** Christopher J. Rhodes, National Heart and Lung Institute, Medicine, Imperial College London, London, W12 0NN, UK. E-mail: crhodes@imperial.ac.uk

Medical Centre (VUMC) and applied MR to investigate whether being genetically predisposed to higher levels of RDW increases the odds of developing PAH.

## Methods

### *Definition of PAH*

PAH was defined by internationally agreed criteria [26], specifically mean pulmonary artery pressure >25 mmHg, pulmonary vascular resistance >3 Woods units and pulmonary capillary wedge pressure <15 mmHg. Patients with concurrent diseases known to cause pulmonary hypertension were excluded, such that all were considered to have idiopathic, heritable or anorexigen-induced PAH.

### *Data: genetic and phenotype data in contributing studies*

For our analyses we used both individual-level and summary-level data. Individual-level data including clinical laboratory RDW were available for 524 PAH patients from the Imperial College Pulmonary Hypertension Biobank and the multicentre UK PAH cohort, a study that facilitates collaboration and the sharing of data and samples between the specialist pulmonary hypertension centres in the UK ([www.ipahcohort.com](http://www.ipahcohort.com)).

In addition, the hospital population-based VUMC study provided longitudinal clinical laboratory RDW measurements, detailed clinical diagnoses and genome-wide genotype array data (genotyping platform: Illumina MEGAex) for an additional 118 PAH patients and 15889 common disease controls (supplementary table S1). VUMC hosts a collection of electronic medical records linked to genetic data derived from blood collected during routine clinical assessment in outpatient clinics where all patients are shown the consent form during check-in [27] (<https://vict.vanderbilt.edu/pub/biovu/>) The exclusion criteria (supplementary figure S1) and the imputation of genotype array data are described in the supplementary methods.

Summary-level data for both RDW and PAH susceptibility were used in the MR analyses to maximise the sample size and therefore the statistical power. Genetic instruments serving as a proxy for RDW were obtained from the largest-to-date (>170000 individuals) genome-wide association study (GWAS) on haematological traits (here referred to as the “RDW GWAS”) in a population-based study by *ASTLE et al.* [25].

A large proportion of the idiopathic and familial PAH cases from the UK PAH cohort were whole-genome sequenced as part of the UK National Institute for Health Research BioResource (NIHRBR) [4, 28] study. For genetic effects on PAH susceptibility we used a recent study published by our group and others investigating, in the largest-to-date GWAS, the effects of common genetic variation on PAH risk (“PAH GWAS”) involving 11744 individuals, of which 2085 were PAH cases. The results of the PAH GWAS were combined through the meta-analysis of four independent studies, one of which is the NIHRBR, with 847 PAH cases and 5048 rare disease controls. The other major contributing study, the US PAH Biobank (PAHB), used a control population with mixed common diseases recruited from outpatient clinics (694 PAH cases and 1560 controls), while the two smaller studies used population-based controls, with 269 and 275 PAH cases and 1068 and 1983 controls in the Pulmonary Hypertension Allele-Associated Risk and British Heart Foundation Pulmonary Arterial Hypertension studies, respectively [4].

### *Statistical analyses*

To confirm and refine the estimate for the association between RDW and PAH, we combined PAH patients from the Imperial College Pulmonary Hypertension Biobank, the UK PAH Cohort and the VUMC study and compared them to common disease controls from the VUMC recruited in outpatient clinics (supplementary methods and supplementary table S1).

To test for causality between RDW and PAH, we applied a two-sample MR strategy that requires effect estimates for the genetic instrument on the risk factor (here RDW) and the outcome (here PAH) from two non-overlapping datasets (supplementary methods and supplementary figure S2). The genetic instrument for RDW comprised genetic variants associated with RDW levels in the RDW GWAS at a study-specific genome-wide level of significance ( $p < 8.31 \times 10^{-9}$ ). In the PAH GWAS, 179 variants (inclusive of 12 proxy variants with a minimum  $r^2$  of 0.8) out of the 212 independent ( $r^2 < 0.01$ ) RDW quantitative trait loci (QTL) were available after excluding 13 palindromic variants (A/T or C/G) with intermediate allele frequencies (minor allele frequency >45%) to ensure that the effects of the variants for the two traits can be harmonised to the same allele. To obtain the causal estimate, we applied the inverse variance weighted (IVW) [29] and weighted median [30] methods. We assessed heterogeneity between the causal estimates from each QTL using Cochran’s Q-test (supplementary methods).

In the primary MR analysis, we included all available genome-wide significant RDW QTL ( $n=179$ ). The secondary analysis explicitly tested the role of iron deficiency in the RDW–PAH association using five RDW QTL mapped to genes involved in iron homeostasis (*HFE*, *TMPRSS6*, *TFRC* and *TFR2*) from the full set of genome-wide significant RDW QTL (supplementary figure S2). All of these five RDW QTL concomitantly increase serum iron, ferritin and transferrin saturation and decrease transferrin, reflecting systemic iron status (supplementary table S2) and were first reported by an independent GWAS (the Genetics of Iron Status) as genome-wide significant signals for at least two of the aforementioned iron status biomarkers [31]. These five RDW QTL are among the signals which explain the highest proportion of variance in the RDW GWAS, highlighting the importance of iron availability in RDW levels.

Furthermore, we validated the RDW genetic instrument as a proxy for RDW levels in our common disease controls from VUMC. To do so, we regressed the first available RDW measurement on the RDW genetic risk score (GRS) constructed for each individual (supplementary methods) and obtained the proportion of variance explained ( $R^2$ ). The  $R^2$  for the same RDW GRS in the RDW GWAS study populations (UK Biobank (UKB) and INTERVAL) was calculated from the summary statistics of the RDW–QTL associations (supplementary methods).

Since genetic variants typically explain a small proportion of the variability in the associated trait, MR studies often require large sample sizes to detect a non-zero causal effect. Our power to detect a causal association in the current MR analyses was calculated [32] using the  $R^2$  values estimated in VUMC and those estimated in the RDW GWAS populations (supplementary methods).

## Results

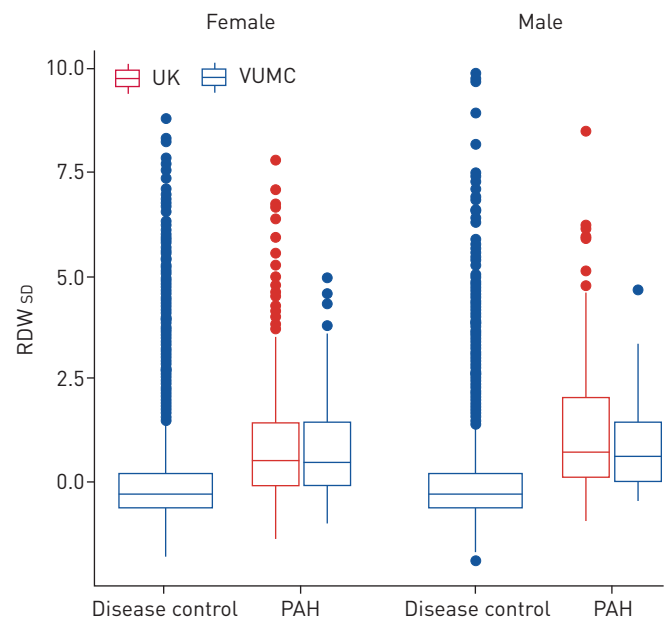
### Defining the association of RDW and PAH

Within this observational analysis, each standard unit (1.4%) increase in RDW was associated with 90% higher odds of prevalent PAH after adjusting for the effects of age and sex (OR 1.90, 95% CI 1.80–2.01). There were no marked between-cohort differences in RDW levels (figure 1 and supplementary table S1).

### Genetic risk score using RDW QTL

We estimated that the 179 RDW QTL would explain  $>12\%$  of the variability ( $R^2$  12.7%, 95% CI 12.32%–12.99%) in RDW levels in the RDW GWAS population (UK Biobank and INTERVAL) in which they were discovered. The RDW GRS constructed for the VUMC controls explained 2.6% (95% CI 2.17–3.19%) of the variability in the first available RDW measurement (supplementary table S3). The set of five variants specific to iron status explain an estimated 1.7% (95% CI 1.62–1.87%) of the RDW variability in the UKB and INTERVAL populations while they explain 0.7% (95% CI 0.43–0.92%) of the total variability in RDW in the VUMC controls (supplementary table S3). The RDW GWAS study populations had a lower mean RDW level than our common disease controls (UKB and INTERVAL 13.45; VUMC 13.60) and lower standard deviation (UKB and INTERVAL 0.82; VUMC 1.40), which could explain in part differences between the  $R^2$  estimates in these studies.

FIGURE 1 Boxplot of red cell distribution width (RDW) levels in the merged cohort of pulmonary arterial hypertension (PAH) cases ( $n=642$ ) and the disease control cohort ( $n=15889$ ). The bottom and the top lines of the box indicate the 25th and 75th percentiles, while the centre line indicates the median value. The whiskers extend to 1.5 times the interquartile range from both ends of the box with individual points being more extreme observations. VUMC: Vanderbilt University Medical Center.



### Statistical power to detect causal effect

With a genetic instrument that explains a relatively high proportion ( $R^2$  12%) of the total variation in RDW levels (figure 2, red curve) we have 80% power to detect a causal effect as small as 1.25 (odds ratio). If the true variance explained lies closer to that estimated in the VUMC controls ( $R^2$  2.6%; figure 2, green curve), this changes to 1.52. When the variance explained by the genetic instrument is small ( $R^2$  1.7%; figure 2, blue curve), we are limited, with our sample size, to an odds ratio of  $\geq 1.7$ . However, if the effect of RDW calculated in our observational analysis (OR 1.80–2.01) was causal in nature, either of the two MR analyses, based on  $R^2$  estimates from the RDW GWAS, would have >80% power to detect an effect of that magnitude. Using the estimates based on the VUMC data, the analysis using all 179 RDW QTL (figure 2, green curve) would have sufficient (>80%) power in our sample, while the iron-specific model (figure 2, purple curve) would be underpowered.

### RDW-PAH causal relationship

We tested for a causal effect of RDW on development of PAH in our primary MR analysis using 179 RDW QTL and found no significant relationship (IVW  $OR_{\text{causal}}$  1.07, 95% CI 0.93–1.23; Q p-value 0.57; figure 3). A secondary MR analysis based on five RDW QTL provided no evidence for a causal effect of iron status on PAH (IVW  $OR_{\text{causal}}$  1.09, 95% CI 0.77–1.54; Q p-value 0.91; figure 3). The weighted median method, which is more robust to violations of MR instrument assumptions, yielded similar estimates for both the primary ( $OR_{\text{causal}}$  1.11, 95% CI 0.89–1.38) and the secondary ( $OR_{\text{causal}}$  1.04, 95% CI 0.68–1.59) MR analyses.

If the odds ratio estimated in the primary MR analysis was indicative of the magnitude of a real causal effect, the number of PAH cases needed to detect such a causal effect with a genetic instrument that explains 10% of the variance in RDW would be  $\sim 20600$  (with the same 1:4.6 ratio of cases:controls as in this study) to achieve 80% power at a false-positive rate of 5% ( $p=0.05$ ). We tested for heterogeneity between causal effects estimated in the four studies contributing to the PAH GWAS separately (supplementary figures S1 and S2) to assess if differences in the nature of their control populations yielded heterogeneous effect estimates for the instrumental variants. The two heterogeneity tests on the IVW

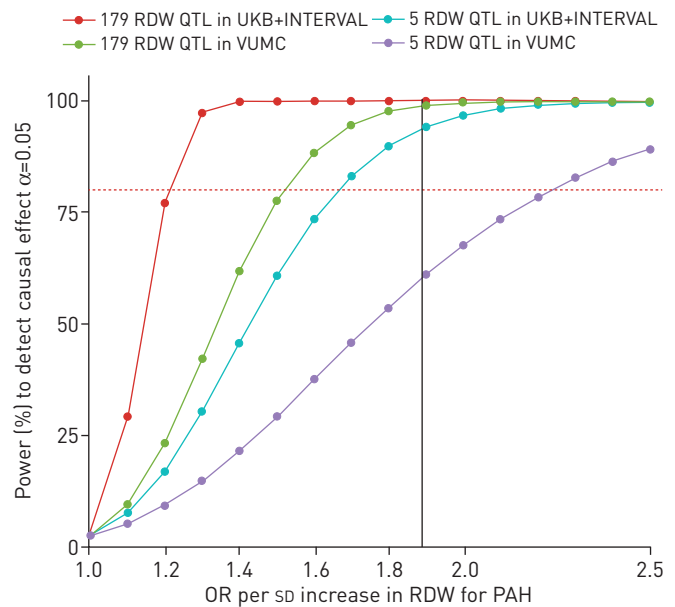


FIGURE 2 Power (%) to detect a causal association (y-axis) given the size of the true underlying causal effect of one standard unit increase in red cell distribution width (RDW) on pulmonary arterial hypertension (PAH) risk (x-axis).  $n=11\,744$  (2085 cases). The effect estimate obtained from the observational study is indicated with the vertical black line at OR 1.90 while the red dotted line marks the desired power of 80%. Red curve: Mendelian randomisation (MR) using all overlapping genome-wide significant variants from the RDW genome-wide association study (GWAS), given the true  $R^2$  12% as per estimated in the UK Biobank (UKB) and INTERVAL cohorts; green curve: MR using all genome-wide significant quantitative trait loci (QTL) from the RDW GWAS, given the true  $R^2$  2.6% as per estimated in our Vanderbilt University Medical Center (VUMC) control cohort; blue curve: MR using five genome-wide significant variants from the RDW GWAS reflecting systemic iron status, given the true  $R^2$  1.7% as per estimated in the UKB and INTERVAL cohorts; purple curve: MR using five genome-wide significant QTL from the RDW GWAS reflecting systemic iron status, given the true  $R^2$  0.7% as per estimated in our VUMC control cohort.

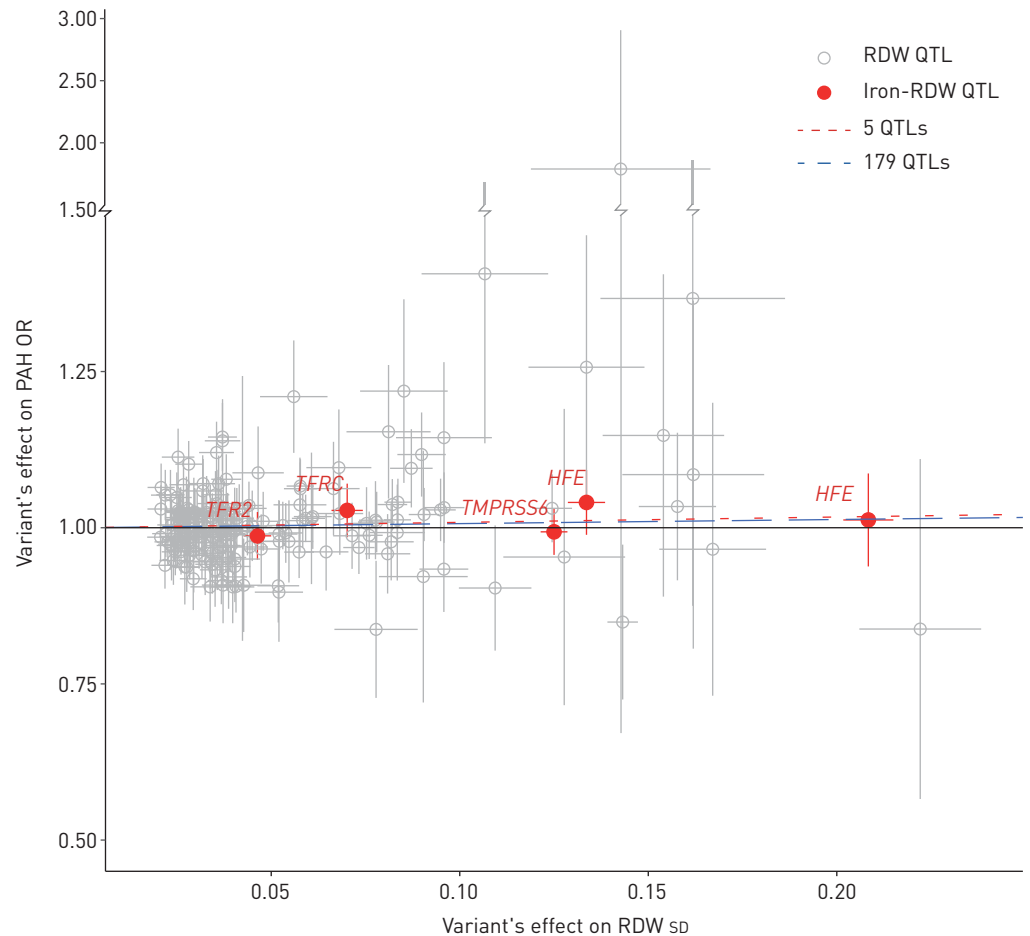


FIGURE 3 Scatterplot of variant (red cell distribution width (RDW)) associations (x-axis) plotted against variant (pulmonary arterial hypertension (PAH)) associations (y-axis) where each dot represents a single RDW quantitative trait locus (QTL). The effect estimates and their standard errors (grey bars) are given in standard units for RDW and in odds ratios for PAH. The solid black line denotes an odds ratio of 1 (no effect), while the dashed blue line is the overall causal effect from the inverse variance weighted regression using all 179 RDW QTL. The five iron-specific RDW QTL used as instruments in the secondary Mendelian randomisation analysis are labelled with their corresponding gene names and the red dotted line denotes the corresponding causal effect.

estimates (main MR:  $Q=0.83$ ,  $df=3$ ,  $p=0.84$ ; secondary MR:  $Q=0.55$ ,  $df=3$ ,  $p=0.91$ ) did not detect considerable variability between the causal effects in the four contributing studies based on a random-effects model (supplementary figures S3 and S4).

## Discussion

We applied a two-sample MR approach to test whether the epidemiological relationship between elevated RDW levels, which are associated with iron deficiency, and PAH is causal in nature. We estimated the effect of RDW on PAH in a large sample of cases and common disease controls. By using genetic variants as instruments for RDW, we found no evidence for a causal effect of RDW on PAH of the magnitude suggested by observational studies.

Previous work has shown that iron deficiency is common in PAH and associated with a poor prognosis, reduced exercise capacity and worsening haemodynamics [9, 11, 13]. A physiological link has been described in healthy volunteers where iron infusion attenuated the rise in pulmonary artery pressure induced by acute hypoxia [33, 34] and in rats where chronic iron deficiency results in pulmonary hypertension [35]. This relationship could be driven by the role of iron in destabilising the hypoxia-inducible factor, thereby deficiency can mimic the hypoxic state [36]. Our study confirmed the association of raised RDW with PAH using controls from a hospital population and cases from multiple centres (with an effect size 1.90 for one standard unit RDW, 1.4%), supporting a recent hypothesis-free phenome-wide analysis which indicated, among several disease descriptors, PAH as the most strongly

associated with RDW with a similar effect size (OR 2.0, 95% CI 1.75–2.4 per % increase in RDW) [7]. Our MR analysis was adequately powered to detect a causal role for RDW with an effect of this magnitude. The fact that we did not detect a causal effect at this level suggests that at least some of the observed association is secondary to the disease.

The results of this study may appear to be at odds with previous clinical studies of the efficacy of iron supplementation in PAH patients, which have focused on functional capacity rather than disease pathology. An open-label study of 20 patients with idiopathic PAH with iron deficiency reported improved iron status, 6-min walk distance and quality of life (QoL) 2 months after a single infusion of 1000 mg ferric carboxymaltose [14]. Another open-label study in 15 iron-deficient idiopathic PAH patients reported improvement of iron status, QoL and exercise endurance capacity on cardiopulmonary exercise testing after receiving 1000 mg of intravenous iron [10]. Neither of the clinical studies were placebo controlled, although VIETHEN *et al.* [14] compared their intervention group to a group of matched iron-replete patients who did not receive iron infusion. It remains possible that iron supplementation in PAH could have benefits through mechanisms distinct from those driving the cardiovascular pathology, for example on muscle function [37].

MR studies using data from large consortia support a causal effect of iron status in other diseases. The genetic instruments (two variants in *HFE* and one variant in *TM6RS6*) used in these studies were also used in our secondary MR analysis. GILL and co-workers found iron to have a protective effect against coronary artery disease (IVW  $OR_{causal}$  0.94 per SD change in serum iron) [38], but increased the risk of cardioembolic stroke (IVW  $OR_{causal}$  1.16 per SD iron) [39]. The authors suggest the opposing effects of iron status on coronary artery disease and stroke might be due to different underlying mechanisms. PICHLER *et al.* [19] have reported that iron protects against the risk of developing Parkinson's disease (IVW  $OR_{causal}$  0.88 per SD iron). Iron deficiency is a common risk factor and these causal effect estimates in common diseases are modest; this study was not powered to detect an effect if this is also true of PAH.

In the light of this MR analysis, alternative explanations for the association of RDW and PAH have to be considered; namely, that PAH causes raised RDW (reverse causation, for example reduced oxygen delivery and/or haemolysis related to PAH may stimulate reticulocytosis, which would increase RDW) or that PAH and elevated RDW are caused by an independent common mechanism. One such mechanism is chronic inflammation, which is a common feature of PAH [40] and leads to intracellular sequestration of iron. Other mechanisms which may modify RDW, such as folate or vitamin B12 could be studied for their association with PAH. To directly test whether PAH is causal for raised RDW levels a stronger genetic instrument for PAH is required than the currently known common variation identified by PAH GWAS.

An important strength of our study lies in the sample size available with phenotype and genetic data achieved through extensive collaboration. Given the rarity of PAH, data had to be pooled from several centres to allow the investigation of common genetic variation in PAH and to test causal relationships. Our study has some limitations. The control population for our observational study was not specifically selected to represent a population at risk of developing PAH. An example of a population at risk of PAH are relatives of patients with pathogenic *BMP2* mutations (and other rare pathogenic variants). Given the reduced penetrance of familial/heritable PAH (~42% in females and ~14% in males carrying known mutations) [41], following-up the families of affected individuals, especially relatives harbouring mutations associated with PAH, would be an invaluable source to identify environmental triggers of PAH.

Despite our efforts to exclude controls with conditions that probably affect RDW and to use the first available RDW measurement, we expect genetic effects of RDW levels to differ between individuals with common diseases and a cohort of healthy volunteers. The  $R^2$  of a genetic variant describes the variance explained in the phenotype in a given population at a given time. Therefore, there can be no single population parameter that applies to multiple populations or the same population at multiple time points. Although the  $R^2$  estimated in the RDW GWAS is likely to be upwardly biased (since it contains the discovery as well as the replication samples), it might reflect better the extent to which genetic variation influences long-term RDW levels in disease-free populations than the  $R^2$  estimate from the VUMC disease controls, who may be expected to have more variable RDW levels due to comorbidities. This highlights a potential challenge in estimating power for MR studies. We now estimate that >20000 PAH patients would be required to detect any likely causal effect of iron deficiency.

It is important to note that in our MR analyses, causal effects were estimated using the results of the PAH GWAS [4]. Four independent studies contributed to the overall results of the PAH GWAS comparing allele frequencies of their PAH cases to control cohorts selected according to different criteria. The NIHRBR study used a control population with a mixture of rare diseases, while the other major contributing study from the United States used mixed common disease controls. The two smaller contributing studies used population-based controls; a preferable design for estimating the effects of

surrogate genetic variants for common risk factors. Selection bias can affect the estimates of the instrumental genetic variants on disease susceptibility. This is especially true if the control cohort is enriched for conditions also affected by the risk factor of interest.

### Conclusions

There is strong observational evidence for an association between elevated RDW, a surrogate for iron deficiency, and PAH. However, this MR analysis does not indicate that RDW is causally linked to disease development. Our study was powered to detect a causal effect similar in size to that observed. A more modest causal effect remains possible, but a significantly larger study population would be required to detect it. Extending international collaborations and careful follow-up of populations at risk will allow increasingly sophisticated study designs to investigate causal relationships, shared underlying mechanisms with other conditions and overall genetic susceptibility in PAH.

**Acknowledgements:** We gratefully acknowledge the participation of patients recruited to the UK National Institute of Health Research BioResource – Rare Diseases (NIHR BR-RD) Study, the UK National Cohort of Idiopathic and Heritable PAH, and the National Institutes of Health/National Heart, Lung, and Blood Institute-sponsored National Biological Sample and Data Repository for PAH (aka PAH Biobank). We thank the physicians, research nurses and coordinators in the UK, Europe and at the 38 pulmonary hypertension centres across the USA involved in the PAH Biobank ([www.pahbiobank.org](http://www.pahbiobank.org)).

**Conflict of interest:** A. Ulrich has nothing to disclose. J. Wharton has nothing to disclose. T.E. Thayer has nothing to disclose. E.M. Swietlik has nothing to disclose. T.R. Assad has nothing to disclose. A.A. Desai has nothing to disclose. S. Gräf has nothing to disclose. L. Harbaum has nothing to disclose. M. Humbert reports grants and personal fees from Bayer and GSK, personal fees from Actelion, Merck and from United Therapeutics, outside the submitted work. N.W. Morrell reports personal fees from Actelion and Morphogen-IX, outside the submitted work. W.C. Nichols has nothing to disclose. F. Soubrier has nothing to disclose. L. Southgate has nothing to disclose. D-A. Trégouët has nothing to disclose. R.C. Trembath reports personal fees for advisory board work from Ipsen Pharmaceuticals, personal fees for non-executive board membership from King's College Hospital NHS Foundation Trust, outside the submitted work. E.L. Brittain reports personal fees for advisory board work from Bayer, outside the submitted work. M.R. Wilkins reports grants from Vifor Pharma, outside the submitted work. I. Prokopenko has nothing to disclose. C.J. Rhodes reports personal fees from Actelion, outside the submitted work.

**Support statement:** The work cited here is supported by funding from the NIHR BR-RD, the British Heart Foundation (SP/12/12/29836), the BHF Cambridge and Imperial Centres of Cardiovascular Research Excellence (RE/18/4/34215), UK Medical Research Council (MR/K020919/1), the Dinosaur Trust, and BHF Programme grants to R.C. Trembath (RG/08/006/25302), N.W. Morrell (RG/13/4/30107) and M.R. Wilkins (RG/10/16/28575). Funding for the PAH Biobank is provided by NIH/NHLBI HL105333. Vanderbilt University Medical Center's BioVU is supported by numerous sources: institutional funding, private agencies, and federal grants that include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at <https://vict.vanderbilt.edu/pub/biovu/>. E.L. Brittain receives funding from the NIH R01 HL146588, American Heart Association Fellow to Faculty Grant (13FTF16070002) and the Gilead PAH Scholars Award Program. The genotyping of the VESPA samples was supported by RC2GM092618. The authors acknowledge use of BRC Core Facilities provided by the financial support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Imperial College NHS Trust, Cambridge University Hospitals and Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust and by NIHR funding to the Imperial NIHR Clinical Research Facility. C.J. Rhodes is supported by a British Heart Foundation Intermediate Basic Science Research Fellowship (FS/15/59/31839). L. Southgate is supported by the Wellcome Trust Institutional Strategic Support Fund (204809/Z/16/Z) awarded to St George's, University of London. I. Prokopenko is supported by the Wellcome Trust (WT205915), and the EU H2020 (DYNAhealth, project number 633595). N.W. Morrell is a British Heart Foundation Professor and National Institute of Health Research (NIHR) Senior Investigator. W.C. Nichols is supported by NIH NHLBI HL105333. A.A. Desai receives support from NIH NHLBI R01HL136603. Funding information for this article has been deposited with the Crossref Funder Registry.

### References

- 1 Hoepfer MM, Humbert M, Souza R, *et al.* A global view of pulmonary hypertension. *Lancet Respir Med* 2016; 4: 306–322.
- 2 Stacher E, Graham BB, Hunt JM, *et al.* Modern age pathology of pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2012; 186: 261–272.
- 3 Evans JDW, Girerd B, Montani D, *et al.* *BMPR2* mutations and survival in pulmonary arterial hypertension: an individual participant data meta-analysis. *Lancet Respir Med* 2016; 4: 129–137.
- 4 Rhodes CJ, Batai K, Bleda M, *et al.* Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *Lancet Respir Med* 2019; 7: 227–238.
- 5 Rhodes CJ, Wharton J, Howard LS, *et al.* Red cell distribution width outperforms other potential circulating biomarkers in predicting survival in idiopathic pulmonary arterial hypertension. *Heart* 2011; 97: 1054–1060.
- 6 Hampole CV, Mehrotra AK, Thenappan T, *et al.* Usefulness of red cell distribution width as a prognostic marker in pulmonary hypertension. *Am J Cardiol* 2009; 104: 868–872.
- 7 Thayer TE, Huang S, Levinson RT, *et al.* Unbiased phenome-wide association studies of red cell distribution width identifies key associations with pulmonary hypertension. *Ann Am Thorac Soc* 2019; 16: 589–598.



- 8 Emans ME, van der Putten K, van Rooijen KL, *et al.* Determinants of red cell distribution width (RDW) in cardiorenal patients: RDW is not related to erythropoietin resistance. *J Card Fail* 2011; 17: 626–633.
- 9 Ruitter G, Lankhorst S, Boonstra A, *et al.* Iron deficiency is common in idiopathic pulmonary arterial hypertension. *Eur Respir J* 2011; 37: 1386–1391.
- 10 Ruitter G, Manders E, Happé CM, *et al.* Intravenous iron therapy in patients with idiopathic pulmonary arterial hypertension and iron deficiency. *Pulm Circ* 2015; 5: 466–472.
- 11 Soon E, Treacy CM, Toshner MR, *et al.* Unexplained iron deficiency in idiopathic and heritable pulmonary arterial hypertension. *Thorax* 2011; 66: 326–332.
- 12 Rhodes CJ, Wharton J, Howard L, *et al.* Iron deficiency in pulmonary arterial hypertension: a potential therapeutic target. *Eur Respir J* 2011; 38: 1453–1460.
- 13 Rhodes CJ, Howard LS, Busbridge M, *et al.* Iron deficiency and raised hepcidin in idiopathic pulmonary arterial hypertension: clinical prevalence, outcomes, and mechanistic insights. *J Am Coll Cardiol* 2011; 58: 300–309.
- 14 Viethen T, Gerhardt F, Dumitrescu D, *et al.* Ferric carboxymaltose improves exercise capacity and quality of life in patients with pulmonary arterial hypertension and iron deficiency: a pilot study. *Int J Cardiol* 2014; 175: 233–239.
- 15 Hennekens CH, Buring JE, Manson JE, *et al.* Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med* 1996; 334: 1145–1149.
- 16 Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; 360: 23–33.
- 17 Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; 23: R89–R98.
- 18 Burgess S, Timpson NJ, Ebrahim S, *et al.* Mendelian randomization: where are we now and where are we going? *Int J Epidemiol* 2015; 44: 379–388.
- 19 Pichler I, Del Greco MF, Gögele M, *et al.* Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. *PLoS Med* 2013; 10: e1001462.
- 20 C Reactive Protein Coronary Heart Disease Genetics Collaboration, Wensley F, Gao P, *et al.* Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ* 2011; 342: d548.
- 21 Interleukin-6 Receptor Mendelian Randomisation Analysis Consortium, Swerdlow DI, Holmes MV, *et al.* The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 2012; 379: 1214–1224.
- 22 Lewis SJ, Araya R, Smith GD, *et al.* Smoking is associated with, but does not cause, depressed mood in pregnancy – a Mendelian randomization study. *PLoS One* 2011; 6: e21689.
- 23 Palmer TM, Nordestgaard BG, Benn M, *et al.* Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts. *BMJ* 2013; 347: f4262.
- 24 King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *BioRxiv* 2019; preprint [https://doi.org/10.1101/513945].
- 25 Astle WJ, Elding H, Jiang T, *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 2016; 167: 1415–1429.
- 26 Galiè N, Humbert M, Vachiery JL, *et al.* 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): Endorsed by Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Heart J* 2016; 37: 67–119.
- 27 Roden DM, Pulley JM, Basford MA, *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008; 84: 362–369.
- 28 Gräf S, Haimel M, Bleda M, *et al.* Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat Commun* 2018; 9: 1416.
- 29 Lawlor DA, Harbord RM, Sterne JA, *et al.* Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; 27: 1133–1163.
- 30 Bowden J, Davey Smith G, Haycock PC, *et al.* Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; 40: 304–314.
- 31 Benyamin B, Esko T, Ried JS, *et al.* Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat Commun* 2014; 5: 4926.
- 32 Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* 2014; 43: 922–929.
- 33 Balanos G, Dorrington K, Robbins P. Desferrioxamine elevates pulmonary vascular resistance in humans: potential for involvement of HIF-1. *J Appl Physiol* 2002; 92: 2501–2507.
- 34 Smith TG, Balanos GM, Croft QPP, *et al.* The increase in pulmonary arterial pressure caused by hypoxia depends on iron status. *J Physiol* 2008; 586: 5999–6005.
- 35 Cotroneo E, Ashek A, Wang L, *et al.* Iron homeostasis and pulmonary hypertension: iron deficiency leads to pulmonary vascular remodeling in the rat. *Circ Res* 2015; 116: 1680–1690.
- 36 Ivan M, Kondo K, Yang H. HIF $\alpha$  targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing. *Science* 2001; 292: 464–468.
- 37 Charles-Edwards G, Amaral N, Sleight A, *et al.* Effect of iron isomaltoside on skeletal muscle energetics in patients with chronic heart failure and iron deficiency. *Circulation* 2019; 139: 2386–2398.
- 38 Gill D, Del Greco MF, Walker AP, *et al.* The effect of iron status on risk of coronary artery disease: a Mendelian randomization study – brief report. *Arterioscler Thromb Vasc Biol* 2017; 37: 1788–1792.
- 39 Gill D, Monori G, Tzoulaki I, *et al.* Iron status and risk of stroke: a Mendelian randomization study. *Stroke* 2018; 49: 2815–2821.
- 40 Hassoun PM. Inflammation in pulmonary arterial hypertension: is it time to quell the fire? *Eur Respir J* 2014; 43: 685–688.
- 41 Machado RD, Pauciulo MW, Thomson JR, *et al.* BMPR2 haploinsufficiency as the inherited molecular mechanism for primary pulmonary hypertension. *Am J Hum Genet* 2001; 68: 92–102.