





Validation of a Genetic Risk Score Combined With Clinical Variables for Predicting Pulmonary Fibrosis in Early Rheumatoid Arthritis

Mikael Brink,¹  Austin Wheeler,²  Bryant R. England,²  and Solbritt Rantapää-Dahlqvist¹ 

Objective. Pulmonary fibrosis (PF) is a severe extra-articular manifestation of rheumatoid arthritis (RA). This study aimed to externally validate a genetic risk score (GRS) and a combined risk score (CRS) for predicting the risk of RA-associated PF in an independent cohort of patients with early RA.

Methods. This study used an inception cohort of 1,118 patients diagnosed with RA from northern Sweden between 1996 and 2016. Clinical data were systematically collected, and genotyping was performed for 12 single-nucleotide polymorphisms (SNPs) associated with idiopathic PF. Statistical analyses, including logistic regression and area under the curve (AUC) assessments, were conducted to evaluate the performance of the GRS and in combination with clinical data as the CRS in predicting RA-PF development.

Results. Of the 1,115 patients with complete data, 60 (5.6%) were diagnosed with PF. PF was significantly associated with age, rheumatoid factor positivity, disease activity, and *MUC5B* (rs35705950) and *FAM13A* (rs2609255) SNPs. The GRS demonstrated a significant association with RA-PF (odds ratio 2.6, 95% confidence interval 1.6–4.5), whereas the CRS exhibited superior performance (AUC 0.75, $P < 0.001$) compared to the GRS alone (AUC 0.62). The combined risk score outperformed the GRS in discriminating RA-PF, indicating its potential utility in clinical practice.

Conclusion. This study provides external validation of the Veterans Affairs Rheumatoid Arthritis Registry interstitial lung disease GRS (VARA-ILD-GRS) and the VARA-ILD-CRS in an RA cohort, demonstrating their generalizability and effectiveness in identifying individuals at high risk for RA-ILD. The findings support the integration of genetic and clinical data in risk stratification models, which could significantly improve screening strategies for patients with RA at risk of developing PF.

INTRODUCTION

Rheumatoid arthritis–associated interstitial lung disease (RA-ILD) is a severe extra-articular manifestation affecting approximately 10% of individuals with RA, leading to significant morbidity and mortality.^{1,2} Despite the identification of several clinical risk factors, such as older age, male sex, smoking history, severe articular disease, and autoantibody seropositivity,^{3,4} particularly in usual interstitial pneumonia,⁵ there remains a lack of standardized tools for risk stratification in RA-ILD. The overlap of genetic risk factors between RA-ILD and idiopathic pulmonary fibrosis (IPF), particularly the *MUC5B* rs35705950 promoter variant, has

demonstrated the potential for genetic risk variants to enhance the prediction of RA-ILD.

Wheeler et al developed and internally validated a clinical and genetic risk score (GRS) for RA-ILD using data from the Veterans Affairs Rheumatoid Arthritis Registry (VARA), a multi-center national US cohort.⁶ Their study demonstrated that a combined clinical and genetic risk model outperformed clinical factors alone in discriminating RA-ILD, highlighting the utility of incorporating genetic information into risk stratification models. This approach aligns with previous findings that genetic variants associated with IPF also contribute to RA-ILD risk or pulmonary fibrosis (PF).^{7–9}

Supported by the Swedish Research Council (2018-02551), King Gustaf V's 80-Year Fund, the Swedish Rheumatism Association, and Umeå University. Dr Wheeler's work was supported by the Rheumatology Research Foundation. Dr England's work was supported by the US Department of Veterans Affairs (CX002203), and the Rheumatology Research Foundation.

¹Department of Public Health and Clinical Medicine, Rheumatology, Umeå University, Umeå, Sweden; ²Division of Rheumatology & Immunology, Department of Internal Medicine, University of Nebraska Medical Center & VA Nebraska-Western Iowa Health Care System, Omaha, Nebraska.

Additional supplementary information cited in this article can be found online in the Supporting Information section (<https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25696>).

Author disclosures and graphical abstract are available at <https://onlinelibrary.wiley.com/doi/10.1002/acr.25696>.

Address correspondence via email to Solbritt Rantapää-Dahlqvist, PhD, MD, at solbritt.rantapaa.dahlqvist@umu.se.

Submitted for publication June 10, 2025; accepted in revised form October 21, 2025.

SIGNIFICANCE & INNOVATIONS

- The study successfully validated a genetic risk score (GRS) and a combined risk score predicting rheumatoid arthritis-associated pulmonary fibrosis.
- Combining a GRS with clinical risk factors performed significantly better than the GRS alone.
- These risk scores can identify at-risk patients with RA-associated interstitial lung disease, improving the selection of patients for screening.

In this study, we aimed to externally validate the GRS developed by Wheeler et al in an independent RA cohort. By using the same genetic loci and clinical data collection methods, we sought to test the robustness and generalizability of the VARA RA-ILD GRS for predicting development of PF in an RA population. With validation of the clinical utility of the GRS in identifying individuals at high risk for RA-ILD, these tools could support targeted screening and early intervention strategies.

PATIENTS AND METHODS

Patients and collection of clinical data. The cohort of patients with early RA (eRA) has previously been used to investigate the associations between selected genotypes and PF.⁸ Briefly, this inception cohort included patients diagnosed with eRA according to the American College of Rheumatology classification criteria.¹⁰ Patients were consecutively included at the time point of RA diagnosis between January 1, 1996, and December 31, 2016, at five rheumatology clinics in northern Sweden. All patients were observed until December 31, 2016, or death, with a mean follow-up time of 9.4 (SD 4.9) years. Clinical data, including the Disease Activity Score in 28 joints (DAS28),¹¹ pharmacological treatments, and smoking history, were systematically recorded at baseline and at 6, 12, 18, and 24 months after diagnosis and subsequently at all clinical visits. The presence of rheumatoid factor (RF) was assessed at baseline using routine laboratory methods, and the presence of anti-cyclic citrullinated peptide (anti-CCP) antibodies was assessed using an anti-CCP2 test (Euro-Diagnostica AB), as previously described.⁸

Genotyping. DNA samples were collected at baseline and stored at -80°C . Genotyping was performed using the Global Screen Assay (Illumina) to analyze 571,151 genome-wide single-nucleotide polymorphisms (SNPs) at deCode Genetics, as previously reported.¹² Information for selected gene variants was extracted from the genotype dataset. The genome-wide association study (GWAS) data used were stored by the National Bioinformatics Infrastructure Sweden at SciLifeLab. The SNP for *MUC5B* (rs35705950) was imputed, and the sequence variants for imputation were identified through whole-genome sequencing of 67,645 individuals of Icelandic, Danish, Norwegian, and Swedish

origin.¹² Samples ($n = 193$) showing quality scores $<90\%$ were reanalyzed using quantitative polymerase chain reaction (qPCR) assays and TaqMan SNP Genotyping Assays (Applied Biosystems) specific for the rs35705950 SNP. For this study, genotype data at 12 loci were extracted: rs35705950 (*MUC5B*), rs2736100 (*TERT*), rs111521887 and rs5743890 (*TOLLIP*), rs2076295 (*DSP*), rs7887 (*EHMT2*), rs2034650 (*IVD*), rs2609255 (*FAM13A*), rs4727443 (*LOC100128334/LOC105375423*), rs11191865 (*OBFC1*), rs1278769 (*ATP11A*), rs6793295 (*LRRC34*), and rs12610495 (*DPP9*). Genetic ancestry analysis was performed for Danish, Swedish, and Norwegian sample sets separately included in the GWAS.¹²

Pulmonary examinations. Radiographs of the lungs were obtained at baseline and during follow-up if patients presented any symptoms of cough, dyspnea, or chest pain or before initiating biologic disease-modifying antirheumatic drugs. High-resolution computed tomography (HRCT) of the chest was performed in cases with signs of pathology on routine plain radiographs or clinically suspected pulmonary involvement. The diagnosis of PF was based on HRCT findings, with reticular patterns, honeycombing, or traction bronchiectasis and occasionally with ground-glass in a few cases being indicative of PF.¹³ In five patients, the diagnosis of PF was already evident and diagnosed on the routine plain radiograph examinations. In this study, no further diagnostic evaluation was performed to separate usual interstitial pneumonia and nonspecific interstitial pneumonia disease. Clinical and pulmonary examinations have previously been presented in detail.⁸

Statistical analysis. Statistical analyses were conducted using R software version 4.4.1 (R Foundation for Statistical Computing).¹⁴ Descriptive data were summarized as proportions, means, or medians. Frequencies were compared using Pearson's chi-square test with Yates' continuity correction, and for continuous data, Student's *t*-test was used. A Mann-Whitney U test was conducted to compare the different GRS between RA-PF and RA without PF. Associations between PF and potential predictors, including genetic markers, were analyzed using logistic regression analyses, presented as odds ratios (ORs) with 95% confidence intervals (CIs). Areas under the curve (AUCs) were compared using DeLong's test for two correlated receiver operating characteristic curves using the pROC library v.1.18.5.¹⁵ For eight individuals, the mean DAS28 value for the first 24 months ($\text{DAS28}_{\text{mean24m}}$) was imputed using predictive mean matching, and ever smoking ($\text{smoking}_{\text{ever}}$) was imputed for 19 individuals using logistic regression by using the mice library v. 3.16.0.¹⁶

Calculation of genetic and clinical risk scores. All risk scores were calculated as previously reported by Wheeler et al.⁶ In brief, the GRS was calculated with each of the five SNPs

represented, assuming autosomal dominant inheritance as 0 (wild type) or 1 (variant allele): VARA-ILD-GRS = $MUC5B \times \ln(2.47) + DSP \times \ln(1.21) + LRR34 \times \ln(1.08) + OBFC1 \times \ln(1.19) + FAM13A \times \ln(0.74)$.

Similarly, the VARA-ILD combined risk score (VARA-ILD-CRS) model was calculated as reported by Wheeler et al, but with small modifications. In the original score, the variable β coefficient (ie, \ln OR) for categorical variables (smoking history, male sex, RF positivity) was multiplied by 1 if the factor was present or 0 if the factor was absent. For age, 69.5 from the individual's current age was subtracted, and the difference was multiplied by the β coefficient, and the individual's mean DAS28 using the C-reactive protein level (DAS28-CRP) was multiplied by the β coefficient. Modifications of this original score were performed for disease activity and age. In our validation cohort, we instead used the mean DAS28 using the erythrocyte sedimentation rate (DAS28-ESR) over the first 24 months after diagnosis to catch the burden of early inflammation, and instead of using age at the most recent encounter, we used age at the time point of diagnosis of RA (continuing to center age by subtracting 69.5). The modified VARA-ILD-CRS (mVARA-ILD-CRS) = $0.9857579 \times \text{GRS} + 0.0249686 \times (\text{age}_{\text{inclusion}} - 69.5) + 0.2992501 \times \text{sex}_{\text{male}} + 0.3963635 \times \text{smoking}_{\text{ever}} + 0.2376407 \times \text{DAS28}_{\text{mean24m}} + 0.610918 \times \text{RF}_{\text{positive}} + (-4.442266)$. The probability was then, as in the original publication, obtained by calculating $\exp(\text{mVARA-ILD-CRS}) / (1 + \exp(\text{mVARA-ILD-CRS}))$.

Ethics. The study complies with the Declaration of Helsinki, and the Regional Ethics Committees at Umeå University, Sweden, approved this study (no. Dnr 2017-432-32M, 2019-02039). The patients provided informed consent to participate.

RESULTS

Patient characteristics and variant allele frequencies associated with RA-PF. Of the 1,118 patients with eRA included in our inception cohort, a GRS could successfully be calculated for 1,115 individuals with complete data, of whom 60 (5.6%) were diagnosed with PF at the time of RA diagnosis or developed PF during follow-up. Variables related to development of PF were age at inclusion ($P < 0.001$), RF positivity ($P < 0.05$), and $\text{DAS28}_{\text{mean24m}}$ ($P < 0.05$) (Table 1). Of the 12 analyzed SNPs, two were significantly associated with RA-PF in unadjusted analyses: *MUC5B* (rs35705950) (OR 3.70 [95% CI 2.18–6.32], $P < 0.001$) and *FAM13A* (rs2609255) (OR 1.91 [95% CI 1.13–3.25], $P < 0.05$) (Supplementary Table 1).

Model performance of GRS in the validation cohort.

The five included SNPs (*MUC5B* [rs35705950], *DSP* [rs2076295], *LRR34* [rs6793295], *OBFC1* [rs11191865], and *FAM13A* [rs2609255]) in the GRS constructed by Wheeler et al⁶

Table 1. Demographic and clinical data of the included patients with eRA, stratified for development of PF*

	eRA with pulmonary fibrosis (n = 60)	eRA without fibrosis (n = 1,058)
Age at RA diagnosis, mean (SD), y	64.8 (10.3) ^a	57.4 (14.1)
Age at PF diagnosis, mean (SD), y	73.9 (9.6)	–
Male sex, n (%)	23 (38.3)	311 (29.4)
Ever smoker, n (%)	43/59 (72.9)	644/1,041 (60.9)
Current smoker, n (%)	11/58 (19.0)	203/1,043 (19.5)
ACPA+, n (%)	43 (80.0)	705 (69.2)
RF+, n (%)	53 (88.3) ^b	770 (72.8)
HLA-SE+, n (%)	23 (63.9)	371 (58.7)
Follow-up time from RA diagnosis until Dec 31, 2016, or death, mean (SD), y	8.7 (4.8)	9.5 (5.0)
BMI, mean (SD) kg/m ²	27.0 (3.7) ^c	26.4 (4.5) ^c
DAS28-ESR (index date), mean (SD)	4.8 (1.4)	4.8 (1.4)
DAS28-ESR AUC _{24m} ^d , mean (SD)	90.0 (18.8) ^e	82.7 (21.5)
Death during study period, n (%)	27 (45.0) ^a	130 (12.3)
Duration of RA at time point of PF diagnosis, mean (min to max), y	5.4 (–0.58 to 15.6)	–
HRCT, n (%)	55/60 (91.7)	–

* ACPA, anti-citrullinated protein/peptide antibody; AUC, area under the curve; BMI, body mass index; DAS28, Disease Activity Score in 28 joints; eRA, early RA; ESR, erythrocyte sedimentation rate; HLA-SE, HLA shared epitope; HRCT, high-resolution computed tomography; IQR, interquartile range; PF, pulmonary fibrosis; RA, rheumatoid arthritis; RF, rheumatoid factor.

^a $P < 0.001$.

^b $P < 0.01$.

^c Data available in 50 and 849 patients, respectively.

^d AUC calculated for the first 24 months after diagnosis.

^e $P < 0.05$.

were analyzed using multivariable logistic regression analyses, with two SNPs continuing to show significant association with RA-PF: rs35705950 (*MUC5B*, OR 3.02 [95% CI 1.95–4.65], $P < 0.001$) and rs2609255 (*FAM13A*, OR 1.82 [95% CI 1.20–2.72], $P < 0.01$).

The GRS calculated according to Wheeler et al showed a significant association with RA-PF in our cohort, with higher values in RA-PF compared to RA without PF, as expected (mean 0.63, median 0.79, range –0.13 to 1.3 and mean 0.41, median 0.27, range –0.3 to 1.3, respectively; $P < 0.01$) (Supplementary Figure 1). The OR for RA-PF increased for each unit of the GRS at 2.6 (95% CI 1.6–4.5; $P < 0.0001$).

Model performance of CRS in the validation cohort.

Univariable logistic regression for the clinical variables showed significant association with RA-PF for age, $\text{DAS28}_{\text{mean24m}}$, and RF positivity ($P < 0.05$ for all) (Table 2). In a multivariable logistic regression analysis including the same clinical variables, the significance remained for age and RF ($P < 0.01$) (data not shown).

Table 2. Multivariable logistic regression models for variables used in the combined risk score for outcome in patients with rheumatoid arthritis with or without pulmonary fibrosis in the validation cohort and in the original cohort according to Wheeler et al^{6*}

Variable	Validation cohort model				Original cohort model	
	Univariable model		Multivariable model		Multivariable model	
	OR (95% CI), <i>P</i> value	β coefficient	OR (95% CI), <i>P</i> value	β coefficient	OR (95% CI), <i>P</i> value	β coefficient
Genetic risk score	2.64 (1.55–4.50), <0.001	0.9719446	2.36 (1.24–4.48), <0.01	0.8576770	2.68 (1.95–3.69), <0.001	0.9857579
Age, per 1-y increase over age 69.5 y	1.05 (1.02–1.07), <0.001	0.0473284	1.04 (1.01–1.07), <0.01	0.0416284	1.03 (1.01–1.04), 0.001	0.0249686
Male sex	1.49 (0.87–2.55), 0.143	0.4008486	1.07 (0.9–3.19), 0.1	0.5298953	1.35 (0.74–2.47), 0.33	0.2992501
Smoking history (ever)	1.56 (0.88–2.78), 0.128	0.4462943	1.40 (0.7–2.82), 0.343	0.3375456	1.49 (1.00–2.20), 0.05	0.3963635
Mean DAS28-ESR during 24 mo/DAS28-CRP	1.42 (1.05–1.9), <0.05	0.3472252	1.42 (0.99–2.02), 0.055	0.3471921	1.27 (1.12–1.44), <0.001	0.2376407
RF positive	2.83 (1.27–6.3), <0.05	1.0409517	2.96 (1.14–7.7), <0.05	1.0859854	1.84 (1.23–2.76), 0.003	0.610918
Model constant	–	–	–	–5.5932484	–	–4.442266

* CI, confidence interval; CRP, C-reactive protein; DAS28, Disease Activity Score in 28 joints; OR, odds ratio; RF, rheumatoid factor.

When also including the GRS as a covariate in the model, the GRS, age, and RF positivity remained significantly predictive of PF ($P < 0.05$) (Table 2).

The GRS was then combined with clinical variables into the VARA-ILD-CRS and validated in our cohort. The CRS was found to be significantly associated with RA-PF development, with higher values in RA-PF compared to RA without PF, as expected (mean 0.12, median 0.12, range 0.03–0.24 and mean 0.076, median 0.064, range 0.0052–0.34, respectively; $P < 0.0001$) (Supplementary Figure 2). The OR for RA-PF using the CRS was 1.72 (95% CI 1.43–2.1, $P < 0.0001$) per 0.05-unit change in the CRS.

The CRS was categorized into quartiles, with the highest quartile (>0.10016) showing an OR for RA-PF of 2.7 (95% CI 1.8–5.0), and that was significantly higher compared to the lowest quartile (<0.04015) ($P < 0.0001$) (data not shown). The CRS discriminated RA-PF with an AUC of 0.75 (95% CI 0.69–0.81), which was significantly better than the GRS showing an AUC of 0.62 (95% CI 0.54–0.69) ($P < 0.001$) (Figure 1). To validate the performance of the CRS, a table was created for different cutoff levels and their sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios (Supplementary Table 2). The optimal cutoff by maximizing the Youden index was ≥ 0.07999 , yielding a sensitivity of 0.789 and a specificity of 0.653. Setting the sensitivity to ≥ 0.9 led to a cutoff of 0.05, as in the derivation study.⁶ Using this cutoff would eliminate 36.1% of the cohort from needing to undergo further testing. In contrast, prioritizing specificity (≥ 0.9) resulted in a sensitivity of 0.316 and a cutoff value of 0.1458237.

Because the components have different scales, we evaluated which variables contributed the most to the CRS. We present the distribution of the contribution for the RA-PF and RA without PF groups separately in Figure 2. The highest average contributions to the CRS were found for DAS28-ESR_{mean24m}, RF, and GRS. In relative terms, the GRS and age variables differed the most between RA-PF and RA without PF.

DISCUSSION

In this study of 1,115 patients with eRA, we externally validated a GRS and a CRS for predicting the risk of RA-PF. The original score was developed for RA-ILD in a male-predominant US veteran RA cohort.⁶ This external validation study evaluated the application of this score in an inception RA cohort from northern Sweden, with a higher proportion of females. Despite the differences in these cohorts, the performances of the GRS and CRS were similar. Thus, this RA-ILD CRS appears to be a promising tool for use to risk stratify RA-ILD and inform RA-ILD screening approaches.

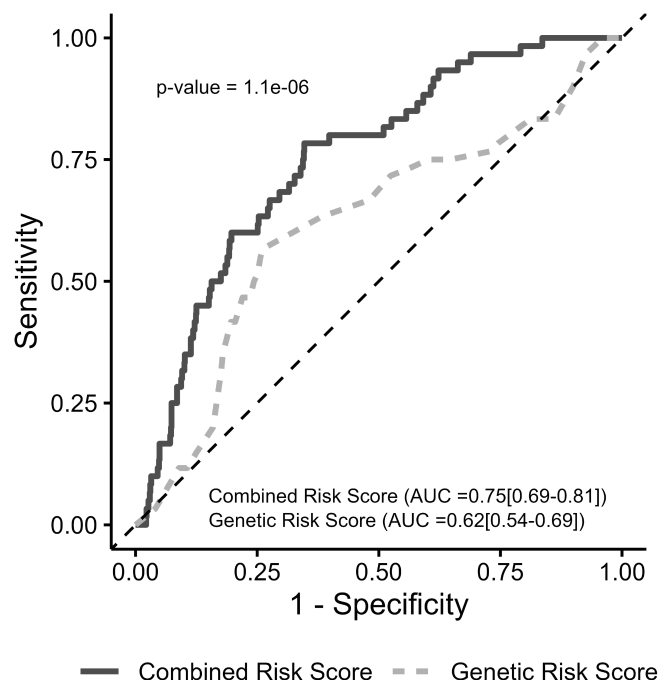


Figure 1. Comparison of receiver operating characteristic curves for combined and genetic risk scores for the presence of rheumatoid arthritis-associated pulmonary fibrosis. AUC, area under the curve.

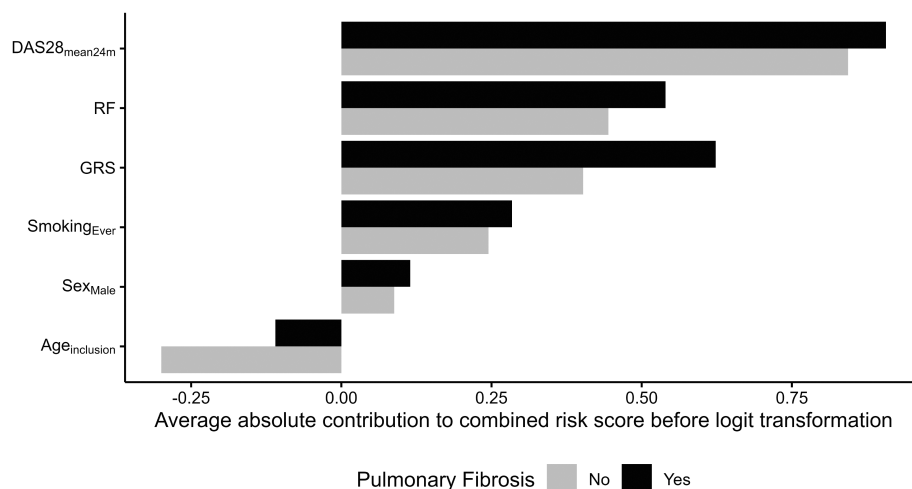


Figure 2. Average contribution of each variable included in the combined risk score, stratified by the presence of pulmonary fibrosis. DAS28, Disease Activity Score in 28 joints; GRS, genetic risk score; RF, rheumatoid factor.

Discrimination of PF in RA with the CRS was higher in our cohort (AUC 0.75 [95% CI 0.69–0.81]) than in the original publication (AUC 0.675 [95% CI 0.639–0.711]), demonstrating the validity of this tool in an alternative study population. As in the original study, we found the CRS to outperform the GRS alone. The similar results are quite remarkable given different frequencies of the genotypes, study populations, and variability of clinical data that were included. Thus, it would be expected for the VARA-ILD-CRS to generalize to other study populations and settings. Despite our patients being of Scandinavian ancestry, there was good agreement between the VARA-ILD-GRS and the VARA-ILD-CRS in RA based on an American population.

We confirmed that the *MUC5B* promotor variant was the strongest genetic risk factor, as was shown in our previous publication.⁸ We also found a significant association with *FAM13A* (rs2609255), which was not significant in the study by Wheeler et al.⁶ This supports the continued need for evaluation of genetic variants among diverse cohorts, as findings for specific SNPs outside of *MUC5B* have varied across populations. This also highlights the value of using a GRS, such as the VARA-ILD-GRS, which allows for a broader representation of our current understanding of genetic risk and moderates the nuanced differences in individual allele frequencies and associations across populations. The performance of the GRS in this study demonstrates the external validity of such approaches.

Polygenic risk scores for IPF, but not RA, were associated with RA-ILD in the COPDgene cohort.¹⁷ In another study evaluation of two large cohorts (MESA and COPDgene), RA-related HLA-DRB1 was not associated with interstitial lung abnormalities.¹⁸

Our cohort was composed of consecutively included patients with eRA between 1996 and 2016, whereas Wheeler et al used a prospective RA cohort of US veterans with more

established disease initiated in 2003. Both cohorts had a high frequency of cigarette smoking history (current cohort 60%–70% vs VARA cohort 77%–85%). Perhaps related to these high cigarette smoking frequencies, we were not able to detect a significant contribution of smoking to RA-PF risk in the multivariable model, despite smoking having previously been established as a risk factor for both RA and RA-PF.^{3,19} Given high background smoking frequency in these cohorts, it is possible that the genetic risk component is more discriminative for RA-PF vs RA alone. Validation of the GRS and CRS in a cohort with less frequent cigarette smoking history would be valuable.

Because of differences in cohorts and data availability, some components of the CRS required modification. Wheeler et al calculated the mean DAS28-CRP preceding ILD diagnosis but also showed consistent performance when using the most recent DAS28-CRP. In adapting the CRS to this cohort, we calculated the DAS28-ESR as a mean value over the first 24 months. This measure of disease activity was similarly associated with RA-PF, suggesting that alternative measures of average disease activity using DAS28 variants are suitable for calculating the CRS. Whether this extends to alternative RA disease activity measures, such as the Clinical Disease Activity Index or Routine Assessment of Patient Index Data 3, and adjustment of these disease activity measure values to maintain consistent CRS values will need to be examined in future studies.

The sensitivity and specificity of the CRS were comparable to those reported previously by Wheeler et al.⁶ At the cutoff of 0.05 proposed previously to ensure a 90% sensitivity for RA-ILD detection, the sensitivity was 93.3% and the specificity was 37.7% in the current cohort. Applying this cutoff would exclude approximately 36% of the patients with eRA in our study from the need for additional testing with HRCT and pulmonary function

tests (PFTs). If providers were comfortable with a test sensitivity of 80%, a cutoff of 0.07 could be used. At this value, the proportion of patients with eRA in our cohort who would avoid the need for additional ILD testing increases to 55%. This illustrates the potential value of an RA-ILD screening strategy that incorporates a highly sensitive VARA-ILD-CRS to identify patients with RA who warrant further ILD testing with HRCT and PFTs.

MUC5B in our cohort was determined from imputation, with a smaller number showing quality scores <90% that were reanalyzed using qPCR assays and TaqMan SNP Genotyping Assays. There was very good agreement between those cases tested by both methods. The sequence variants for the imputation were identified using cases with Scandinavian origin.¹² These findings suggest that imputation is a valid method for ascertaining *MUC5B* status.

A limitation of our study is that the HRCT examinations were not performed randomly or on all included patients. Rather, these were performed among patients with abnormalities on the plain radiographs or who had clinical indications for HRCT with the development of defined symptoms or signs suspicious for RA-PF. The HRCT examinations have been performed over a period of almost 20 years, and methodologic improvements during this time could affect the results. Consequently, we refrained from further diagnostic evaluations besides PF.

In summary, we externally validated the VARA-ILD-GRS and VARA-ILD-CRS among a cohort of patients with eRA based in northern Sweden. Despite differences in the study populations and implementing adaptations to the clinical risk score components, our findings showed a similar ability to discriminate RA-PF and performance metrics at various cut points. These results emphasize the potential role of combined genetic and clinical risk scores to inform RA-ILD screening strategies.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Rantapää-Dahlqvist confirms that all authors have provided the final approval of the version to be published and takes responsibility for the affirmations regarding article submission (eg, not under consideration by another journal), the integrity of the data presented, and the statements regarding compliance with institutional review board/Declaration of Helsinki requirements.

REFERENCES

- Marigliano B, Soriano A, Margiotta D, et al. Lung involvement in connective tissue diseases: a comprehensive review and a focus on rheumatoid arthritis. *Autoimmun Rev* 2013;12(11):1076–1084. doi:<https://doi.org/10.1016/j.autrev.2013.05.001>
- Spagnolo P, Lee JS, Sverzellati N, et al. The lung in rheumatoid arthritis: focus on interstitial lung disease. *Arthritis Rheumatol* 2018;70(10):1544–1554. doi:<https://doi.org/10.1002/art.40574>
- Kelly CA, Saravanan V, Nisar M, et al; British Rheumatoid Interstitial Lung (BRILL) Network. Rheumatoid arthritis-related interstitial lung disease: associations, prognostic factors and physiological and radiological characteristics—a large multicentre UK study. *Rheumatology (Oxford)* 2014;53(9):1676–1682. doi:<https://doi.org/10.1093/rheumatology/keu165>
- Chartrand S, Lee JS, Swigris JJ, et al. Clinical characteristics and natural history of autoimmune forms of interstitial lung disease: a single-center experience. *Lung* 2019;197(6):709–713. doi:<https://doi.org/10.1007/s00408-019-00276-7>
- McDermott GC, Hayashi K, Juge P-A, et al. Impact of sex, serostatus, and smoking on risk for rheumatoid arthritis-associated interstitial lung disease subtypes. *Arthritis Care Res (Hoboken)* 2025;77(2):185–194. doi:<https://doi.org/10.1002/acr.25432>
- Wheeler AM, Baker JF, Riley T, et al. Development and internal validation of a clinical and genetic risk score for rheumatoid arthritis-associated interstitial lung disease. *Rheumatology (Oxford)* 2025; 64(1):268–275. doi:<https://doi.org/10.1093/rheumatology/keae001>
- Juge PA, Lee JS, Ebstein E, et al. *MUC5B* promoter variant and rheumatoid arthritis with interstitial lung disease. *N Engl J Med* 2018; 379(23):2209–2219. doi:<https://doi.org/10.1056/NEJMoa1801562>
- Jönsson E, Ljung L, Norrman E, et al. Pulmonary fibrosis in relation to genetic loci in an inception cohort of patients with early rheumatoid arthritis from northern Sweden. *Rheumatology (Oxford)* 2022;61(3): 943–952. doi:<https://doi.org/10.1093/rheumatology/keab441>
- Klein J, Wheeler AM, Baker JF, et al. *MUC5B* Promoter variant and survival in rheumatoid arthritis-associated interstitial lung disease. *Rheumatology (Oxford)* 2025;64(SI):SI47–SI54. doi:<https://doi.org/10.1093/rheumatology/keae615>
- Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31(3):315–324. doi:<https://doi.org/10.1002/art.1780310302>
- Prevoo MLL, van 't Hof MA, Kuper HH, et al. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38(1):44–48. doi:<https://doi.org/10.1002/art.1780380107>
- Saevarsdottir S, Stefansdottir L, Sulem P, et al; Members of the DBDS Genomic Consortium; Danish RA Genetics Working Group; Swedish Rheumatology Quality Register Biobank Study Group (SRQb). Multiomics analysis of rheumatoid arthritis yields sequence variants that have large effects on risk of the seropositive subset. *Ann Rheum Dis* 2022;81(8):1085–1095. doi:<https://doi.org/10.1136/annrheumdis-2021-221754>
- Raghu G, Remy-Jardin M, Myers JL, et al; American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Society. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018;198(5):e44–e68. doi:<https://doi.org/10.1164/rccm.201807-1255ST>
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2022.
- Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12(1):77. doi:<https://doi.org/10.1186/1471-2105-12-77>
- van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;45(3):1–67. doi:<https://doi.org/10.18637/jss.v045.i03>
- McDermott GC, Moll M, Cho MH, et al. Polygenic risk scores for rheumatoid arthritis and idiopathic pulmonary fibrosis and associations with RA, interstitial lung abnormalities, and quantitative interstitial

- abnormalities among smokers. *Semin Arthritis Rheum* 2025;72:152708. doi:<https://doi.org/10.1016/j.semarthrit.2025.152708>
18. Kim JS, Flack KF, Malik V, et al. Genomic and serological rheumatoid arthritis biomarkers, *MUC5B* promoter variant, and interstitial lung abnormalities. *Ann Am Thorac Soc* 2025;22(1):64–71. doi:<https://doi.org/10.1513/AnnalsATS.202403-238OC>
19. Klareskog L, Stolt P, Lundberg K, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 2006;54(1):38–46. doi:<https://doi.org/10.1002/art.21575>