AMERICAN COLLEGE
of RHEUMATOLOGY
Empowering Rheumatology Professionals

Using Polygenic Risk Scores to Aid Diagnosis of Patients With Early Inflammatory Arthritis: Results From the Norfolk Arthritis Register

Ryan M. Hum,[1] Seema D. Sharma,[1] Michael Stadler,[1] Sebastien Viatte,[1] Pauline Ho,[1] Nisha Nair,[1] Chenfu Shi,[1] Chuan Fu Yap,[1] Mehreen Soomro,[1] Darren Plant,[1] Jenny H. Humphreys,[1] Alexander MacGregor,[2] Max Yates,[2] Suzanne Verstappen,[1] Anne Barton,[1] and John Bowes,[1] on behalf of all NOAR collaborators

Objective. There is growing evidence that genetic data are of benefit in the rheumatology outpatient setting by aiding early diagnosis. A genetic probability tool (G-PROB) has been developed to aid diagnosis has not yet been tested in a real-world setting. Our aim was to assess whether G-PROB could aid diagnosis in the rheumatology outpatient setting using data from the Norfolk Arthritis Register (NOAR), a prospective observational cohort of patients presenting with early inflammatory arthritis.

Methods. Genotypes and clinician diagnoses were obtained from patients from NOAR. Six G-probabilities (0%–100%) were created for each patient based on known disease-associated odds ratios of published genetic risk variants, each corresponding to one disease of rheumatoid arthritis, systemic lupus erythematosus, psoriatic arthritis, spondyloarthropathy, gout, or "other diseases." Performance of the G-probabilities compared with clinician diagnosis was assessed.

Results. We tested G-PROB on 1,047 patients. Calibration of G-probabilities with clinician diagnosis was high, with regression coefficients of 1.047, where 1.00 is ideal. G-probabilities discriminated clinician diagnosis with pooled areas under the curve (95% confidence interval) of 0.85 (0.84–0.86). G-probabilities <5% corresponded to a negative predictive value of 96.0%, for which it was possible to suggest >2 unlikely diseases for 94% of patients and >3 for 53.7% of patients. G-probabilities >50% corresponded to a positive predictive value of 70.4%. In 55.7% of patients, the disease with the highest G-probability corresponded to clinician diagnosis.

Conclusion. G-PROB converts complex genetic information into meaningful and interpretable conditional probabilities, which may be especially helpful at eliminating unlikely diagnoses in the rheumatology outpatient setting.

INTRODUCTION

Given that there has been and continues to be significant investment in the study of human genetics and genomics,[1] it is important that we fully harness the potential value of genetic data in the clinical setting. With initiatives such as the "UK Newborn Genomes Programme," which will whole-genome sequence up to 200,000 newborn babies,[2] and the popularity of direct-to-consumer genotyping companies,[3,4] an ever increasing proportion of the population will be genotyped, meaning that these data will be increasingly available with no added cost. Coupled with the fact that the costs associated with whole-genome genotyping have reduced over time, it is now feasible to have genetic data available for patients during patient appointments.[3,4]

Misdiagnosis of patients first presenting to rheumatology clinic with suspected early inflammatory arthritis can delay treatment and increase the risk of irreversible disability, comorbidity, and death.[5,6] The majority of patients seen in rheumatology outpatient clinics with early inflammatory arthritis will be diagnosed with rheumatoid arthritis (RA),[7] psoriatic arthritis (PsA),[8] systemic lupus erythematosus (SLE),[9] ankylosing spondylitis (SpA),[10] or gout.[11] Furthermore, it is known that these rheumatologic diseases have a strong genetic basis with heritability estimates of approximately 60% for RA and PsA,[12,13] 90% for SpA,[14] 65% for SLE,[15] and 30% for gout,[16] with numerous risk loci identified for each disease.[17–21]

Polygenic risk scores (PRSs) are an estimate of an individual's genetic predisposition to a disease or trait and are typically calculated from the sum of risk alleles carried by the individual weighted by the effect estimate of the risk allele.[22] Although using genetics on a population level has limited value given the low prevalence of rheumatic diseases, using genetics to aid diagnosis in a rheumatology clinic, where all patients will have symptoms and thus have an increased pretest probability for disease, may make probabilistic predictions based on genetics more effective.[23,24]

Based on this premise, a tool to aid diagnosis using existing knowledge of genetic risk loci resulted in the development of a genetic probability tool (G-PROB).[25] G-PROB converts PRSs derived from known risk variants into easy-to-interpret conditional probabilities (G-probabilities) for multiple diseases, assuming that one of the diseases is present.[25] In principle, patients would be genotyped before their first rheumatology clinic visit, and the clinician would be provided with G-probabilities adding up to 100%, indicating a breakdown of which diagnoses are most likely based on the presence of known disease-associated variants. These G-probabilities would then be interpreted, with diseases with low G-probabilities (eg, <5% or <20%) suggesting unlikely diagnoses and diseases with high G-probabilities (eg, >20% or >50%) suggesting a likely diagnosis. This information could complement existing investigations and tools, such as clinical history and physical examination, to aid and improve the accuracy of the first provisional diagnosis.

The authors who developed G-PROB tested the tool in various settings. These initially included applying G-PROB to simulated cohorts of various sizes and disease prevalence and finally on a constructed cohort of 243 patients from a biobank who had presented to a rheumatology outpatient clinic with suspected synovitis using classification criteria to define cases. They concluded that G-probabilities created by G-PROB were well calibrated with diagnosis based on classification criteria, with high negative predictive value (NPV) and modest positive predictive value (PPV).[25] However, G-PROB has not been tested on a real-world cohort of patients from a rheumatology outpatient clinic with real-world disease prevalence, nor has its performance using clinician diagnosis been assessed. Before conducting prospective studies of the G-PROB tool, it is necessary to assess its performance in its intended setting by comparing it with clinician diagnosis in patients presenting to a rheumatology outpatient clinic with early inflammatory arthritis. The aim of our study was to assess G-PROB's ability to aid clinician diagnosis of patients who present with early inflammatory arthritis using prospective data from a large observational cohort in which the clinical heterogeneity and disease prevalence reflect that seen in a typical rheumatology clinic.

METHODS

The Norfolk Arthritis Register. The Norfolk Arthritis Register (NOAR) is a primary care-based inception cohort of patients with early inflammatory arthritis, defined as a minimum of two swollen joints for a period of at least 4 weeks.[26] Patient recruitment began in 1989, and in this data set, follow-up continued for up to 20 years. All patients were from the UK and had self-reported White ancestry. Ethical approval was granted by the Norwich Research Ethics committee, and all patients were recruited after giving informed consent (NOAR - REC Ref 2003/075, December 18, 2003, Norwich Local Research Committee [NHS]).

Clinical diagnosis. Case note review was undertaken for all patients in the NOAR cohort, and the clinical diagnosis made by a consultant rheumatologist at the last available follow-up was recorded. Patients with a documented clinical diagnosis were included in the analysis. Patients with missing case notes were excluded (Figure 1).

The G-PROB tool and calculating G-probabilities. The full method by which G-PROB operates can be found in the original study.[25] Briefly, G-PROB generates G-probabilities, which each correspond to a disease probability based on a weighted genetic risk score. This risk score considers the presence of known genetic variants associated with disease susceptibility and estimated disease prevalence and also assumes the patient has one of the diseases in the model. In theory, G-PROB can be
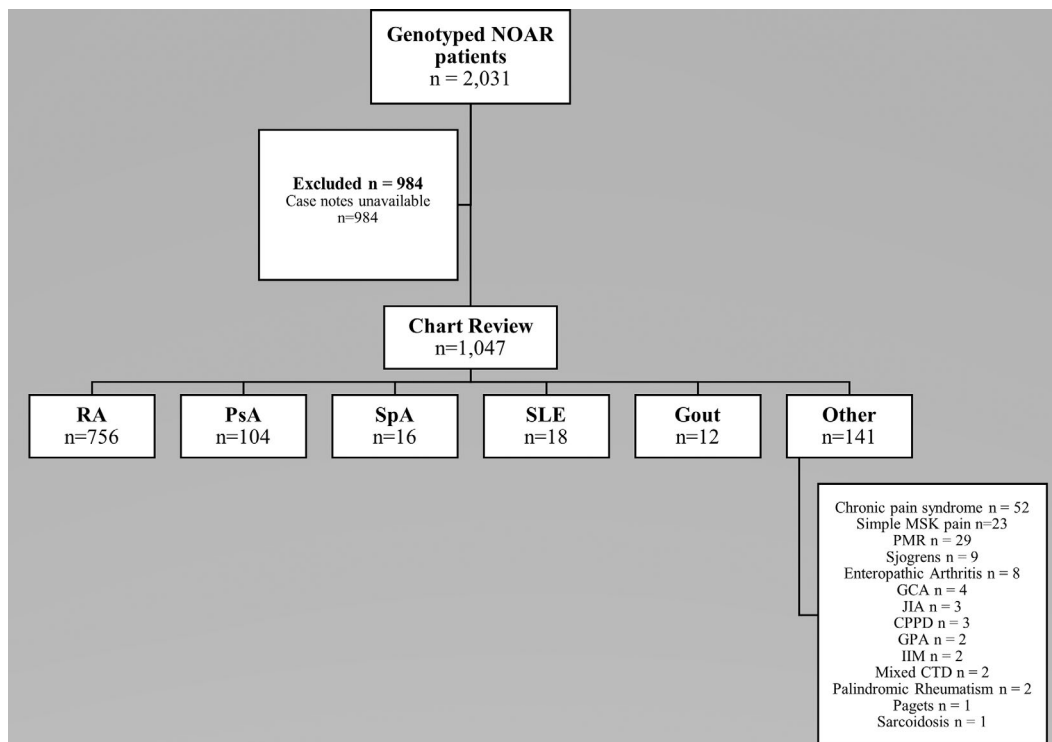
Figure 1. Flowchart of patient selection. CCP, cyclic citrullinated peptide; CPPD, crystal pyrophosphate deposition disease; CTD, connective tissue disease; GCA, giant cell arteritis; GPA, granulomatosis with polyangiitis; IIM, idiopathic inflammatory myopathy; JIA, juvenile idiopathic arthritis; NOAR, Norfolk Arthritis Register; PMR, polymyalgia rheumatica; PsA, psoriatic arthritis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SpA, ankylosing spondylitis.

used to discriminate any sets of diseases with known genetic risk variants. G-PROB requires three inputs to generate G-probabilities: 1) estimated disease prevalence, 2) odds ratios (ORs) of susceptibility for known genetic variants, and 3) genotypes including presence of known genetic variants.

Genotyping and imputation. All samples with sufficient DNA available were genotyped using the Illumina Infinium CoreExome genotyping array. This was performed in accordance with the manufacturer's instructions, in which genotype calling was performed using the GenCall algorithm in the GenomeStudio Data Analysis software platform (Genotyping Module version 1.8.4). Non-HLA imputation was performed using the Michigan Imputation server, in which phasing was performed with Shapeit2, and imputation was performed with the haplotype reference consortium panel. After imputation, single-nucleotide polymorphisms (SNPs) were excluded based on a minor allele frequency of <0.01 and imputation accuracy of $r^2 < 0.5$.

HLA-DRB1 typing. A semiautomated, reverse dot-blot method was used for HLA typing.[27,28] For HLA-DRB1, amino acids at position 11, 71, and 74 were determined, and a four-digit HLA typing corresponding to the full and unambiguous amino acid sequence of the HLA protein was assigned.[28] The estimated association with RA for each of the 16 possible HLA-DRB1

haplotypes have been described previously (Supplementary Table 1).[18,29]

Disease-associated genetic variants. Known genetic risk variants associated with one of five rheumatologic diseases (RA, PsA, SLE, SpA, or gout) at the genome-wide significance level ($P < 5 \times 10^{-8}$) from the most recent genome-wide studies on people of European descent were collated.[17–21,25] Disease-specific ORs for each variant for each of the diseases (except gout) were determined, in which different ORs denote variants that contributed to the susceptibility of several diseases and ORs of 1.0 denote variants that are not associated with a particular disease (Supplementary Table 2).[25] For gout, known genetic variants associated with serum urate concentrations were used to calculate ORs for susceptibility to gout as described previously.[21,25] A total of 208 SNPs outside the HLA region and 42 HLA variants were included in the G-PROB settings for this analysis (Supplementary Table 2).

"Other disease" category. To account for patients with suspected inflammatory arthritis who have a condition other than the five most common aforementioned rheumatologic diseases, a sixth category entitled "Other disease" was devised by the creators of the G-PROB tool.[25] The "Other disease" cohort will include a proportion of patients presenting to rheumatology clinic

that will not have one of the five most common rheumatologic conditions described and will have a variety of other conditions, including chronic pain syndromes (such as fibromyalgia) and simple musculoskeletal injuries, for which there are limited known genetic variants. The ORs for the genetic risk scores of the "Other disease" group will be set to 1.0 to reflect that no data on the genetic risk profile are available for these patients.

Prevalence settings. G-PROB combines the weighted genetic risk score of each disease with an intercept to ensure that the mean probability is equal to the predefined disease prevalence. To enable comparison with the initial study that developed and tested G-PROB, the same estimates of outpatient clinic disease prevalence were used (Supplementary Table 3).[25,30,31] The disease prevalence within the NOAR cohort was also calculated and G-PROB was tested using these disease prevalence settings as well for comparison (Supplementary Table 3).

Statistical analysis. All analyses were performed in R version 4.2.[32] Six G-probabilities were generated for each patient, with one G-probability corresponding to the clinician diagnosis (denoted as a match) and the other five not (denoted as nonmatch). All G-probabilities for all patients were combined into one vector, and a corresponding binary disease match vector indicating clinician diagnosis match or nonmatch was created, with one of the six of the G-probabilities being "matched" and five of six being "nonmatched."

G-PROB performance was assessed in three ways as described previously.[25] First, density plots indicating distribution of G-probabilities that were matched compared with those that were nonmatched were created, for which good performance would find that matched G-probabilities were higher and distributed to the right compared with nonmatched G-probabilities. Mean matched and nonmatched G-probabilities were compared using an independent samples t-test. Second, calibration with clinician diagnosis match was determined by performing a linear regression without intercept, in which G-probabilities were the independent variable and binary disease match was the dependent variable. A regression coefficient (β) was determined for which the ideal calibration would have $\beta = 1.0$.[33] Third, the ability of G-probabilities to correctly classify clinician diagnosis was assessed using the area under the curve (AUC)-receiver operating curve (ROC) of multiclass classifications, in which higher AUCs (from 0.5 to 1.0) indicate better classification.[34] As described previously, AUCs were determined using two-vector data summaries of G-probabilities and clinician diagnosis match/nonmatch (known as micro-AUCs); however, macro-AUCs were also determined for comparison.[25,34,35] Fourth, performance of G-probabilities in suggesting likely and unlikely diagnoses was assessed by determining the NPVs and PPVs at several G-probability thresholds.

G-probabilities can range from 0% to 100%, for which a G-probability of 16.7% for all diseases would suggest all diseases are equally probable. Therefore, to further assess the discriminative abilities of different G-probability thresholds, arbitrary cutoffs of <5% and <20% for NPV and cutoffs of >20% or >50% for PPV, defined in the study in which G-PROB was developed, were assessed.[25] The number of patients for which at least one, two, or three diseases could be considered less likely based on a G-probability threshold of <5% was determined, and the number of patients for which the highest G-probability corresponded to the clinician diagnosis was determined. Finally, scatterplots were created to investigate the PPV and NPV at different G-probability thresholds.

Patient and public involvement. Patients and the public were not specifically involved in this study. Deidentified data are available after reasonable request and ethical approval to the corresponding author.

RESULTS

Patient cohort characteristics and case assignment. From NOAR, 1,047 patients with genotype data and clinician diagnosis following case note review were included in the study (Figure 1). Clinical characteristics of patients according to clinician diagnosis were noted as well as for those with and those without retrievable clinical case notes (Supplementary Table 4).

G-probabilities by disease and distribution for correct and incorrect diagnoses. Six G-probabilities were created for each patient, resulting in 6,282 G-probabilities, of which 1,047 corresponded to clinician diagnosis and 5,235 did not (Table 1). The mean of G-probabilities matching clinician diagnosis was significantly higher than the G-probabilities not matching clinician diagnosis (12.2% vs 39.1%, $P < 2.2 \times 10^{-16}$). Furthermore, the distribution of G-probabilities matching clinician diagnosis was clearly skewed to the right toward higher probabilities compared with the G-probabilities not matching clinician diagnosis (Figure 2). Distributions of G-probabilities for each disease group were also determined (Supplementary Figure 1).

Concordance of G-probabilities with clinician-defined disease. Estimation of the regression line between G-probabilities and clinician diagnosis match using linear regression, constraining the intercept to zero, demonstrated that the magnitude of G-probabilities was well calibrated with clinician diagnosis match status ($\beta = 1.047$, where $\beta = 1.00$ would indicate perfect test performance). It also showed that G-PROB is well calibrated for its intended use, that is, to produce G-probabilities that increase if the clinician diagnosis is more likely (Figure 3).

ROC analysis. Micro-AUC analysis found that the ability of G-PROB to discriminate among different diseases was high, with an AUC of 0.849 (95% confidence interval 0.835–0.863) (Figure 4). ROC analysis on the individual disease level as well as macro-AUC analysis was determined and showed that

Table 1. Performance of G-probabilities in suggesting likely and unlikely diagnoses at different thresholds*

G-probability threshold	Patients with at least one G-probability at the given threshold (n = 1,047), n (%)	G-probabilities at the given threshold (n = 6,282), n (%)	NPV or PPV at the given threshold, %
Thresholds suggesting unlikely diagnoses			
<5%	1,047 (100)	2,597 (41.3)	NPV 96.0
<20%	1,047 (100)	4,159 (66.2)	NPV 95.6
Thresholds suggesting likely diagnosis			
>20%	1,047 (100)	2,123 (33.8)	PPV 40.7
>50%	511 (48.8)	510 (8.1)	PPV 70.3

* Two lower G-probability thresholds (<5% and <20%) for suggesting unlikely diagnoses and two upper G-probability thresholds (>20% and >50%) for suggesting likely diagnosis were evaluated. Each patient has six G-probabilities (one for each disease) generated. The number of patients with at least one G-probability at the given threshold are displayed out of 1,047 patients, as well as the number of G-probabilities overall at the given threshold out of 6,282 G-probabilities generated (6 per patient). NPV, negative predictive value; PPV, positive predictive value.

G-probabilities had low to moderate discriminative capabilities (Supplementary Figure 2).

NPV and PPV of G-probabilities. We found that 41.3% of the G-probabilities (n = 2,597 of 6,282) were <5%, corresponding to an NPV of 96.0% (Table 1). At the <5% threshold, it was possible to suggest that at least one disease was unlikely for 100% of patients (n = 1,047 of 1,047), at least two diseases were unlikely for 94% of patients (n = 984 of 1,047), and three or more diseases were unlikely for 53.7% of patients (n = 562 of 1,047).
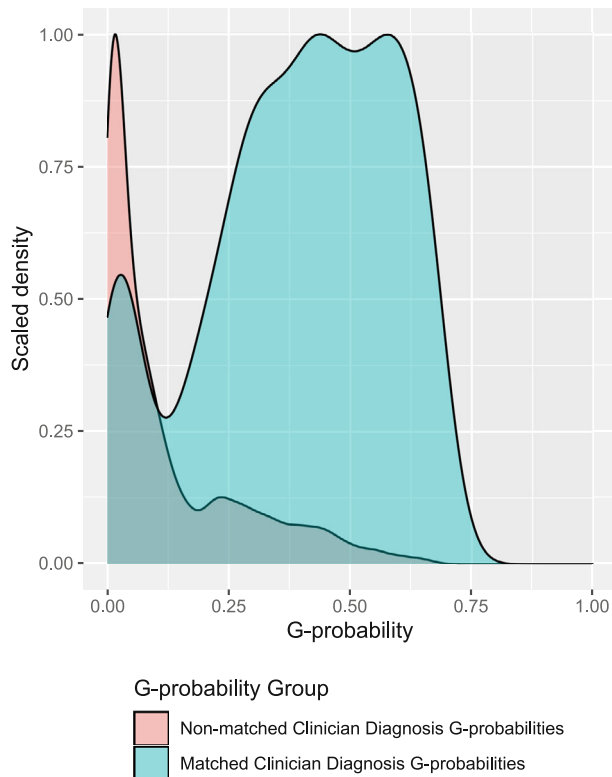
We found that 48.8% of patients (n = 511 of 1,047) had a single G-probability >50%, corresponding to a PPV of 70.3%, whereas using a G-probability threshold of >20% resulted in a lower PPV at 40.7% (Table 1). In 70.3% of patients (n = 359 of 511), the disease with the highest G-probability corresponded to clinician diagnosis. Scatterplots were created to investigate changes
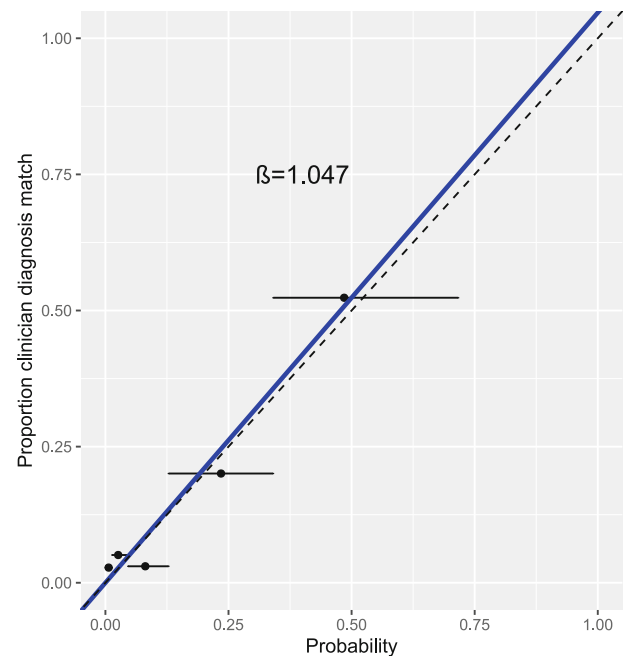

Figure 3. Linear regression model without intercept of concordance of G-probabilities with clinician diagnosis match, with the x-axis showing G-probabilities and the y-axis showing a binary outcome of concordance with clinician diagnosis (y = 1) and nonconcordance with clinician diagnosis (y = 0), with a β (regression coefficient) of 1.047. Ideally, the higher the inferred G-probability for a disease, the more likely that it is the actual diagnosis. Linear regression comparing G-probabilities and disease match (yes/no) constraining the intercept to zero is shown as a solid blue line. A β of one indicating exact calibration is shown as a dashed black line. For visualization, G-probabilities were placed into five equally sized bins, and we plotted the proportion of instances in which predicted disease is concordant with clinician diagnosis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.42760/abstract.

Figure 2. Distribution of G-probabilities that matched clinician diagnosis (blue) and G-probabilities that did not match clinician diagnosis (red). Ideally, the correct diagnosis should have higher probabilities, and the distribution should be skewed to the right, with incorrect diagnoses having lower probabilities with a distribution skewed to the left.
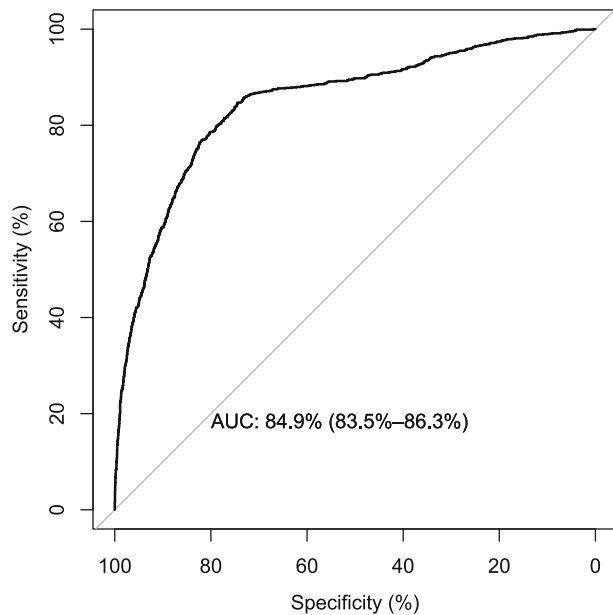
Figure 4. Receiver operative curve analysis assessing the discriminative ability of G-probabilities to correspond with clinician diagnosis. The AUC summarizes the AUC (micro-AUC) of the pooled data for all diseases. AUC, area under the curve.

in NPV and PPV at different G-probability thresholds at which NPV and PPV were inversely correlated, with higher NPV and lower PPV at lower G-probability thresholds (Supplementary Figure 3).

DISCUSSION

In the first real-world validation of the G-PROB tool, we found first, that G-PROB produces meaningful, easily interpretable outputs; second, that G-PROB is especially helpful at suggesting unlikely diagnoses; and finally, that G-PROB showed good, although not perfect, agreement with likely clinician diagnoses. Our results demonstrate that the G-PROB tool can be easily set up and used to generate G-probabilities. These G-probabilities are easy to interpret, and we demonstrate that they are meaningful based on G-probabilities matching clinician diagnosis being significantly higher compared with those that do not match and very good calibration of G-probabilities with clinician diagnosis, indicating that increases in G-probabilities are concordant with increases in proportions of clinician diagnosis match, as would be expected (Figure 2). Furthermore, ROC-AUC analysis showed that G-probabilities had good performance at discriminating clinician diagnosis. Overall, these assessments suggest that the G-probabilities generated by G-PROB are meaningful, and that increases and decreases in G-probabilities occur in a way that is appropriate.

We found that NPV was high at G-probability thresholds of <5% and <20%. This suggests that, when interpreting G-PROB results in clinic, we can be especially confident that diseases with G-probabilities below these thresholds are unlikely. However, the

PPV was modest, which demonstrates that genetics alone cannot accurately suggest or make diagnoses; this is not unexpected given that common rheumatologic conditions are typically only between 60% and 90% heritable.[12–16] It is important to note that the tool itself, and therefore its current performance, is directly linked to how much is known about the genetic risks of the diseases assessed, so as more susceptibility loci are discovered for each disease, performance may improve further. Furthermore, integrating clinical risk factors such as demographics, serology, and comorbidities with PRS has the potential to improve predictive performance, including PPV, as has been demonstrated in studies in atherosclerotic cardiovascular disease.[36]

There are several strengths of our study including the use of data from the NOAR cohort, which is a large, real-world observational cohort of patients who present with early inflammatory arthritis and are suspected to have a rheumatologic disease. Use of this cohort addresses several limitations of the original study, which assessed performance of G-PROB in a smaller sample size and used patients from a biobank.[25] Furthermore, we have compared G-PROB's performance with clinician diagnosis as opposed to classification criteria, which is more appropriate in the real-world setting given that patients in the real world may not satisfy classification criteria but will still receive diagnoses based on expert clinical opinion. Overall, the fact that G-PROB performs well in the NOAR cohort and using clinician diagnosis strengthens the evidence that the tool may be helpful in real-world clinical practice.

We acknowledge several limitations of our study. First, the NOAR cohort only included patients who identified as White; therefore, we extrapolate performance to patients of other ethnicities. As methods to improve generalizability of known genetic risk variants using transancestry data are developed, this issue may be resolved in the future, but testing of the tool in patients from other ancestries is required.[37] Future genetic studies such as functional genomics studies, which discover new disease-associated variants and functionally annotate known genetic risk variants with target genes and regulatory elements, can be incorporated into further iterations of the G-PROB tool, which may further improve its cross-ancestry performance and utility.[38,39]

A second limitation relates to power; despite including a sample size nearly four times larger than the original study,[25] we still had relatively low numbers of patients with diagnoses less commonly made in patients with suspected early inflammatory arthritis (including gout) (Figure 1). In the context of NOAR and UK clinical practice, this is likely because patients with gout are diagnosed by clinicians outside of the rheumatology outpatient setting (eg, in primary care or in hospital). Low numbers of cases in certain categories potentially overestimate the tool's ability (especially the NPV) by overemphasizing its ability to rule out uncommon diagnoses. Furthermore, because the selection of disease prevalence settings is crucial in the set-up and generation of G-probabilities, care must be taken when implementing the

G-PROB tool in clinical practice to ensure that the appropriate disease prevalence is selected to reflect the disease prevalence in a given clinic population.

In the initial study describing G-PROB, it was found in a cohort of 197 patients, of which 35% were initially misdiagnosed on first presentation to a rheumatology clinic, that for 65% of patients, the G-probability of the correct disease was higher than the G-probability of the initial diagnosis of the rheumatologist.[25] Unfortunately, the NOAR study was not set up to record misdiagnoses, and therefore, a third limitation of our study is that we could not validate these findings because "initial diagnosis" and subsequent "follow-up/final diagnosis" were not recorded.

Finally, our study has demonstrated that there is strong correlation between the G-probabilities generated by genetic data from known genetic risk alleles and the final clinician diagnosis; however, in both our study and the initial study describing G-PROB, there is a heterogeneous group of patients in the "Other disease" category. In the "Other disease" category, there are patients with inflammatory conditions (such as polymyalgia rheumatica) as well as those with noninflammatory conditions (such as chronic pain syndromes). Therefore, a fourth limitation that neither our nor the initial G-PROB study could address was whether G-PROB can assist in improving the accuracy of diagnosis for patients who exist in the heterogeneous "Other disease" category.

In conclusion, we have shown that the G-PROB tool and G-probabilities are meaningful using assessments based on real-world observational data and clinician diagnoses. Further work is needed in the form of prospective studies to assess the added value of genetic tools such as G-PROB in addition to nongenetic factors such as serology and other clinical information. Additionally, studies aimed at evaluating the views of clinicians and patients with regards to the acceptability and usefulness of using genetic information in clinical practice as well as studies assessing the cost–benefit of genotyping patients and using genetic information in the clinical setting are needed before this technology is integrated in the modern outpatient clinic.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Drs. Hum and Sharma had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Hum, Sharma, Stadler, Viatte, Ho, Barton, Bowes.
Acquisition of data. Hum, Sharma, Stadler, Viatte, Ho, Nair, MacGregor, Yates, Verstappen, Barton, Bowes.
Analysis and interpretation of data. Hum, Sharma, Stadler, Viatte, Ho, Nair, Shi, Yap, Soomro, Plant, Humphreys, MacGregor, Yates, Barton, Bowes.

REFERENCES

1. Tripp S, Grueber M. The Economic Impact and Functional Applications of Human Genetics and Genomics. American Society of Human Genetics; 2021.

2. Burki T. UK explores whole-genome sequencing for newborn babies. Lancet 2022;400(10348):260–261.

3. 23andMe. Accessed [September 1, 2021]. https://www.23andme.com

4. Ancestry. Accessed [September 1, 2021]. https://www.ancestry.com

5. Van der Linden MPM, le Cessie S, Raza K, et al. Long-term impact of delay in assessment of patients with early arthritis. Arthritis Rheum 2010;62(12):3537–3546.

6. Korpela M, Laasonen L, Hannonen P, et al. Retardation of joint damage in patients with early rheumatoid arthritis by initial aggressive treatment with disease-modifying antirheumatic drugs: five-year experience from the FIN-RACo study. Arthritis Rheum 2004;50(7):2072–2081.

7. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62(9):2569–2581.

8. Taylor W, Gladman D, Helliwell P, et al; CASPAR Study Group. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. Arthritis Rheum 2006;54(8):2665–2673.

9. Petri M, Orbai AM, Alarcón GS, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64(8):2677–2686.

10. Van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27(4):361–368.

11. Neogi T, Jansen TLTA, Dalbeth N, et al. 2015 Gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2015;74(10):1789–1798.

12. MacGregor AJ, Snieder H, Rigby AS, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43(1):30–37.

13. Soomro M, Stadler M, Dand N, et al; BABDIR Study Group; BSTOP study group. Comparative genetic analysis of psoriatic arthritis and psoriasis for the discovery of genetic risk factors and risk prediction modeling. Arthritis Rheumatol 2022;74(9):1535–1543.

14. Brown MA, Kennedy LG, Macgregor AJ, et al. Susceptibility to anky-losing spondylitis in twins: the role of genes, HLA, and the environment. Arthritis Rheum 1997;40(10):1823–1828.

15. Deapen D, Escalante A, Weinrib L, et al. A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum 1992;35(3):311–318.

16. Cadzow M, Merriman TR, Dalbeth N. Performance of gout definitions for genetic epidemiological studies: analysis of UK Biobank. Arthritis Res Ther 2017;19(1):181.

17. Bentham J, Morris DL, Cunninghame Graham DS, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 2015;47(12):1457–1464.

18. Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and sero-positive rheumatoid arthritis. Nat Genet 2012;44(3):291–296.

19. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506(7488): 376–381.

20. Stuart PE, Nair RP, Tsoi LC, et al. Genome-wide association analysis of psoriatic arthritis and cutaneous psoriasis reveals differences in their genetic architecture. Am J Hum Genet 2015;97(6):816–836.

21. Köttgen A, Albrecht E, Teumer A, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat Genet 2013;45(2):145–154.

22. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores [review]. Nat Rev Genet 2018;19(9):581–590.

23. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments [review]. Genome Med 2020;12(1):44.

24. Soomro M, Hum R, Barton A, et al. Genetic studies investigating susceptibility to psoriatic arthritis: a narrative review [review]. Clin Ther. 2023;45(9):810–815.

25. Knevel R, le Cessie S, Terao CC, et al. Using genetics to prioritize diagnoses for rheumatology outpatients with inflammatory arthritis. Sci Transl Med 2020;12(545):eaay1548.

26. Symmons DP, Silman AJ. The Norfolk Arthritis Register (NOAR) [review]. Clin Exp Rheumatol 2003;21(5 Suppl 31):S94–S99.

27. Thomson W, Harrison B, Ollier B, et al. Quantifying the exact role of HLA-DRB1 alleles in susceptibility to inflammatory polyarthritis: results from a large, population-based study. Arthritis Rheum 1999;42(4): 757–762.

28. Sharma S, Plant D, Bowes J, et al. HLA-DRB1 haplotypes predict cardiovascular mortality in inflammatory polyarthritis independent of CRP and anti-CCP status. Arthritis Res Ther 2022;24(1):90.

29. Han B, Diogo D, Eyre S, et al. Fine mapping seronegative and sero-positive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. Am J Hum Genet 2014; 94(4):522–532.

30. Miedema HS, van der Linden SM, Rasker JJ, et al. National database of patients visiting rheumatologists in The Netherlands: the standard diagnosis register of rheumatic diseases. A report and preliminary analysis. Br J Rheumatol 1998;37(5):555–561.

31. Zink A, Listing J, Klindworth C, et al; German Collaborative Arthritis Centres. The national database of the German Collaborative Arthritis Centres: I. Structure, aims, and patients. Ann Rheum Dis 2001;60(3):199–206.

32. R Computing Team. R: A language and environment for statistical computing. 2015. https://www.r-project.org/

33. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1–48.

34. 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.

35. Wei R, Wang J, Jia W. multiROC: calculating and visualizing ROC and PR curves across multi-class classifications. 2018. https://cran.r-project.org/package=multiROC

36. Weale ME, Riveros-Mckay F, Selzam S, et al. Validation of an integrated risk tool, including polygenic risk score, for atherosclerotic cardiovascular disease in multiple ethnicities and ancestries. Am J Cardiol 2021;148:157–164.

37. Amariuta T, Ishigaki K, Sugishita H, et al. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat Genet 2020;52(12):1346–1354.

38. Shi C, Rattray M, Barton A, et al. Using functional genomics to advance the understanding of psoriatic arthritis [review]. Rheumatology (Oxford) 2020;59(11):3137–3146.

39. González-Serna D, Shi C, Kerick M, et al. Functional genomics in primary T cells and monocytes identifies mechanisms by which genetic susceptibility loci influence systemic sclerosis risk. Arthritis Rheumatol 2023;75:1007–1020.