

Original Paper

Evaluating the Accuracy of the Frysian Questionnaire for Differentiation of Musculoskeletal Complaints for Triage of Musculoskeletal Diseases: Algorithm Development and Validation Study

Tjardo Daniël Maarseveen¹, MSc; Floor Reimann², MD; Ahmed al Hasan¹, MD; Annemarie Schilder², MD; Dan Zhang², MD; Freke Wink², MD; Lidy Hendriks², MD; Rachel Knevel^{1,3*}, PhD; Reinhard Bos^{2*}, PhD

¹Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

²Department of Rheumatology, Frisius Medical Center Leeuwarden, The Netherlands

³Department of Rheumatology, Newcastle University, Newcastle upon Tyne, United Kingdom

*these authors contributed equally

Corresponding Author:

Tjardo Daniël Maarseveen, MSc
Department of Rheumatology
Leiden University Medical Center
Albinusdreef 2
Leiden 2333ZA
The Netherlands
Phone: 31 0715269111
Email: t.d.maarseveen@lumc.nl

Abstract

Background: Inflammatory rheumatic diseases (IRDs) affect 5% of the general population, whereas 35% of the population experiences musculoskeletal concerns. IRDs cause early disability, reduced life expectancy, and considerable health care costs. Early diagnosis is essential to prevent long-term damage. Similarly important is the early identification of patients with musculoskeletal concerns without IRDs to prevent unnecessary health care expenses. Of the population referred to the rheumatologist, 60% have noninflammatory musculoskeletal concerns, whereas only 20% of patients with an IRD see a rheumatologist within 3 months of symptom onset. The need for digital predictive (triage) tools for rheumatic and musculoskeletal diseases led to the development of the Frysian Questionnaire for Differentiation of Musculoskeletal Complaints (FRYQ).

Objective: This study aimed to assess whether the FRYQ can distinguish IRD from noninflammatory musculoskeletal concerns in general, and rheumatoid arthritis and fibromyalgia specifically, in newly referred patients.

Methods: The FRYQ is an 87-item tool (20 open-ended and 67 closed-ended questions) used to triage new rheumatology patients at Frisius Medical Center in the Netherlands. We analyzed data from 2 sources: dataset A with 728 outpatient clinic patients and dataset B with 373 patients from the Joint Pain Assessment Scoring Tool study. We built a classifier using Extreme Gradient Boosting to distinguish inflammatory from noninflammatory conditions based on closed-ended questions. Using elastic net regularization, we identified the most informative questions. We evaluated classification using receiver operating characteristic curve analysis and assessed feature importance through Shapley Additive Explanation analysis. To test generalizability, we replicated our analysis on dataset B. Finally, we examined whether the questions of the FRYQ could be used to identify specific conditions beyond the general categories of IRD and non-IRD, specifically for detecting fibromyalgia and rheumatoid arthritis.

Results: Feature selection reduced the questionnaire from 67 to 28 items while maintaining discriminative power. After initial development, the model achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.72 (95% CI 0.67-0.78) for distinguishing inflammatory from noninflammatory conditions in an external validation set. Using a probability threshold of 0.30, the model achieved 71% sensitivity and 56% specificity on external validation. The FRYQ demonstrated stronger performance in identifying specific diagnoses such as fibromyalgia (AUC-ROC=0.81) and rheumatoid arthritis (AUC-ROC=0.77). Key discriminating features included symptom duration, pain response to movement, and anti-inflammatory medication effectiveness.

Conclusions: The FRYQ effectively distinguishes inflammatory from noninflammatory rheumatic conditions before specialist consultation and shows particular strength in identifying fibromyalgia and rheumatoid arthritis. This tool could improve rheumatology triage by prioritizing referrals with high likelihood of IRD for early rheumatologist evaluation while directing other patients to appropriate alternative resources. Prospective studies are needed to determine the FRYQ's impact on clinical outcomes and health care efficiency.

*JMIR Med Inform*2025;13:e77345; doi: [10.2196/77345](https://doi.org/10.2196/77345)

Keywords: medical informatics; questionnaire; machine learning; fibromyalgia; rheumatic diseases; rheumatoid arthritis; triage; decision support system; symptom checker

Introduction

Inflammatory rheumatic diseases (IRDs) affect 5% of the general population in the Netherlands, whereas 35% of the population experience musculoskeletal concerns [1]. Global prevalence of both IRDs and musculoskeletal concerns varies with age and region but is comparable to that in the Netherlands [2]. IRDs can cause early disability, reduced life expectancy, and considerable health care costs.

In the Dutch health care system, as in many European countries, patients with musculoskeletal concerns first consult a general practitioner (GP) before referral to a specialist. To facilitate appropriate and timely care, the Dutch College of General Practitioners (*Nederlands Huisartsen Genootschap*) has developed national guidelines and referral protocols [3], helping ensure that patients receive specialist attention when necessary.

Early diagnosis of rheumatic diseases such as rheumatoid arthritis (RA), ankylosing spondylitis, or psoriatic arthritis by specialists is essential to prevent irreversible damage [4-7]. Similarly important is the early identification of patients with musculoskeletal concerns without IRDs who can be treated in primary care to prevent unnecessary health care expenses. This need is particularly pressing in the Netherlands, which has one of the most expensive health care systems in Europe—ranking third after Austria and Germany [8].

Approximately 109 new patients with musculoskeletal concerns per 1000 people are seen by GPs annually [9]. Of the population referred to a rheumatologist, 60% have noninflammatory musculoskeletal concerns, whereas only 20% of patients with an IRD see a rheumatologist within 3 months of symptom onset [10]. For the target disease, RA, we observe that these patients have to wait—on average—4 weeks from referral to their visit to a rheumatologist [11].

The urgent need for more efficient management of referred patients led to the development of the Frysian Questionnaire for Differentiation of Musculoskeletal Complaints (FRYQ).

The FRYQ is an 87-item questionnaire consisting of 20 open-ended and 67 close-ended questions that is used in regular care in the rheumatology outpatient clinic of Frisius Medical Center to triage newly referred patients. It was developed to improve diagnostic speed for referred patients by facilitating early appropriate blood testing.

Our study had three main objectives: (1) evaluate the FRYQ's ability to differentiate IRDs from non-IRDs in newly referred patients using machine learning techniques; (2) identify which questions are most informative for distinguishing inflammatory conditions and develop a streamlined version without losing diagnostic value; and (3) assess whether the FRYQ can be used to identify specific conditions of interest, particularly fibromyalgia and RA, which represent important diagnostic targets in rheumatology practice.

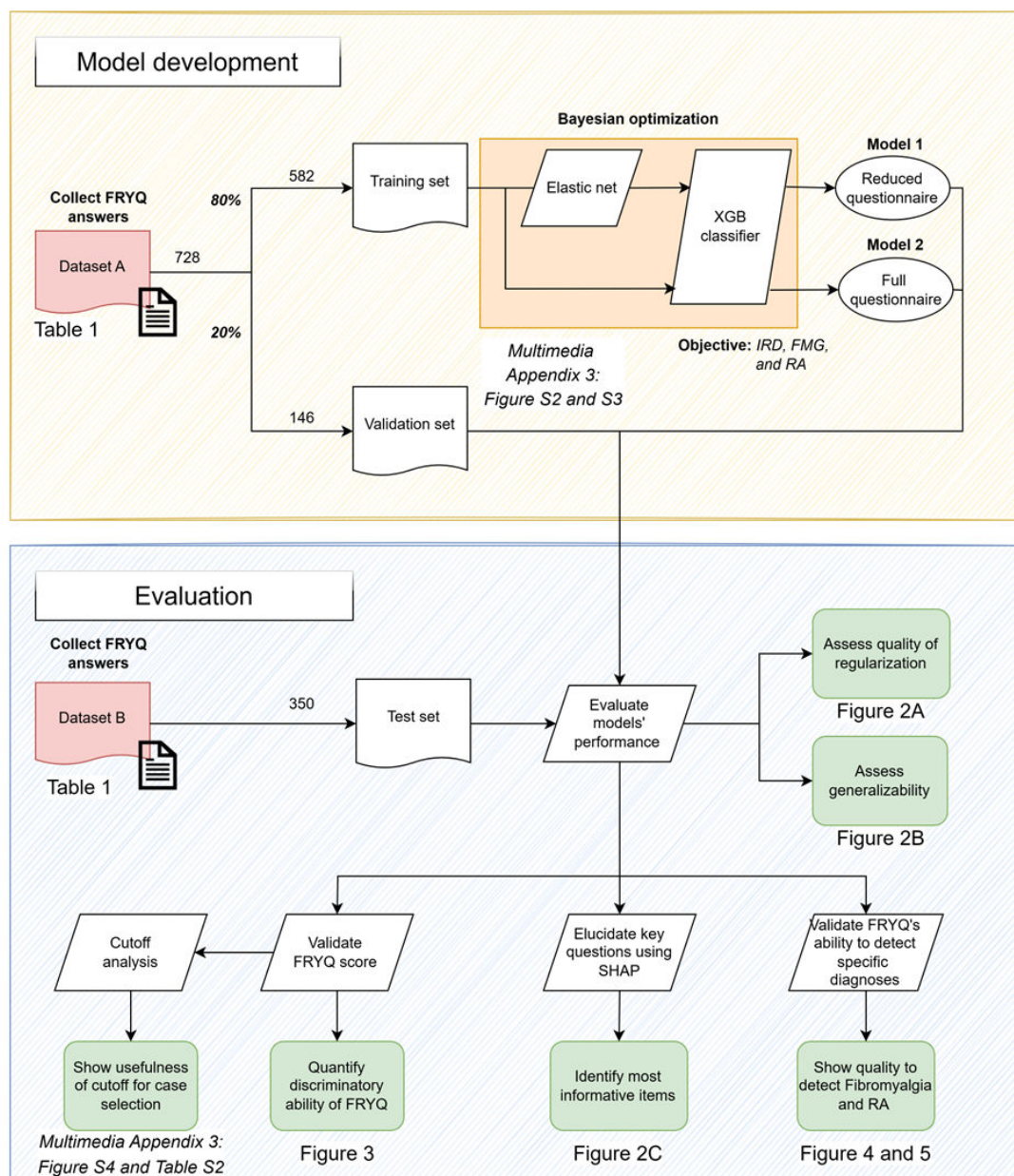
Methods

Study Design and Source of Data

In this study, we conducted a retrospective analysis of newly referred patients to the rheumatology department at Frisius Medical Center, Leeuwarden, the Netherlands (the pipeline is shown in [Figure 1](#)). We developed and validated machine learning prediction models to assess the FRYQ's ability to differentiate inflammatory from noninflammatory rheumatic conditions using only preconsultation patient responses. [Multimedia Appendix 1](#) provides the Dutch version of the questionnaire, and [Multimedia Appendix 2](#) provides the translated version in English.

We used two independent cohorts: (1) a development dataset (dataset A), which comprised 728 adult patients who completed the FRYQ between August 2017 and June 2019 as part of routine care, and (2) an external validation dataset (dataset B), which comprised 373 adult patients from the Joint Pain Assessment Scoring Tool (JPAST) study cohort [12] visiting the clinic between January 2023 and March 2024 who completed the FRYQ as an additional research instrument.

Figure 1. Study workflow consisting of two phases: (1) model development, where we built a classifier to detect inflammatory rheumatic diseases (IRDs), fibromyalgia, and rheumatoid arthritis (RA) using only Frysian Questionnaire for Differentiation of Musculoskeletal Complaints (FRYQ) responses+sex based on the data from the first pilot (n=728), and (2) model evaluation, where we deployed the model on unseen data from a different study (n=350) to assess generalizability. FMG: fibromyalgia; SHAP: Shapley Additive Explanations; XGB: Extreme Gradient Boosting.



Ethical Considerations

The FRYQ was developed by the rheumatology clinic at the Frisius Medical Center to assist in prioritizing newly referred patients. The data in dataset A were gathered as part of standard clinical care. Informed consent for the use of anonymized data in research was secured through the opt-out procedure, whereby all patient data were eligible for inclusion unless an individual filed a formal objection. According to Article 1(b) of the Dutch Medical Research Involving Human Subjects Act (Wet medisch-wetenschappelijk onderzoek met mensen [WMO]), ethical review is required only for studies that aim to obtain medical-scientific knowledge and in which participants are subjected to an intervention or are required to follow a behavioral regime [13]. As the FRYQ questionnaire was administered as part of routine care and did not involve

any intervention or imposed behavior, the study does not fall under the scope of the WMO and therefore did not require approval by a medical ethics review committee.

In contrast, dataset B consisted of patients enrolled in the JPAST study, which involved time-consuming study procedures. Therefore, the use of the FRYQ in this dataset was reviewed and approved by the medical ethics committee of the Frisius Medical Center (reference number RTPO-1104) as part of the JPAST study protocol. Since participation in the JPAST study involved procedures beyond standard care, each participant provided written informed consent authorizing the use of their data for research.

For both datasets, patient data security was ensured by assigning each patient a unique sequential identifier through

pseudonymization. Data were stored and analyzed in a secure, password-protected database accessible only to authorized personnel. Participants retained the right to opt out of the use of their data for research purposes and did not receive financial compensation for their involvement in this study. All study procedures were conducted in accordance with the principles of the Declaration of Helsinki.

Participants

Eligibility Criteria

All adult patients newly referred to the rheumatology outpatient clinic who completed the FRYQ were eligible. From dataset A, 854 patients were initially invited, with 728 (85.2%) having complete data for analysis. From dataset B, all patients with complete FRYQ responses and diagnosis data were included.

Study Setting

Data collection took place in a nonacademic hospital setting (Frisius Medical Center) during routine clinical care for dataset A and as part of the international JPAST validation study for dataset B.

Outcome

Primary Outcome

The primary outcome was the presence of an IRD versus a noninflammatory musculoskeletal condition (non-IRD) as diagnosed by a rheumatologist after clinical evaluation. This determination represented the clinical ground truth against which the model predictions were evaluated. The outcome was assessed by extracting the final diagnosis from electronic health records.

Secondary Outcomes

Specific diagnoses of fibromyalgia and RA were evaluated as secondary outcomes in additional predictive models.

Predictors (Input Variables)

The primary input variables consisted of responses to the 67 closed-ended questions from the FRYQ supplemented by patient age and sex. The FRYQ is an 87-item tool (including 20 open-ended questions not used in this analysis) designed by rheumatologists for triaging rheumatology patients. Question topics covered pain characteristics and patterns, stiffness symptoms and duration, functional limitations, associated symptoms, comorbid conditions, medication responses, and patient behaviors and characteristics. No laboratory test results or imaging data were included as predictors in the model.

Missing Data

Patients with incomplete demographic and questionnaire data were removed from dataset A (127/862, 14.7%) and dataset B (166/552, 30.1%; Figure S1 in [Multimedia Appendix 3](#)). Furthermore, we had some patients with inconsistent diagnostic labels removed from dataset A (7/862, 0.8%) and a

few with missing diagnosis information removed from dataset B (36/552, 6.5%).

Model Development

Data Preprocessing

No significant preprocessing was required as the questionnaire items were already structured. Patient sex was included as a binary variable, and patient age was included as a continuous variable.

Feature Selection

We used elastic net regularization to identify the most informative subset of questions from the original 67 closed-ended items. The regularization parameters (alpha and L1 ratio) were optimized using Bayesian methods with a tree-structured Parzen estimator across 1000 iterations [14]. This approach allowed us to reduce dimensionality while retaining predictive power.

Model Specification

We used the Extreme Gradient Boosting technique [15] to build a model for IRD prediction using only the questions selected by the elastic net. Extreme Gradient Boosting was chosen for its ability to capture complex relationships between variables and its robust performance in previous prediction tasks [16]. All machine learning analyses were performed in Python (version 3.6.13; Python Software Foundation).

Model Training

The model was trained on 80% of dataset A. We used Bayesian optimization across 1000 iterations to tune model hyperparameters, optimizing for root mean square logarithmic error to emphasize sensitivity to inflammatory conditions.

Model Performance

Measures of Model Performance

We evaluated model discrimination using the area under the receiver operating characteristic curve (AUC-ROC). Calibration was assessed using calibration plots comparing predicted probabilities with observed outcome frequencies.

Validation

The model was first validated on the reserved 20% of dataset A (internal validation) and then on dataset B (external validation) to assess generalizability. Both the reduced model (with selected features) and the full model (with all 67 questions) were evaluated to assess the impact of feature selection.

Clinical Utility

We conducted threshold analysis to identify an optimal probability cutoff for distinguishing IRD from non-IRD cases in clinical practice, evaluating sensitivity, specificity, positive predictive value, and negative predictive value at various thresholds.

Model Interpretation

To understand feature importance and explain model predictions, we conducted Shapley additive explanation (SHAP) analysis [17]. This approach quantified the contribution of each question to the model output, allowing us to identify which questions were most informative for detecting inflammatory conditions.

Results

Overview

Dataset A comprised 728 patients, of whom 239 (32.8%) were male and 489 (67.2%) were female (Table 1). Nearly half (257/582, 44.2%) of the patients in the training dataset

had IRDs. The most common disorders associated with IRD were RA (62/582, 10.7%), gout (45/582, 7.7%), and spondyloarthritis (43/582, 7.4%). Among noninflammatory arthropathies, the most frequent were osteoarthritis (119/582, 20.4%), tendinopathy (92/582, 15.8%), and fibromyalgia (91/582, 15.6%).

The external test dataset B included 350 patients with complete information, comprising 106 (30.3%) male and 244 (69.7%) female individuals. Only 26.6% (93/350) of the patients were diagnosed with an IRD, which was notably lower than those in the training and validation partitions of dataset A (257/582, 44.2% and 67/146, 45.9%, respectively). In contrast, the JPAST data contained a higher proportion of patients with an “other” diagnosis, indicating that this dataset represented a somewhat different patient population.

Table 1. Patient characteristics in both datasets.

	Dataset A		Dataset B
	Training (n=582)	Hold-out (n=146)	Replication (n=350)
Sex (female), n (%)	308 (52.9)	97 (66.4)	244 (69.7)
Age (y), mean (SD)	52 (17)	55 (17)	53 (16)
IRD ^a , n (%)	257 (44.2)	67 (45.9)	93 (26.6)
Rheumatoid arthritis	62 (10.7)	25 (17.1)	15 (4.3)
Polymyalgia rheumatica	20 (3.4)	7 (4.8)	12 (3.4)
Gout	45 (7.7)	4 (2.7)	11 (3.1)
Systemic lupus erythematosus	3 (0.5)	2 (1.4)	0 (0)
Psoriatic arthritis	20 (3.4)	9 (6.2)	4 (1.1)
Spondyloarthritis	43 (7.4)	9 (6.2)	13 (3.7)
Sjögrenn disease	5 (0.9)	1 (0.7)	1 (0.3)
Reactive arthritis	9 (1.5)	3 (2.1)	2 (0.6)
Undifferentiated arthritis	9 (1.5)	1 (0.7)	14 (4)
Systemic sclerosis	4 (0.7)	1 (0.7)	1 (0.3)
CPPD ^b	10 (1.7)	3 (2.1)	2 (0.6)
Probable RMD ^c in development	17 (2.9)	1 (0.7)	18 (5.1)
Sarcoidosis	9 (1.5)	1 (0.7)	0 (0)
ANCA ^d -associated vasculitis	1 (0.2)	0 (0)	0 (0)
No IRD, n (%)	323 (55.5)	79 (54.1)	185 (52.9)
Osteoarthritis	119 (20.4)	33 (22.6)	73 (20.9)
Tendinopathy	92 (15.8)	25 (17.1)	39 (11.1)
Fibromyalgia	91 (15.6)	19 (13)	68 (19.4)
Pernio	0 (0)	0 (0)	4 (1.1)
Hypermobility	21 (3.6)	2 (1.4)	1 (0.3)
Ambiguous, n (%)	2 (0.3)	0 (0)	72 (20.6)
Other	0 (0)	0 (0)	65 (18.6)
Unknown	2 (0.3)	0 (0)	7 (2)

^aIRD: inflammatory rheumatic disease.

^bCPPD: calcium pyrophosphate crystal deposition disease.

^cRMD: rheumatic and musculoskeletal disease.

^dANCA: antineutrophil cytoplasmic antibody.

Prediction of IRD

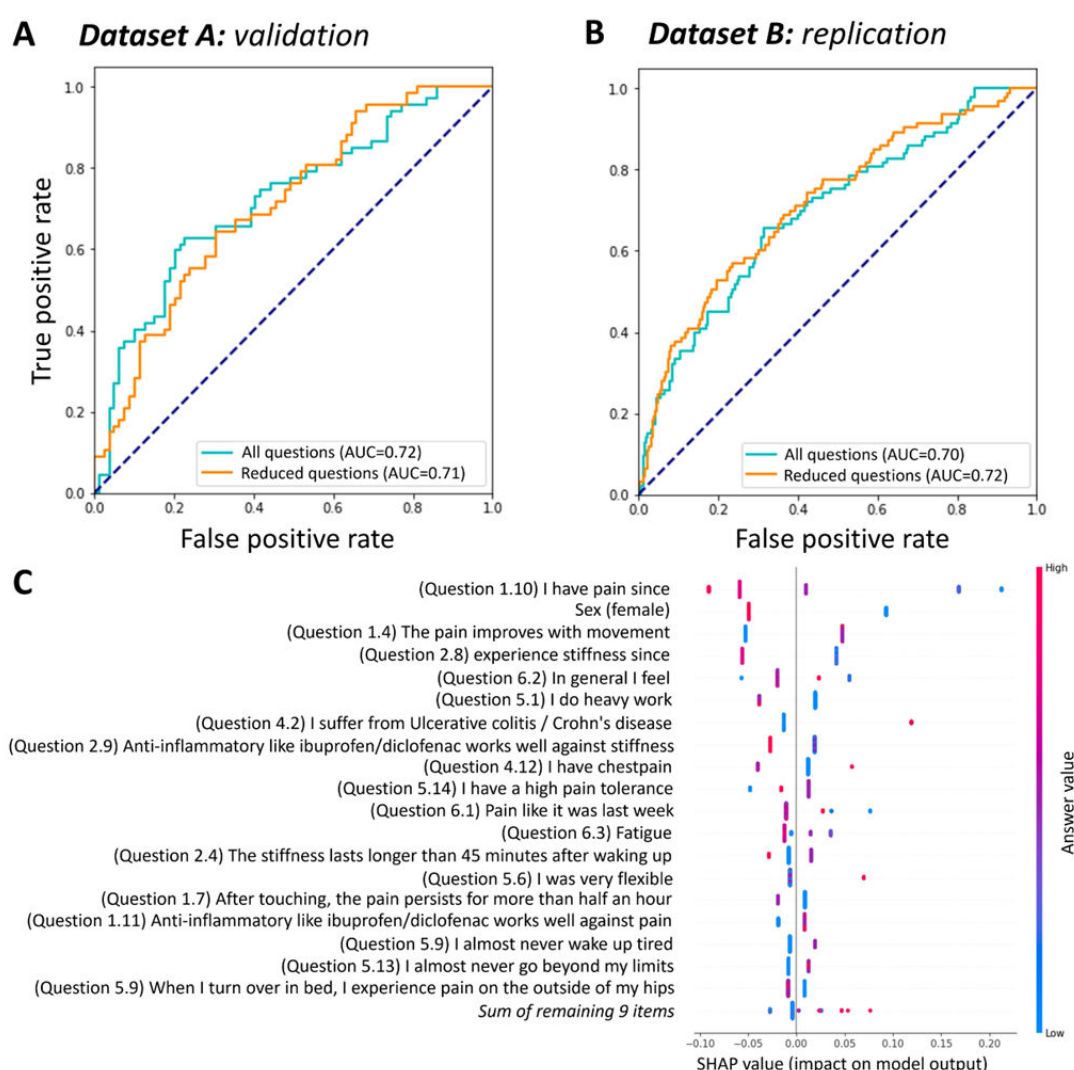
After hyperparameter optimization ($\alpha=.07$; L1-ratio=0.16; Figures S2 and S3 in [Multimedia Appendix 3](#)) and feature selection, our model reduced the 67 closed-ended questions to 28 key items (Table S1 in [Multimedia Appendix 3](#)). This streamlined model effectively distinguished IRD from non-IRD across multiple validation steps.

In the development dataset A, the reduced model achieved an AUC-ROC of 0.70 (95% CI 0.67-0.74) in the training set and 0.71 (95% CI 0.64-0.78) in the hold-out validation set,

performing comparably to the full model (AUC-ROC of 0.67, 95% CI 0.63-0.72 and 0.72, 95% CI 0.64-0.78, respectively; [Figure 2A](#)).

Importantly, when tested in the independent dataset B, the model demonstrated robust generalizability, with an AUC-ROC of 0.72 (95% CI 0.67-0.78) for the reduced model and 0.68 (95% CI 0.63-0.75) for the full model ([Figure 2B](#)), confirming the model's stability across different patient populations.

Figure 2. Performance of the inflammatory rheumatic disease classifier: (A) receiver operating characteristic (ROC) curves showing performance before and after regularization in the validation set, (B) ROC curves showing performance before and after regularization in the replication set, and (C) Shapley Additive Explanation (SHAP) analysis results highlighting the impact and direction of the most important features contributing to the classifier's predictions. AUC: area under the curve.



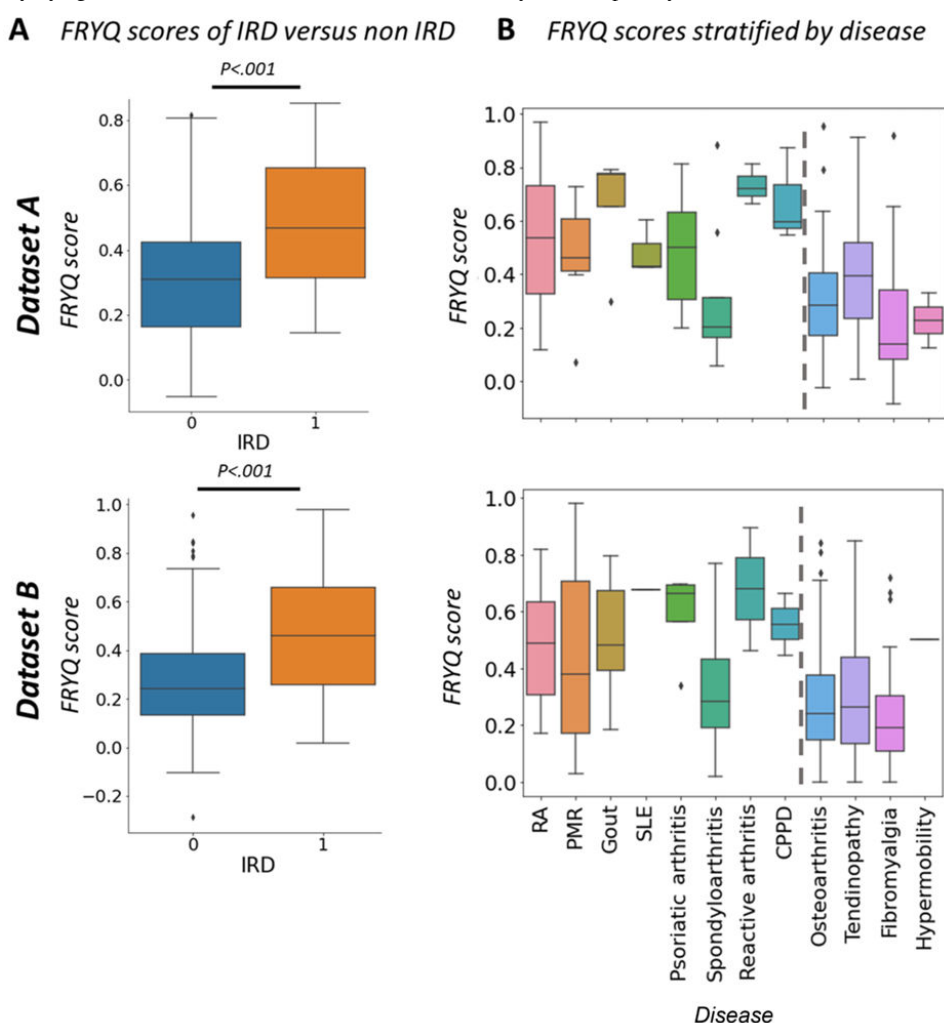
The FRYQ score (model-predicted probability) consistently distinguished IRD patients from non-IRD patients across both datasets ($P<.001$; [Figure 3](#)). In the development set A, the median FRYQ score was 0.47 (IQR 0.31-0.65) for patients with IRD, versus 0.30 (IQR 0.16-0.42) for patients without IRD. Similarly, in replication set B, the median score was 0.46 (IQR 0.26-0.66) for patients with IRD, compared to 0.24 (IQR 0.13-0.39) for patients without IRD. When

examining specific conditions, most IRDs exhibited higher scores than non-IRDs, with 2 notable exceptions: spondyloarthritis consistently scored lower among IRDs, whereas tendinopathy scored relatively high (>0.35) among non-IRDs, suggesting areas where diagnostic discrimination was more challenging. Fibromyalgia consistently showed very low scores (<0.2), making it a promising target for specialized detection.

The key features for identifying an IRD in the newly referred population (Figure 2C) were questions related to duration of concerns (questions 1.10 and 2.8), male sex, and decrease of pain upon movement (question 1.4). Moreover, the presence of certain comorbidities influenced the model's assessment. Ulcerative colitis and Crohn disease (question

4.2) showed a positive association with IRDs, whereas chest pain (question 4.12) showed a negative association. Additionally, respondents diagnosed with an IRD more frequently reported that nonsteroidal anti-inflammatory drugs such as ibuprofen and diclofenac were effective in managing their symptoms (questions 2.9 and 1.11).

Figure 3. Overview of Frysian Questionnaire for Differentiation of Musculoskeletal Complaints (FRYQ) score distributions (A) between inflammatory rheumatic disease (IRD) and non-IRD and (B) more specifically by disease, with IRDs on the left of the dashed line and non-IRDs on the right. The comparison between IRDs and non-IRDs was conducted using a 2-tailed Student *t* test. CPPD: calcium pyrophosphate crystal deposition disease (pseudogout); PMR: polymyalgia rheumatica; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus. * $P<.05$; ** $P<.01$; *** $P<.001$.



Next, we defined the optimal clinical threshold for case selection. On the basis of the training data, we set the cutoff at 0.30 (Figure S4 in Multimedia Appendix 3) as it captured more than three-fourths of IRD cases (sensitivity=0.76) while ensuring that half of the cases without IRDs were excluded (specificity=0.49). This threshold showed similar sensitivity (0.71) and specificity (0.56) in the external validation dataset. The positive predictive value was lower in the external validation dataset (0.39 vs 0.56 in dataset A), attributable to the lower IRD prevalence (93/350, 26.6% vs 257/582, 44.2%; Table S4 in Multimedia Appendix 3).

The calibration curve indicated that the model slightly overestimated IRD probability in the external test dataset (Figure S5 in Multimedia Appendix 3), reflecting the difference in disease prevalence between the training and

validation populations. Despite this, the discriminatory ability remained strong, with IRDs consistently receiving higher scores than non-IRDs across both datasets, reinforcing the FRYQ's potential value as a triage tool.

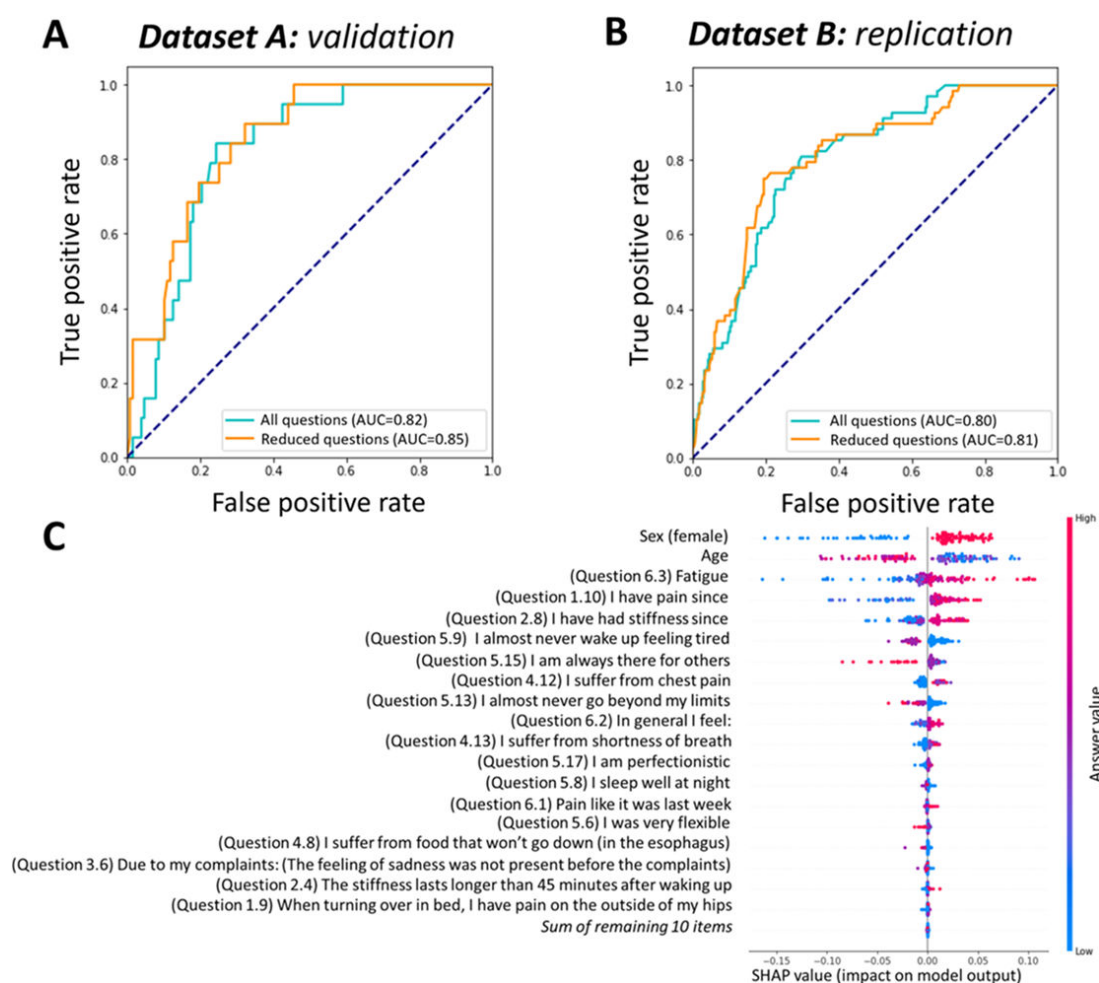
Prediction of Specific Diagnoses: Fibromyalgia and RA

Our analysis extended beyond distinguishing inflammatory from noninflammatory conditions to evaluate the FRYQ's ability to identify specific diagnoses. Patients with fibromyalgia showed particularly distinctive response patterns compared to those with inflammatory conditions (Figure 3), making this a compelling diagnostic target with sufficient representation in our datasets.

The fibromyalgia-specific model demonstrated strong discriminative ability, achieving an AUC-ROC of 0.85 (95% CI 0.78-0.90) in the development dataset and maintaining robust performance in the external validation dataset with an AUC-ROC of 0.81 (95% CI 0.75-0.86; [Figure 4](#)). SHAP analysis identified key predictive features for fibromyalgia: female sex, younger to middle age, fatigue (question 6.3),

chest pain (question 4.12), extended symptom duration (questions 1.10 and 2.8), and waking up tired (question 5.9). The model effectively differentiated fibromyalgia from RA, although some misclassification occurred with noninflammatory conditions such as hypermobility ([Figure 6A in Multimedia Appendix 3](#)).

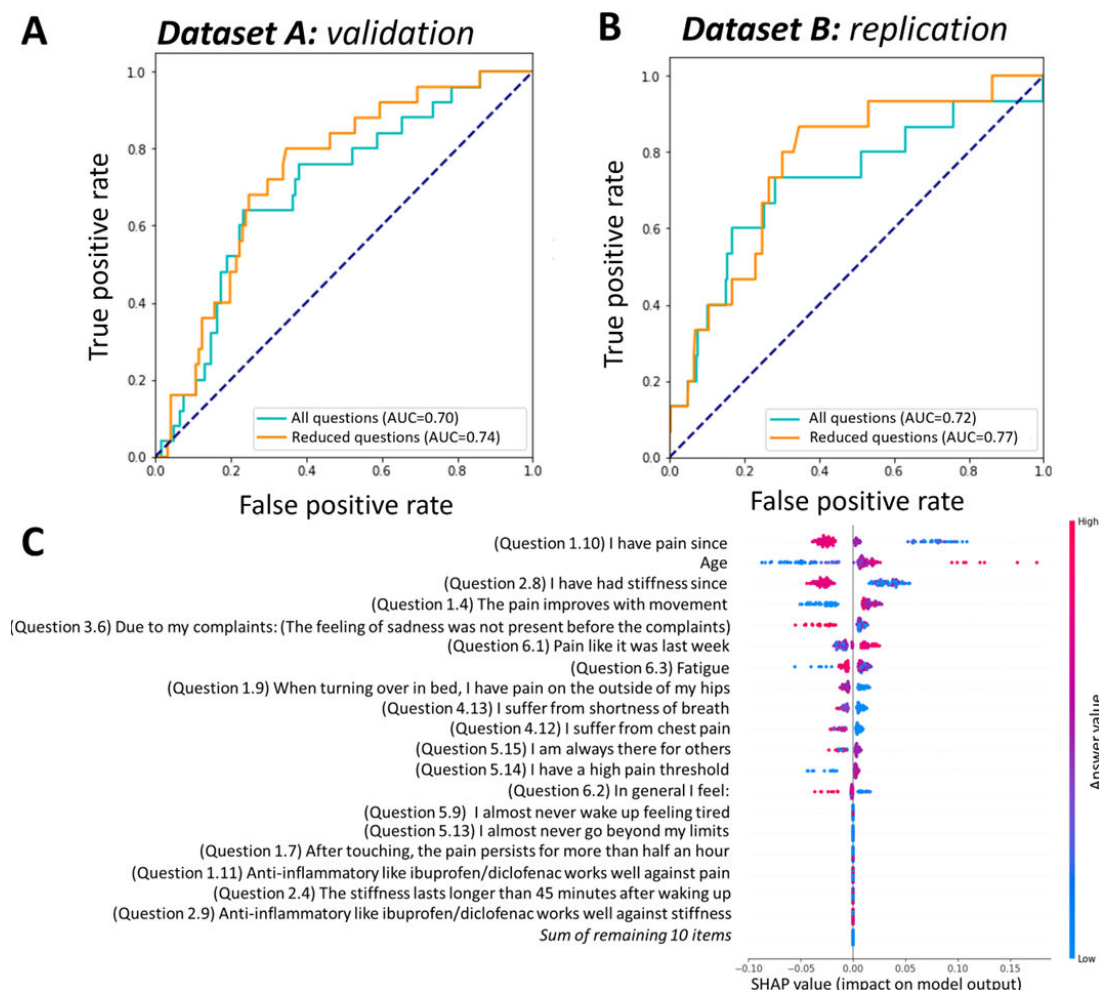
Figure 4. Performance of the fibromyalgia classifier: (A) receiver operating characteristic (ROC) curves showing performance before and after regularization in the validation set, (B) ROC curves showing performance before and after regularization in the replication set, and (C) Shapley Additive Explanation (SHAP) analysis results highlighting the impact and direction of the most important features contributing to the classifier's predictions. AUC: area under the curve.



For RA—a critical target for early rheumatology intervention—our model showed moderate discrimination, with AUC-ROC values of 0.74 (95% CI 0.63-0.83) in the development set and 0.77 (95% CI 0.66-0.86) in the external validation set ([Figure 5](#)). Key predictors included shorter disease duration (questions 1.10 and 2.8), pain improvement with movement (question 1.4), recent severe pain (question

6.1), onset at middle age or later, and fewer respiratory or chest symptoms (questions 4.13 and 4.12) or hip concerns (question 1.9). The model's primary challenge was differentiating RA from other IRDs with similar presentations, such as reactive arthritis, polymyalgia rheumatica, and gout ([Figure 6B in Multimedia Appendix 3](#)).

Figure 5. Performance of the rheumatoid arthritis classifier: (A) receiver operating characteristic (ROC) curves showing performance before and after regularization in the validation set, (B) ROC curves showing performance before and after regularization in the replication set, and (C) Shapley Additive Explanation (SHAP) analysis results highlighting the impact and direction of the most important features contributing to the classifier's predictions. AUC: area under the curve.



Discussion

Principal Findings

Our results demonstrate that the FRYQ can effectively distinguish between IRDs and noninflammatory musculoskeletal concerns before patients' first rheumatology consultation. The model achieved robust performance, with an AUC-ROC of 0.71 in the development dataset and 0.72 in the external validation dataset, indicating good generalizability. Importantly, we found that a reduced set of just 28 questions from the original 67 closed-ended items maintained equivalent discriminative power, suggesting potential for a more efficient screening tool.

Beyond the primary classification task, our analysis revealed that the FRYQ demonstrates even stronger performance in identifying specific conditions, particularly fibromyalgia (AUC-ROC of 0.85 and 0.81 in the development and replication datasets, respectively). The FRYQ also showed adequate ability to detect RA cases (AUC-ROC of 0.77 in the external validation dataset), although it had more difficulty distinguishing RA from other inflammatory conditions.

These findings have important clinical applications. In settings with limited rheumatology resources, the FRYQ could serve as a triage tool to streamline the large influx of referrals (overreferral) by prioritizing the referred patients with higher probability of inflammatory disease for early specialist intervention. This could result in more efficient allocation of specialist resources, potentially reducing waiting lists for a substantial number of patients with IRD. In light of the anticipated shortage of clinical professionals, it is increasingly important to identify which patients truly need rheumatologists' care versus those who can be managed elsewhere [18].

For example, the FRYQ's ability to identify fibromyalgia among the rheumatology waiting list could help redirect these patients to alternative care providers instead (pain specialists, physiotherapists, or dedicated programs), freeing capacity at the clinic for patients needing rheumatological care. Currently, rheumatologists spend substantial time on patients they cannot effectively treat as approximately 60% of rheumatology referrals involve noninflammatory conditions (such as fibromyalgia) [10]. This optimized allocation may also offer an additional benefit: the extra capacity created could reduce primary care physicians' reluctance to refer

clinically ambiguous cases that would benefit from specialist input.

Finally, the FRYQ could aid clinicians in their decision to request extra laboratory workup. When responses suggest RA, clinicians could order relevant serological markers (anticitrullinated protein antibody, rheumatoid factor, C-reactive protein, and erythrocyte sedimentation rate) before the initial visit, potentially streamlining the diagnostic process and reducing time to treatment for conditions requiring early intervention [4-7].

The clinical threshold of 0.30 that we identified balances sensitivity (capturing 71%-76% of inflammatory cases) with reasonable specificity (49%-56%), making it suitable for initial screening where missing IRD cases would be more concerning than false positives. As reported by Graydon and Thompson [19], primary care physicians and rheumatologists frequently disagree on referral urgency (47%), with rheumatologists upgrading 17% of referrals to urgent status—cases initially overlooked by GPs. This underlines the potential value of applying such a threshold in practice, although adjustments may be required depending on local context, such as disease prevalence and health care setting.

Our findings align with the growing recognition that structured questionnaires can support diagnostic decision-making in rheumatology by systematically capturing and integrating patients' own experiences and symptom descriptions. While over 100 symptom checkers currently exist [20], their adoption has remained limited due to lack of validation or suboptimal diagnostic accuracy [20,21]. Therefore, the availability of a validated tool such as the FRYQ may fulfill a need in the field. As far as we know, this is the first questionnaire developed in the Netherlands for triaging musculoskeletal concerns that has been validated using real-world patient data.

The predictive factors identified in our SHAP analysis align with established clinical knowledge. Pain improvement with movement, shorter morning stiffness duration, and good response to nonsteroidal anti-inflammatory drugs all emerged as important indicators of inflammatory disease [22-24]. The association between inflammatory bowel disease and IRDs supports known connections between these conditions [25]. For fibromyalgia, we observe that patients report sleep problems, persistent pain, and fatigue [26].

Beyond questionnaires, there have been other digital approaches proposed in the literature for improving rheumatology triage. Natural language processing (NLP) of referral letters has shown promising results in studies from the United Kingdom [27] and the Netherlands [28], potentially complementing questionnaire-based methods. Similarly, patient-reported outcomes have been investigated, although Schäfer et al [29] found these to be less effective at distinguishing inflammatory from noninflammatory rheumatic conditions.

Study Limitations

Several limitations of this study should be acknowledged. First, while we validated our model in an external dataset,

both cohorts originated from the same medical center, potentially limiting generalizability to different health care settings or patient populations. Second, we observed a rather big difference in prevalence of inflammatory conditions (257/582, 44.2% vs 93/350, 26.6%). This affected the model's calibration in external validation, suggesting the need for recalibration when applying to populations with different disease prevalence.

These prevalence differences likely stem from distinct patient selection mechanisms. Dataset A included patients with prescheduled outpatient appointments who completed the questionnaire during their first visit as part of routine care in Medical Center Leeuwarden (approximately 70% of referrals). In contrast, dataset B comprised volunteers from the JPAST study (approximately 40% of referrals). The time-consuming study procedures of the JPAST study likely led to an overrepresentation of nonurgent (non-IRD) cases among the volunteers.

We should also acknowledge that our model showed moderate discriminative ability (AUC of approximately 0.72) rather than excellent performance. Therefore, it could not reliably distinguish certain conditions such as spondyloarthritis from noninflammatory diseases. Rare inflammatory conditions were also underrepresented.

Crucially, our study was tailored to patients already referred to rheumatologists with the goal of optimizing the time from referral to specialist consultation. This means that the FRYQ cannot be used as a prescreening tool at GP offices. Importantly, it was not designed to exclude patients from care but rather to guide referred patients to the most appropriate care providers and streamline referral pathways.

Future Directions

To definitively establish the FRYQ's clinical utility, we propose conducting a randomized clinical trial in the future to assess whether FRYQ-guided triage leads to faster appropriate treatment, reduced unnecessary appointments, and improved patient satisfaction. Furthermore, our study would benefit from additional replication in other countries to assess how health care system differences affect its effectiveness. Referral rates and treatment protocols vary substantially between countries—for instance, 21% of patients with musculoskeletal conditions are referred to specialists in Spain compared to 11% in the Netherlands and 5% in Sweden [30].

In the future, we also plan to incorporate NLP techniques to leverage open-ended questions as they may capture nuanced symptom descriptions that structured items tend to miss. Modern NLP methods, including large language models, show potential for supporting triage [31] and the identification of rare or unrecognized diseases using free-text data [32,33]. For spondyloarthritis in particular, large language models could be prompted with educational information on axial spondyloarthritis recognition to improve detection.

Finally, we may enhance the questionnaire by incorporating questions on occupational stress given that it is a known

predictor of non-IRD [34,35]. Further refinement could also involve integrating FRYQ scores with serological markers and imaging to create a comprehensive multimodal prediction model. This multimodal approach might overcome some limitations of questionnaire-based triage alone, particularly for discerning conditions with overlapping symptomatology (ie, overlapping symptoms of osteoarthritis and early RA [36]).

Conclusions

The FRYQ shows promise in distinguishing inflammatory from noninflammatory rheumatic conditions in newly referred

patients, particularly fibromyalgia and RA, before specialist assessment. While not a substitute for clinical judgment, this tool could help optimize the time from referral to specialist consultation by prioritizing cases likely to involve IRDs and guiding others toward alternative care pathways or additional laboratory testing. These potential benefits merit evaluation in prospective implementation studies, especially given the projected shortage of rheumatology professionals.

Acknowledgments

The authors thank all participants of the study for filling out the Frysian Questionnaire for Differentiation of Musculoskeletal Complaints. The authors also want to extend their thanks to the data managers and the outpatient clinic secretariat.

Funding

This study received funding from the Netherlands Organisation for Health Research and Development Clinical Fellowship (40-00703-97-19069), and the SPIDeRR-NL project (Stratification of Patients using advanced Integrative modeling of Data Routinely acquired for diagnosing Rheumatic complaints in the Netherlands), which was granted by the Netherlands Organisation for Health Research and Development Open Competition (activity 09120012110075).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. The corresponding code is publicly available on a GitHub repository [37].

Authors' Contributions

FR, AaH, AS, DZ, FW, LH, and RB were responsible for data collection and processing. TDM conducted the machine learning analysis. RK, TDM, RB, and FR interpreted the results. TDM and FR drafted the initial version of the manuscript, which was further developed by RK and RB. All authors reviewed and approved the final version of the manuscript for submission.

Conflicts of Interest

RB has received research grants from Galapagos and Sanofi; has served on the speaker bureau for Galapagos and Janssen; and has received consulting fees from AbbVie, Galapagos, Pfizer, and UCB. All other authors declare no other conflicts of interest.

Multimedia Appendix 1

The original Frysian Questionnaire for Differentiation of Musculoskeletal Complaints in Dutch.
[DOCX File (Microsoft Word File), 90 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Table comprising the English translation of all original Dutch questions from the Frysian Questionnaire for Differentiation of Musculoskeletal Complaints.

[DOCX File (Microsoft Word File), 20 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Supplementary results supporting the main analysis, including patient selection details, model optimization procedures, comprehensive performance metrics and calibration analyses, the final questionnaire items selected by the model, and results from disease-specific classification models.

[PDF File (Adobe File), 839 KB-Multimedia Appendix 3]

References

1. van der Linden MW, Westert GP, Schellevis F. Tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk: klachten en aandoeningen in de bevolking en in de huisartspraktijk [Report in Dutch]. NIVEL/RIVM; 2004. URL: https://www.nivel.nl/sites/default/files/bestanden/ns2_r1_h02.pdf [Accessed 2025-11-05]
2. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet. Oct 17, 2020;396(10258):1204-1222. [doi: [10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)] [Medline: [33069326](https://pubmed.ncbi.nlm.nih.gov/33069326/)]

3. NHG-richtlijnen [Web page in Dutch]. Nederlands Huisartsen Genootschap. 2023. URL: <https://richtlijnen.nhg.org/> [Accessed 2025-07-28]
4. Saalfeld W, Mixon AM, Zelig J, Lydon EJ. Differentiating psoriatic arthritis from osteoarthritis and rheumatoid arthritis: a narrative review and guide for advanced practice providers. *Rheumatol Ther*. Dec 2021;8(4):1493-1517. [doi: [10.1007/s40744-021-00365-1](https://doi.org/10.1007/s40744-021-00365-1)] [Medline: [34519965](https://pubmed.ncbi.nlm.nih.gov/34519965/)]
5. Nell VPK, Machold KP, Eberl G, Stamm TA, Uffmann M, Smolen JS. Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. *Rheumatology (Oxford)*. Jul 2004;43(7):906-914. [doi: [10.1093/rheumatology/keh199](https://doi.org/10.1093/rheumatology/keh199)] [Medline: [15113999](https://pubmed.ncbi.nlm.nih.gov/15113999/)]
6. Malaviya AP, Ostor AJK. Early diagnosis crucial in ankylosing spondylitis. *Practitioner*. Dec 2011;255(1746):21-24, [Medline: [22272527](https://pubmed.ncbi.nlm.nih.gov/22272527/)]
7. 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*. Dec 2010;62(12):3537-3546. [doi: [10.1002/art.27692](https://doi.org/10.1002/art.27692)] [Medline: [20722031](https://pubmed.ncbi.nlm.nih.gov/20722031/)]
8. Spending on health care up by 8.1 percent in 2024. Centraal Bureau voor de Statistiek. 2025. URL: <https://www.cbs.nl/en-gb/news/2025/22/spending-on-health-care-up-by-8-1-percent-in-2024> [Accessed 2025-08-29]
9. van der Waal JM, Bot SDM, Terwee CB, van der Windt D, Bouter LM, Dekker J. Determinants of the clinical course of musculoskeletal complaints in general practice: design of a cohort study. *BMC Musculoskelet Disord*. Feb 24, 2003;4:3. [doi: [10.1186/1471-2474-4-3](https://doi.org/10.1186/1471-2474-4-3)] [Medline: [12600275](https://pubmed.ncbi.nlm.nih.gov/12600275/)]
10. Stack RJ, Nightingale P, Jinks C, et al. Delays between the onset of symptoms and first rheumatology consultation in patients with rheumatoid arthritis in the UK: an observational study. *BMJ Open*. Mar 4, 2019;9(3):e024361. [doi: [10.1136/bmjopen-2018-024361](https://doi.org/10.1136/bmjopen-2018-024361)] [Medline: [30837252](https://pubmed.ncbi.nlm.nih.gov/30837252/)]
11. Raza K, Stack R, Kumar K, et al. Delays in assessment of patients with rheumatoid arthritis: variations across Europe. *Ann Rheum Dis*. Oct 2011;70(10):1822-1825. [doi: [10.1136/ard.2011.151902](https://doi.org/10.1136/ard.2011.151902)] [Medline: [21821867](https://pubmed.ncbi.nlm.nih.gov/21821867/)]
12. Knitza J, Knevel R, Raza K, et al. Toward earlier diagnosis using combined eHealth tools in rheumatology: the joint pain assessment scoring tool (JPAST) project. *JMIR Mhealth Uhealth*. May 15, 2020;8(5):e17507. [doi: [10.2196/17507](https://doi.org/10.2196/17507)] [Medline: [32348258](https://pubmed.ncbi.nlm.nih.gov/32348258/)]
13. Act of 26 February 1998 on medical-scientific research involving human subjects (Medical Research Involving Human Subjects Act, WMO). Dutch Official Gazette. 1998. URL: <https://wetten.overheid.nl/BWBR0009408/2025-01-01> [Accessed 2025-11-04]
14. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. arXiv. Preprint posted online on Jun 13, 2012. [doi: [10.48550/arXiv.1206.2944](https://doi.org/10.48550/arXiv.1206.2944)]
15. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Krishnapuram B, Shah M, editors. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 785-794. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)] ISBN: 9781450342322
16. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022:507-520. ISBN: 9781713871088
17. Roth AE, editor. The Shapley Value: Essays in Honor of Lloyd S Shapley. Cambridge: Cambridge University Press; 1988. [doi: [10.1017/CBO9780511528446](https://doi.org/10.1017/CBO9780511528446)] ISBN: 9780511528446
18. Miloslavsky EM, Marston B. The challenge of addressing the rheumatology workforce shortage. *J Rheumatol*. Jun 2022;49(6):555-557. [doi: [10.3899/jrheum.220300](https://doi.org/10.3899/jrheum.220300)] [Medline: [35428725](https://pubmed.ncbi.nlm.nih.gov/35428725/)]
19. Graydon SL, Thompson AE. Triage of referrals to an outpatient rheumatology clinic: analysis of referral information and triage. *J Rheumatol*. Jul 2008;35(7):1378-1383. [Medline: [18484689](https://pubmed.ncbi.nlm.nih.gov/18484689/)]
20. Hügler M, Omoumi P, van Laar JM, Boedecker J, Hügler T. Applied machine learning and artificial intelligence in rheumatology. *Rheumatol Adv Pract*. 2020;4(1):rkaa005. [doi: [10.1093/rap/rkaa005](https://doi.org/10.1093/rap/rkaa005)] [Medline: [32296743](https://pubmed.ncbi.nlm.nih.gov/32296743/)]
21. Knitza J, Mohn J, Bergmann C, et al. Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. *Arthritis Res Ther*. Apr 13, 2021;23(1):112. [doi: [10.1186/s13075-021-02498-8](https://doi.org/10.1186/s13075-021-02498-8)] [Medline: [33849654](https://pubmed.ncbi.nlm.nih.gov/33849654/)]
22. Brorsson S, Hilliges M, Sollerman C, Nilsson A. A six-week hand exercise programme improves strength and hand function in patients with rheumatoid arthritis. *J Rehabil Med*. Apr 2009;41(5):338-342. [doi: [10.2340/16501977-0334](https://doi.org/10.2340/16501977-0334)] [Medline: [19363566](https://pubmed.ncbi.nlm.nih.gov/19363566/)]
23. Baillet A, Payraud E, Niderprim VA, et al. A dynamic exercise programme to improve patients' disability in rheumatoid arthritis: a prospective randomized controlled trial. *Rheumatology (Oxford)*. Apr 2009;48(4):410-415. [doi: [10.1093/rheumatology/ken511](https://doi.org/10.1093/rheumatology/ken511)] [Medline: [19211654](https://pubmed.ncbi.nlm.nih.gov/19211654/)]

24. Crofford LJ. Use of NSAIDs in treating patients with arthritis. *Arthritis Res Ther*. 2013;15 Suppl 3(Suppl 3):S2. [doi: [10.1186/ar4174](https://doi.org/10.1186/ar4174)] [Medline: [24267197](https://pubmed.ncbi.nlm.nih.gov/24267197/)]
25. Salvarani C, Vlachonikolis IG, van der Heijde DM, et al. Musculoskeletal manifestations in a population-based cohort of inflammatory bowel disease patients. *Scand J Gastroenterol*. Dec 2001;36(12):1307-1313. [doi: [10.1080/003655201317097173](https://doi.org/10.1080/003655201317097173)] [Medline: [11761022](https://pubmed.ncbi.nlm.nih.gov/11761022/)]
26. Chinn S, Caldwell W, Gritsenko K. Fibromyalgia pathogenesis and treatment options update. *Curr Pain Headache Rep*. Apr 2016;20(4):25. [doi: [10.1007/s11916-016-0556-x](https://doi.org/10.1007/s11916-016-0556-x)] [Medline: [26922414](https://pubmed.ncbi.nlm.nih.gov/26922414/)]
27. Wang B, Li W, Bradlow A, Bazuaye E, Chan ATY. Improving triaging from primary care into secondary care using heterogeneous data-driven hybrid machine learning. *Decis Support Syst*. Mar 2023;166:113899. [doi: [10.1016/j.dss.2022.113899](https://doi.org/10.1016/j.dss.2022.113899)]
28. Maarseveen TD, Glas HK, Veris-van Dieren J, van den Akker E, Knevel R. Improving musculoskeletal care with AI enhanced triage through data driven screening of referral letters. *NPJ Digit Med*. Feb 14, 2025;8(1):98. [doi: [10.1038/s41746-025-01495-4](https://doi.org/10.1038/s41746-025-01495-4)] [Medline: [39948271](https://pubmed.ncbi.nlm.nih.gov/39948271/)]
29. Schäfer A, Kovacs MS, Nigg A, Feuchtenberger M. Patient-reported outcomes of depression and fibromyalgia symptoms do not predict non-inflammatory versus inflammatory diagnoses at initial rheumatology consultation. *Healthcare (Basel)*. Sep 29, 2024;12(19):1948. [doi: [10.3390/healthcare12191948](https://doi.org/10.3390/healthcare12191948)] [Medline: [39408128](https://pubmed.ncbi.nlm.nih.gov/39408128/)]
30. Gomon G, Raffray M, Pérez - Sancristóbal I, et al. POS0433 comparison of musculoskeletal referral pathways across EUROPE. *Ann Rheum Dis*. Jun 2025;84(Suppl 1):664-665. [doi: [10.1016/j.ard.2025.05.819](https://doi.org/10.1016/j.ard.2025.05.819)]
31. Krusche M, Callhoff J, Knitz J, Ruffer N. Diagnostic accuracy of a large language model in rheumatology: comparison of physician and ChatGPT-4. *Rheumatol Int*. Feb 2024;44(2):303-306. [doi: [10.1007/s00296-023-05464-6](https://doi.org/10.1007/s00296-023-05464-6)] [Medline: [37742280](https://pubmed.ncbi.nlm.nih.gov/37742280/)]
32. Wu J, Dong H, Li Z, et al. A hybrid framework with large language models for rare disease phenotyping. *BMC Med Inform Decis Mak*. Oct 8, 2024;24(1):289. [doi: [10.1186/s12911-024-02698-7](https://doi.org/10.1186/s12911-024-02698-7)] [Medline: [39375687](https://pubmed.ncbi.nlm.nih.gov/39375687/)]
33. Shyr C, Hu Y, Bastarache L, et al. Identifying and extracting rare diseases and their phenotypes with large language models. *J Healthc Inform Res*. Jun 2024;8(2):438-461. [doi: [10.1007/s41666-023-00155-0](https://doi.org/10.1007/s41666-023-00155-0)] [Medline: [38681753](https://pubmed.ncbi.nlm.nih.gov/38681753/)]
34. Soteriades ES, Psalta L, Leka S, Spanoudis G. Occupational stress and musculoskeletal symptoms in firefighters. *Int J Occup Med Environ Health*. Jun 14, 2019;32(3):341-352. [doi: [10.13075/ijomeh.1896.01268](https://doi.org/10.13075/ijomeh.1896.01268)] [Medline: [30938370](https://pubmed.ncbi.nlm.nih.gov/30938370/)]
35. Kivimäki M, Leino-Arjas P, Virtanen M, et al. Work stress and incidence of newly diagnosed fibromyalgia: prospective cohort study. *J Psychosom Res*. Nov 2004;57(5):417-422. [doi: [10.1016/j.jpsychores.2003.10.013](https://doi.org/10.1016/j.jpsychores.2003.10.013)] [Medline: [15581643](https://pubmed.ncbi.nlm.nih.gov/15581643/)]
36. Lewis KL, Battaglia PJ. Differentiating bilateral symptomatic hand osteoarthritis from rheumatoid arthritis using sonography when clinical and radiographic features are nonspecific: a case report. *J Chiropr Med*. Mar 2019;18(1):56-60. [doi: [10.1016/j.jcm.2018.07.005](https://doi.org/10.1016/j.jcm.2018.07.005)] [Medline: [31193241](https://pubmed.ncbi.nlm.nih.gov/31193241/)]
37. Maarseveen TD. FRYQ. Github. URL: <https://github.com/levrex/FRYQ> [Accessed 2025-11-13]

Abbreviations

AUC-ROC: area under the receiver operating characteristic curve

FRYQ: Frysian Questionnaire for Differentiation of Musculoskeletal Complaints

GP: general practitioner

IRD: inflammatory rheumatic disease

JPAST: Joint Pain Assessment Scoring Tool

NLP: natural language processing

RA: rheumatoid arthritis

SHAP: Shapley Additive Explanations

WMO: Wet medisch-wetenschappelijk onderzoek met mensen

Edited by Jeffrey Klann; peer-reviewed by Maxime Raffray, Oliver Sander; submitted 13.May.2025; final revised version received 01.Oct.2025; accepted 01.Oct.2025; published 17.Nov.2025

Please cite as:

Maarseveen TD, Reimann F, al Hasan A, Schilder A, Zhang D, Wink F, Hendriks L, Knevel R, Bos R

Evaluating the Accuracy of the Frysian Questionnaire for Differentiation of Musculoskeletal Complaints for Triage of Musculoskeletal Diseases: Algorithm Development and Validation Study

*JMIR Med Inform*2025;13:e77345

URL: <https://medinform.jmir.org/2025/1/e77345>

doi: [10.2196/77345](https://doi.org/10.2196/77345)

© Tjardo Daniël Maarseveen, Floor Reimann, Ahmed al Hasan, Annemarie Schilder, Dan Zhang, Freke Wink, Lidy Hendriks, Rachel Knevel, Reinhard Bos. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 17.Nov.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.