Master Health Sciences

Combining breath analysis and clinical variables for improved diagnosis of rheumatoid arthritis

Master thesis

Silvie van Grotel
S2255251
June 14, 2024

Name:	Silvie van Grotel
Studentnumber:	S2255251
Date:	June 14, 2024
Assignmentnumber:	HS-2023067-4
Daily supervisor:	Demy Gerritsen
First supervisor:	Prof. dr. Harald Vonkeman
Second supervisor:	Prof. dr. Job van der Palen
Organisation:	Medisch Spectrum Twente, Enschede
Department:	Rheumatology
University:	University of Twente
Programme:	Master Health Sciences
Specialisation:	Optimisation of Healthcare Processes
Faculty:	Science and Technology
Course code:	202300226

Preface

Before you lies the thesis “Combining breath analysis and clinical variables for improved diagnosis of rheumatoid arthritis”. The research for this thesis was conducted at Medisch Spectrum Twente in Enschede. I wrote this thesis as part of my graduation from the master Health Sciences at the University of Twente. I worked on this entirely from home, except for seven days, during the period from January 2024 to June 2024.

Together with my supervisors, Demy Gerritsen, Harald Vonkeman and Job van der Palen, I came up with the topic of my thesis. My thesis is part of several studies with the aeoNose in the rheumatology department of Medisch Spectrum Twente. Since I was working from home and could not engage in conducting research at the hospital, I contributed to the research from a distance. I worked on checking all study participants in the online dataset and whether they matched the actual data.

Within this research, I experienced how dependent you can be on others and my patience has been tested with that. My graduation period also taught me that I am ready to enter the field of work and that studying has been enough for now. I am glad that I started and also finished this master’s thesis and therefore hereby conclude my student life.

I would like to thank my supervisors for their excellent guidance and support during my graduation. Firstly, I would like to thank Demy Gerritsen for answering my questions promptly, collecting the data I used and keeping me updated on the research. I would also like to thank Harald Vonkeman and Job van der Palen for making my graduation possible quickly, their expert advice and quick answers to my questions. Not to mention that all three of them gave me the opportunity to graduate almost entirely from home and, as a result, conduct almost all contact digitally. I am also very grateful to the other students who engaged in research with the aeoNose, with them I was able to discuss things we were running into and thus come to a solution. This all gave me new ideas that helped me to bring this thesis to a good end.

Enjoy reading my thesis!

Silvie van Grotel

Riethoven, June 14, 2024

Abstract

Rheumatoid arthritis is a chronic inflammatory disease, which is characterised by severe damage to joints and bones. A specialist performs an examination on an individual with suspected rheumatoid arthritis to establish the diagnosis, this includes a medical history, physical examination, and blood tests. When treating rheumatoid arthritis, it is important for a patient to start treatment early to avoid irreversible damage.

The aeoNose is a device from The eNose Company in which an individual must breathe in and out through the mouthpiece of the device for 5 minutes. The aeoNose measures volatile organic compounds in exhaled air and, through the use of an algorithm, identifies different breath profiles. By analysing these breath profiles, the aeoNose's algorithm can be trained to better distinguish between individuals with a particular diagnosis and individuals without that particular disease.

Within this current study, sociodemographic variables will be added to the analyses to see if this will allow the model to better distinguish between individuals with rheumatoid arthritis and without rheumatoid arthritis. These are the variables age, gender, body mass index and smoking status. First, the differences in sociodemographic variables between control group, individuals without rheumatoid arthritis, and patient group, individuals with rheumatoid arthritis patients, are assessed. The eNose Company then calculates the probability of the presence of rheumatoid arthritis based on the volatile organic compounds in the exhaled air from the aeoNose measurements. Next, univariate and multivariate logistic regression analyses are used to calculate the probability of rheumatoid arthritis based on the aeoNose data and sociodemographic variables alone and combined. Based on the sensitivity, specificity, positive predictive value, negative predictive value, receiver operating characteristic curve and area under the curve, the different models are compared.

The aim of this study is to see whether the aeoNose can contribute to the diagnosis of rheumatoid arthritis, with the aim of deploying the aeoNose in general practices. Because of this, the focus is on the diagnostic parameters of sensitivity and negative predictive value, as it is important not to miss rheumatoid arthritis patients who should be receiving treatment. Therefore, the number of false negatives should be as low as possible.

The results show that there is a negligible difference between the area under the curve from the receiver operating characteristic of the model with the aeoNose classification value and the model with the sociodemographic variables. The model that combines the aeoNose classification value with the sociodemographic variables outperforms the other two models. Indeed, this model has an area under the curve from the receiver operating characteristic of 0.750 which is 0.057 and 0.073 more than of the models with only the sociodemographic variables and only the aeoNose classification value respectively.

Ultimately, it can be concluded that adding the sociodemographic variables; age and gender, to the analysis of aeoNose provides better distinguishing individuals with and without rheumatoid arthritis. The sociodemographic variables combined with aeoNose classification value gives the best area under the curve from the receiver operating characteristic relative to both separately. This means that the aeoNose can be used in combination with age and gender to help general practitioners better refer individuals with suspected rheumatoid arthritis to a specialist.

Keywords: *aeoNose, breath analysis, rheumatoid arthritis, sociodemographic variables.*

Table of Contents

Preface 2
Abstract 3
Abbreviations 5
Introduction 6
Method 9
Results 11
Discussion and conclusion 14
Strengths and limitations 14
Future research 15
Conclusion 15
Bibliography 16
Appendix A. aeoNose 18
Appendix B. Informed consent 19

Abbreviations

Abbreviation	Definition
ACR	American College of Rheumatology
Anti-CCP	Anti-cyclic citrullinated peptides
AUC	Area under the curve
BMI	Body mass index
BSE	Bezinkingssnelheid van de erytrocyten
COPD	Chronic obstructive pulmonary disease
CRP	C-reactive protein
eNose	Electronic nose
ESR	Erythrocyte sedimentation rate
EULAR	European Alliance of Associations for Rheumatology
FN	False negatives
FP	False positives
GP	General practitioner
MST	Medisch Spectrum Twente
NPV	Negative predictive value
PPV	Positive predictive value
RA	Rheumatoid arthritis
RF	Rheumatoid factor
ROC	Receiver operating characteristic
TN	True negatives
TP	True positives
VOCs	Volatile organic compounds

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory disease characterised by severe joint and bone damage [1–3]. In RA there is a heightened autoimmune response at joint sites. In 2021, there were nearly 270,000 people diagnosed with RA in the Netherlands, of whom about 170,000 were female and 100,000 were male [4]. While the exact cause of RA is unknown, it is more common and severe in individuals with a specific genetic predisposition [2,5]. RA is also more prevalent in females, hormones could play a role in the course and development of the disease [6]. Thereby, smoking is a major risk factor in the development of the disease and can make antirheumatic medication less effective.

The American College of Rheumatology (ACR) and the European Alliance of Associations for Rheumatology (EULAR) established criteria which allow classification of RA [7]. A specialist will perform an examination on an individual with suspected RA. This involves examining the number of involved joints, serology; this involves rheumatoid factor (RF) and anti-cyclic citrullinated peptides (anti-CCP), the acute-phase reactant; in this the C-reactive protein (CRP) and the erythrocyte sedimentation rate (ESR, in Dutch: BSE) of the blood and the duration of symptoms is considered; a distinction is made between less or more than 6 weeks.

Persistent joint inflammation in RA patients can lead to joint and bone damage, which can lead to deterioration in performance of daily activities and other physical activities [8,9]. Early treatment is crucial as joint damage can occur within the first year [10,11]. Joint damage is strongly related to disease activity, reducing inflammation promptly is vital to prevent irreversible damage [11,12]. Besides medication, maintaining physical activity and a healthy lifestyle is essential for RA patients.

The patient journey of an individual with RA will usually start at a general practitioner (GP); in case of symptoms, a person will initially report to a GP. As soon as the GP suspects RA, the GP should refer the patient to the rheumatologist as soon as possible [12]. Seeing a rheumatologist within 6 weeks of the onset of initial symptoms increases the chances for achieving sustained drug-free remission, compared to referral between 7 and 12 weeks [13]. Prompt referral to a rheumatologist is particularly important because joint damage can occur in the first year of the disease [10,11]. A study by Scott et al. found that referrals are not always made quickly and therefore examined the barriers for quick referral [14]. This research among English GPs showed that a patient's medical history had the most influence on a referral to a specialist. In contrast, RF and anti-CCP had the least influence on referral. It was also found that GPs want to perform tests, such as measuring RF, CRP and BSE, before referral to support their clinical opinion when referring. As a result, referral to a rheumatologist will take longer than it could and should have.

When an individual with suspected RA reports to the GP, the GP has several options. The GP can, based on a diagnostic test, such as the application of certain criteria, refer an individual to a rheumatologist with certainty (very high prior probability of RA), definitely not refer an individual to a rheumatologist (very low prior probability of RA), and in addition, there is also a group with in-between prior probability of RA. For this in-between group, it is important that those with actual RA are referred to a rheumatologist as soon as possible to start treatment quickly. Within this group, there are also individuals who do not have RA but are still referred to a rheumatologist by a GP, which can be considered as false positives (FP). Conversely, there are also individuals who actually have RA but are not referred to a rheumatologist, these can be considered as false negatives (FN). For the best treatment of the individual with RA, it is important that these individuals are referred to a rheumatologist as soon as possible to start treatment. So, there should be as few false negatives as possible. The number of false positives should also be as low as possible, as this costs time and money for the entire care chain and thus puts more pressure on the care chain. Within diagnostic tests, sensitivity and specificity are considered, using the gold standard as the best test [15]. The gold

standard is the examination performed by a specialist using the established criteria by the ACR and EULAR [7]. Sensitivity is the ability of a test to correctly classify an individual with the disease. Specificity, on the other hand, is the ability of a test to correctly classify an individual free of the disease. Furthermore, positive and negative predictive value should also be considered. The positive predictive value (PPV) is the percentage of individuals with a positive test who have the disease according to the gold standard. The percentage of individuals with a negative test who are free of the disease, according to the gold standard, is the negative predictive value (NPV). See figure 1 for a clarification of the terms.

	Positive diagnosis (gold standard)	**Negative diagnosis (gold standard)**	**Measures**
Positive prediction	True positive (TP)	False positive (FP)	$PPV = \frac{TP}{TP + FP}$
Negative prediction	False negative (FN)	True negative (TN)	$NPV = \frac{TN}{FN + TN}$
Measures	$Sensitivity = \frac{TP}{TP + FN}$	$Specificity = \frac{TN}{FP + TN}$	

Figure 1: clarification of diagnostic parameters, PPV means positive predictive value and NPV means negative predictive value [16].

A receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier [16]. Figure 2 illustrates a ROC curve for clarification. The horizontal axis shows the 1-specificity and the vertical axis the sensitivity. The area under the curve (AUC) represents the performance measure for a classification model, a higher value signifies a better performance. An AUC value of 1 represents a perfect distinction between true positives and true negatives.

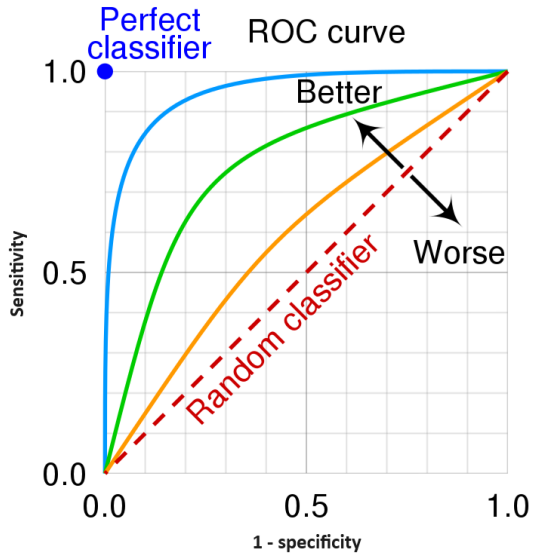

Figure 2: ROC curve with clarification [17].

In the Netherlands, healthcare costs have risen by several billions in recent years [18]. Adding to this, there is also a staff shortage within the care sector [19]. It is therefore important that as few false positives as possible are referred to a rheumatologist to reduce the pressure on the care sector. In this regard, it is also important to minimise false negatives to prevent patients from not starting treatment early enough. Thus, it would be useful for the whole patient journey to find a way to ensure that the GP can refer individuals with suspected RA to the rheumatologist with more accuracy.

Recently, a new possibility has emerged that could potentially help diagnose RA. Namely the aeoNose™ (The eNose Company, Zutphen, the Netherlands) that can analyse exhaled air based on the presence of volatile organic compounds (VOCs) [20–23]. Exhaled air contains more than 1,500 different VOCs that reflect the metabolism of the human body [24]. The aeoNose measures VOCs in exhaled air and, through the use of an algorithm, recognises and diagnoses various conditions by analysing different breath profiles. Appendix A shows an explanation of aeoNose.

In 2023, Demy Gerritsen conducted a study on the aeoNose's ability to diagnose RA using breath analysis [25]. Gerritsen's study showed that the training set had a sensitivity of 72%, a specificity 57%, a PPV of 64% and a NPV of 66%. The blind set had a sensitivity of 72%, a specificity of 48%, a PPV of 58% and a NPV of 63%. The AUC for the training set was 0.66 and for the blind set 0.70, meaning the blind set is a better performing model. In Gerritsen´s study, blind and training data were used, the training data was used to train an Artificial Neural Network [21]. The blind data was used to assess the model's true ability to classify data that the model has not seen before, this validates the model.

The aim of this study is to see whether the aeoNose can contribute to the diagnosis of rheumatoid arthritis, with the aim of deploying the aeoNose in general practices. Because of this, the focus is on the diagnostic parameters of sensitivity and NPV. This is done by combining aeoNose classification value with the analysis sociodemographic characteristics, such as age and gender. Research by Kort et al. showed that adding variables to aeoNose analysis, in the detection of lung cancer, improved diagnostic performance [22]. That study included the variables age, gender, smoking status, smoking exposure in pack-years, chronic obstructive pulmonary disease (COPD) and body mass index (BMI). This current study examines whether the aeoNose diagnosis of RA can also be improved by adding variables. Sociodemographic characteristics such as gender, age, BMI and smoking status known in both the patient and control group will be examined. The outcome of the analysis from the aeoNose data and sociodemographic characteristics will be compared. For aeoNose and sociodemographic variables separately, there is a desired sensitivity of at least 72%. So that this is equal or greater than the sensitivity in Gerritsen´s study mentioned earlier [25]. The aeoNose and variables combined should have a desired sensitivity of at least 80% with a corresponding specificity, by choosing a relevant threshold value in the ROC curve. Here, a sensitivity of at least 80% is desired because this is seen as an acceptable sensitivity [26,27]. The following question is therefore answered in this study: "How can sociodemographic variables contribute to improving the analysis of the aeoNose breath test in distinguishing individuals with and without rheumatoid arthritis?".

Method

In this study, individuals aged 18 years or older and officially diagnosed with RA by a rheumatologist were included in the patient group of the study. The control group consisted of individuals without RA, mainly employees of Medisch Spectrum Twente (MST) or family, friends or caretakers of patients attending the rheumatology department of MST; they also had to be 18 years or older. Individuals with insufficient understanding of the Dutch language to understand and sign informed consent form or individuals with insufficient capacity to perform breath test on aeoNose were excluded from both the patient and control group. For the current analysis, if an individual had another rheumatic condition, that person was excluded from the control group. The aeoNose measurements took place at MST, either in the rheumatology department or at the blood collection office. Before an individual could participate in the study, the informed consent form in Dutch (see appendix B) had to be signed by the participant and the researcher.

Study participants in the patient group were recruited through the rheumatology department's consultation hours. Once someone arrived at the outpatient department, prior to their appointment, the person was informed about the study, invited to participate and handed the informed consent form. When the individual had completed their appointment, they were approached again and asked if they wanted to participate and agree to the informed consent. If the individual agreed to the informed consent, he or she was given a detailed explanation of how the aeoNose works. Once this explanation was complete, breath analysis started, involving 5 minutes of unforced inhaling and exhaling through the aeoNose. In this, the participant was given a nose clamp to ensure that all air was breathed in and out through the mouth into the aeoNose. After the procedure, the participant was queried on sociodemographic data, medical data, and their experience of the aeoNose procedure. The procedure for participants in the control group was largely similar, except for having a consultation with the rheumatologist. If an MST employee was asked to participate in the study, the same steps followed from informed consent. The following data were gathered from the study participants in the patient group:

1. diagnosis of rheumatic disease;
2. RA erosivity;
3. number of swollen joints;
4. duration of illness longer than 6 weeks;
5. positive or negative anti-CCP;
6. positive or negative rheumatoid factor;
7. BSE value and date;
8. CRP value and date;
9. active disease state;
10. date of birth;
11. age;
12. gender;
13. height;
14. weight;
15. smoking status, if still smoking how long and number of cigarettes per day, if stopped indicated since when;
16. comorbidities;
17. medication use;
18. eaten less than or more than 3 hours before taking the test;
19. whether individual would participate in the test again;
20. discomfort score and
21. success of the test.

For the control group, only variables 10 to 21 were known.

Regarding sociodemographic data, only age, gender, BMI and smoking status were considered. BMI was used instead of weight and height because these measures say little about a person's physique separately from each other. Continuous variables were reported as mean with corresponding standard deviation or as median with interquartile range. The nominal variables were reported as numbers with corresponding percentages. For normally distributed continuous variables a t-test was applied; for skewed distributed continuous or ordinal variables, a Mann-Whitney U-test was applied; and for nominal and categorical variables, a Chi-squared test was applied to assess differences between groups.

Several models were used for the analyses in this study namely:

1. only the aeoNose data;
2. only the sociodemographic characteristics and
3. sociodemographic characteristics combined with aeoNose data.

The participants' exhaled air measured by the aeoNose was analysed by The eNose Company. This revealed a probability between -1 and 1 for the presence of RA in each individual according to the aeoNose model 1. Based on this, a logistic regression was performed, then the sensitivity, specificity, PPV and NPV were calculated. Lastly, the ROC curve and AUC were composed. For models 2 and 3, the analysis was carried out in the following way. First, a univariate logistic regression was performed for each variable. Next, a multivariate logistic regression was performed. The sensitivity, specificity, PPV, NPV, ROC curve and AUC were then calculated again for the diagnosis of RA according to multivariate logistic regression analyses. For the three models, a relevant threshold value in the ROC curve was chosen, focusing on high sensitivity and high NPV. The results from the different models were compared. All statistical tests had a significance level of 0.05 and were performed in SPSS.

Results

The study population included 207 controls and 222 RA patients. Study participants were included from June 2021 through February 2024. See table 1 for the sociodemographic characteristics of the study population. For smoking status and BMI, some data are missing due to data storage errors, therefore these individuals were excluded from the study. These were 2 individuals in the control group and 7 in the patient group. The continuous variables age and BMI were visually observed to be normally distributed, therefore a t-test was performed. Finally, the Chi-squared test was performed for the variables gender and smoking status.

Table 1: sociodemographic variables control group, RA patients and total study population.

Variable	Control group	RA patients	Total study population	P-value
Subjects	205 (48.8%)	215 (51.2%)	420 (100%)	
Age in years	55 (22.0)	65 (18.0)	60 (21)	<0.001
Female gender	122 (59.5%)	150 (69.8%)	272 (64.8%)	0.028
BMI	26.3 (6.2)	26.4 (5.1)	26.3 (5.6)	0.924
Smoking status				0.348
Never smoker	98 (47.8%)	92 (42.8%)	190 (45.2%)	
Stopped	78 (38,0%)	82 (38,1%)	160 (38,1%)	
Current smoker	29 (14.1%)	41 (19,1%)	70 (16.7%)	
Data are presented as n (%) or median (interquartile range). The percentages are relative to the respective group; control group, patient group or total study population.				

Table 1 shows that RA patients were significantly older and more likely to be female. There is no significant difference in BMI and smoking status between controls and patients.

Table 2: results of the univariate and multivariate logistic regression analyses from the sociodemographic characteristics for diagnosing RA.

Variable	Univariate analysis	Multivariate analysis	Regression coefficient *
Age	1.040 (1.027, 1.055)	1.047 (1.032, 1.062)	0.046
Female gender	1.570 (1.049, 2.349)	1.929 (1.242, 2.995)	0.657
BMI	0.998 (0.959, 1.038)	0.982 (0.941, 1.025)	-0.018
Smoking status			
Never smoker	Reference	Reference	
Stopped	0.664 (0.382, 1.156)	0.883 (0.554, 1.407)	-0.125
Current smoker	0.744 (0.421, 1.312)	1.665 (0.921, 3.012)	0.510
Data are presented as odds ratio (95% confidence interval) unless otherwise stated. ** Constant -2.616*			

Table 2 shows that each additional year of age was associated with a 4.7% higher chance of having RA. Females have a 92.9% higher chance of having RA. Someone with a higher BMI has a lower chance of getting RA, however, this was not statistically significant. Individuals who stopped smoking also have a lower chance of RA than never smoking, with current smoking being more likely to cause RA than never smoked. However, these were also not statistically significant. This is shown by the multivariate analysis using only the sociodemographic variables. The univariate analysis showed a negative association between current smoker and presence of RA, however, the multivariate analysis showed a positive association. This may involve a confounding variable where current smoker may be associated with another independent variable such as age or gender. Correcting for these confounders in the multivariate analysis may result in an odds ratio greater than 1.

By choosing a relevant threshold value in the ROC curve, focusing on high sensitivity and high NPV, the multivariate logistic regression analysis, based on sociodemographic variables only, showed a sensitivity of 72.1%, a specificity of 50.7%, a PPV of 60.5% and a NPV of 63.4%. Where 0.483 represents the probability of RA within this model. This corresponded with an AUC-ROC of 0.693 (95% confidence interval 0.643, 0.743).

The logistic regression analysis based on the aeoNose data only showed a sensitivity of 72.6%, a specificity of 52.7%, a PPV of 61.7% and a NPV of 64.7%. Within this model, 0.463 represents the probability of RA. This corresponded with an AUC-ROC of 0.677 (95% confidence interval 0.626, 0.727), which means that this model performs worse than the model above with only the sociodemographic variables.

Table 3: results of the univariate and multivariate logistic regression analyses from the sociodemographic characteristics and aeoNose data for diagnosing RA.

Variable	**Univariate analysis**	**Multivariate analysis**	**Regression coefficient ***
Age	1.040 (1.027, 1.055)	1.045 (1.030, 1.061)	0.044
Female gender	1.570 (1.049, 2.349)	2.068 (1.302, 3.286)	0.727
BMI	0.998 (0.959, 1.038)	0.989 (0.946, 1.035)	-0.11
Smoking status			
Never smoker	Reference	Reference	
Stopped	0.664 (0.382, 1.156)	0.984 (0.604, 1.602)	-0.16
Current smoker	0.744 (0.421, 1.312)	1.767 (0.948, 3.292)	0.569
AeoNose classification value	10.140 (4.779, 21.542)	9.961 (4.501, 22.045)	2.299
Data are presented as odds ratio (95% confidence interval) unless otherwise stated. ** Constant -2.815*			

The multivariate analysis with the sociodemographic variables and aeoNose classification value in table 3 shows that each additional year of age was associated with a 4.5% higher chance of having RA. Females have a two-fold increased chance of having RA. The association between the presence of RA and BMI and smoking status were both not statistically significant. The classification value of the aeoNose is strongly associated with the presence of RA.

The multivariate logistic regression analysis, based on sociodemographic variables and aeoNose classification value, showed a sensitivity of 80.0%, a specificity of 52.7%, a PPV of 63.9% and a NPV of 71.5%. Within this model, the probability of RA is 0.430. This corresponded with an AUC-ROC of 0.750 (95% confidence interval 0.704, 0.796), which means this model outperforms the two aforementioned models.

Figure 3 shows the ROC curves of the sociodemographic variables, the aeoNose data and these combined. This figure shows the improved performance of the multivariate model with the sociodemographic variables and the aeoNose classification value, relative to both separately.

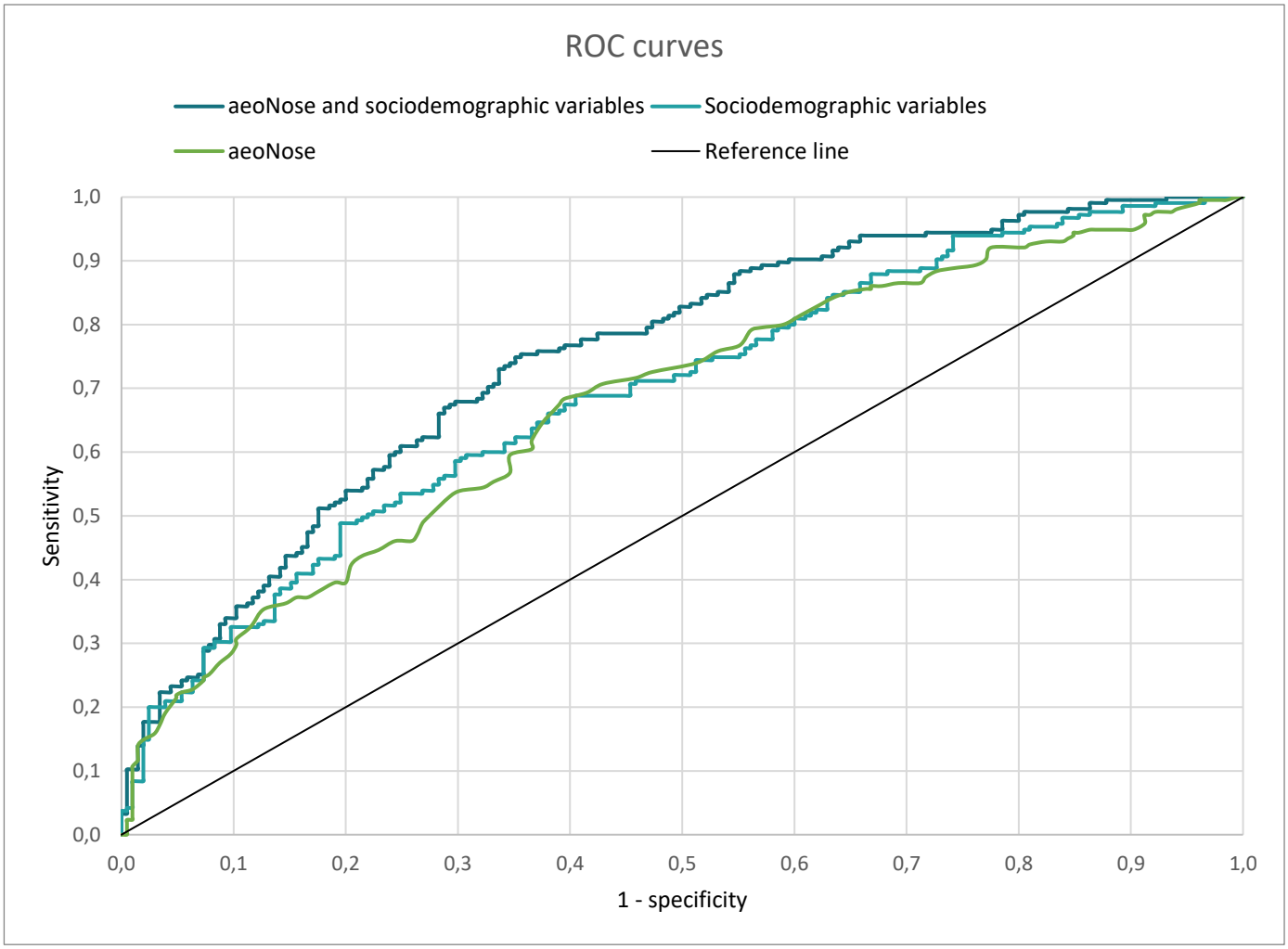

Figure 3: ROC curves combined showing three predictions for having RA: aeoNose and sociodemographic variables and both separately.

Discussion and conclusion

This study shows that the aeoNose classification value is strongly associated with the presence of RA, the univariate and multivariate analysis showed. The multivariate analysis shows that each additional year of age is associated with a 4.5% higher chance of having RA. Research by Chen et al. also shows that each increasing year of life increases the risk of RA of between 1% and 10% [28]. The analysis in current study shows that females have a two-fold increased chance of having RA than males. This is consistent with other research showing that RA is two or even three times more common in females than in males [29]. BMI and smoking status did not significantly affect the risk of developing RA, according to the analyses. Other studies show that a higher BMI may contribute to a higher risk of developing RA, so this is not consistent with current research [30,31]. However, these do indicate that there may be confounding factors, such as female gender. Previous research shows in contrast to current research that smoking is related to RA development and risk [32]. Although, this considers smoking intensity and duration, which was not fully examined within the current study.

The AUC-ROC of the model with the aeoNose classification value is lower than that of the model with the sociodemographic variables. This means a poorer performance of the model with the aeoNose classification value for distinguishing RA. However, this is a very small difference of 0.016 which is negligible. The results show that the aeoNose classification value is strongly associated with the presence of RA. In 2016, a study was conducted using an electronic nose (eNose), which also showed that an eNose can distinguish breath profiles from patients with active RA and controls without RA [23]. The model that combines the aeoNose classification value with the sociodemographic variables outperforms the other two models. Indeed, this model has a AUC-ROC of 0.750 which is 0.057 and 0.073 more than of the models with only the sociodemographic variables and aeoNose classification value respectively. Thus, the model of aeoNose classification value together with sociodemographic variables is a better predictor of the presence of RA than both models separately. Research on aeoNose combined with clinical variables in diagnosing lung cancer also showed that clinical variables and aeoNose together give better diagnosis of lung cancer than each separately [21,22]. Due to higher sensitivity and NPV in the model in current study in which the aeoNose classification value was combined with the variables than in the models separately, it can be said that using the variables and aeoNose classification value together in the analysis reduces the number of false negatives. As a result, fewer patients with RA who could otherwise have been treated are missed. The specificity of the aeoNose classification value alone was better than of the variables alone. However, the specificity of the model with only the aeoNose classification value did not increase or decrease after adding the variables to the analysis. Although, these values are determined by choosing certain probabilities of RA within each model separately. Therefore, AUC-ROC is considered to assess model performance. On that basis, there is a negligible difference between the sociodemographic variables alone and the aeoNose classification value alone, but the model combining the two performs better in distinguishing individuals with and without RA.

Strengths and limitations

A strength of aeoNose is that a breath test only takes 5 minutes. Thereby, the aeoNose is a very user-friendly device whose participants experienced little to no discomfort. Study participants gave the breath test with the aeoNose a mean discomfort score of 2.4 on a scale of 0 to 10 where 0 means no discomfort at all and 10 stands for unbearable.

This study showed that RA patients were significantly older than controls without RA and more likely to be a female. Previous research have also shown that RA is more common in females [29]. However, within this study, the significant differences in age and gender cannot be directly traced to the presence or absence of RA. When recruiting study participants, there was no consideration of whether the samples of controls without RA or RA patients were representative of the entire population. It appears

that many young people, working in MST, have taken a breath test who belong to the control group. As a result, the control group may be younger than the entire population of individuals without RA, making it not directly traceable to the presence of RA.

This study found that there was no significant association between current smokers, quit smoking or never smoked and the presence of RA. Previous research does show that smoking intensity and duration is related to the development and risk of RA [32]. A limitation of current study is that individuals who indicated they had stopped smoking were not asked how long they had smoked and how much they smoked then. Among those who stopped smoking, only when they stopped smoking is known. The smoking individuals are known how long they have been smoking and how many cigarettes they smoke per day. For those smoking, the number of pack-years could be calculated, which is a clinical quantification to measure a person's exposure to tobacco [22,32]. The number of pack-years provides more information about the intensity and duration of smoking than smoking status; current smoker, stopped smoking or never smoked. This may explain that in the current study no association was found between smoking status and the presence of RA but possibly when using pack-years an association was found.

Future research

Future research can be conducted with RA related variables. Thus, the variables number of swollen joints, anti-CCP, RF, BSE and CRP can also be examined and tested in control participants. If these variables are then known in the control and RA patient group, similar analyses can be performed as in this study. Also, the RA related variables can then be analysed according to their diagnostic performance for distinguishing RA. Then, the RA related variables can also be combined in the analysis with the sociodemographic variables alone, aeoNose classification value alone and with both together. This way, more can become clear about the possible contribution of RA related variables in combination with sociodemographic variables and the aeoNose classification value in diagnosing RA. The downside to this is that these tests cannot be performed at the GP's office and therefore do not add value when using aeoNose within GP practices. However, it could contribute in diagnosis by a specialist. This would require further research as to whether this has advantages over the current method of diagnosis.

Other future research can be done by conducting a similar research. Only that when recruiting study participants, make sure the sample is representative of the population. This can also be investigated further in this study. This is to more confidently trace differences between individuals with and without RA to the disease.

Since previous research shows that the intensity and duration of smoking affects the risk and development of RA, it is interesting to investigate this further [32]. Further queries on time smoked and amount of smoked could then be asked of all study participants. This could then be used to calculate the number of pack-years among all participants. The number of pack-years can then be added to the model as a variable. The advantage of this is that it can also be asked out by a GP. If pack-years contributes to better differentiation of individuals with and without RA by the aeoNose, then this could be of added value within GP practices when using the aeoNose in diagnosing RA.

Conclusion

Ultimately, it can be concluded that adding the sociodemographic variables; age and gender, to the analysis of aeoNose provides better distinguishing individuals with and without RA. The sociodemographic variables combined with aeoNose classification value gives the best AUC-ROC relative to both separately. This means that a combination of both can add value in the diagnosis of RA. The aeoNose combined with age and gender could be used in GP practices to support referral of suspected RA patients to a specialist.

Bibliography

[1] Lau CS, Chia F, Harrison A, Hsieh T, Jain R, Jung SM, et al. APLAR rheumatoid arthritis treatment recommendations. Int J Rheum Dis 2015;18:685–713. https://doi.org/10.1111/1756-185X.12754.

[2] Calabresi E, Petrelli F, Bonifacio AF, Puxeddu I, Alunno A. One year in review 2018: Pathogenesis of rheumatoid arthritis. Clin Exp Rheumatol 2018;36:175–84.

[3] Prasad P, Verma S, Surbhi, Ganguly NK, Chaturvedi V, Mittal SA. Rheumatoid arthritis: advances in treatment strategies. Mol Cell Biochem 2023;478:69–88. https://doi.org/10.1007/s11010-022-04492-3.

[4] Nivel zorgregistraties. Reumatoïde artritis (RA) | Leeftijd en geslacht | Volksgezondheid en Zorg. Vzinfo RIVM 2022. https://www.vzinfo.nl/reumatoide-artritis/Leeftijd-en-geslacht (accessed January 30, 2024).

[5] Hulshof T. Reumatoïde artritis (RA) | Oorzaken en gevolgen | Volksgezondheid en Zorg. Vzinfo RIVM n.d. https://www.vzinfo.nl/reumatoide-artritis/oorzaken-en-gevolgen# (accessed January 30, 2024).

[6] Alpízar-Rodríguez D, Finckh A. Environmental factors and hormones in the development of rheumatoid arthritis. Semin Immunopathol 2017;39:461–8. https://doi.org/10.1007/s00281-017-0624-2.

[7] Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. Rheumatology 2012;51:vi5–9. https://doi.org/10.1093/rheumatology/kes279.

[8] Matcham F, Scott IC, Rayner L, Hotopf M, Kingsley GH, Norton S, et al. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: A systematic review and meta-analysis. Semin Arthritis Rheum 2014;44:123–30. https://doi.org/10.1016/j.semarthrit.2014.05.001.

[9] Metsios GS, Lemmey A. Exercise as Medicine in Rheumatoid Arthritis: Effects on Function, Body Composition, and Cardiovascular Disease Risk. J Clin Exerc Physiol 2015;4:14–22. https://doi.org/10.31189/2165-6193-4.1.14.

[10] Richtlijnendatabase. Optimale behandelstrategie bij reumatoïde artritis - Richtlijn - Richtlijnendatabase 2019. https://richtlijnendatabase.nl/richtlijn/reumato_de_artritis_ra/optimale_behandelstrategie_bij_reumatoide_artritis.html (accessed February 19, 2024).

[11] Janssens H, Lopuhaä D, Schaafstra A, Shackleton D, Van der Helm-Van Mil A, Van der Spruit R, et al. Artritis | NHG-Richtlijnen. 2017.

[12] Zorginstituut Nederland. Reumatoïde artritis n.d. https://www.farmacotherapeutischkompas.nl/bladeren/indicatieteksten/reumatoide_artritis (accessed February 15, 2024).

[13] Niemantsverdriet E, Dougados M, Combe B, van der Helm-van Mil AHM. Referring early arthritis patients within 6 weeks versus 12 weeks after symptom onset: an observational cohort study. Lancet Rheumatol 2020;2:e332–8. https://doi.org/10.1016/S2665-9913(20)30061-8.

[14] Scott IC, Mangat N, MacGregor A, Raza K, Mallen CD, Hider SL. Primary care challenges in diagnosing and referring patients with suspected rheumatoid arthritis: a national cross-sectional GP survey. Rheumatol Adv Pract 2018;2. https://doi.org/10.1093/rap/rky012.

[15] Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008;56:45. https://doi.org/10.4103/0301-4738.37595.

[16] Twisk JWR. Inleiding in de toegepaste biostatistiek. 4th ed. Houten: Bohn Stafleu van Loghum; 2016. https://doi.org/10.1007/978-90-368-1534-5.

[17] Cmglee, MartinThoma. ROC curve image 2021.

[18] Zorginstituut Nederland. Totale zorgkosten Zorgverzekeringswet | Zorgcijfersdatabank.nl.

ZorgcijfersdatabankNl 2023. https://www.zorgcijfersdatabank.nl/databank?infotype=zvw&label=00-totaal&geg_zvw=jjaarNEW&geg_wlz=jjaarNEW&meta_tabel=kosten&tabel=B_kost&item= (accessed March 20, 2024).

[19] Dashboard Arbeidsmarkt Zorg en Welzijn (AZW). Dashboard Arbeidsmarkt Zorg en Welzijn (AZW). DashboardsCbsNl n.d. https://dashboards.cbs.nl/v4/AZWDashboard/ (accessed March 20, 2024).

[20] Kort S, Brusse-Keizer M, Gerritsen J-W, van der Palen J. Data analysis of electronic nose technology in lung cancer: generating prediction models by means of Aethena. J Breath Res 2017;11:026006. https://doi.org/10.1088/1752-7163/aa6b08.

[21] Kort S, Brusse-Keizer M, Schouwink H, Citgez E, de Jongh FH, van Putten JWG, et al. Diagnosing Non-Small Cell Lung Cancer by Exhaled Breath Profiling Using an Electronic Nose. Chest 2023;163:697–706. https://doi.org/10.1016/j.chest.2022.09.042.

[22] Kort S, Brusse-Keizer M, Gerritsen JW, Schouwink H, Citgez E, de Jongh F, et al. Improving lung cancer diagnosis by combining exhaled-breath data and clinical parameters. ERJ Open Res 2020;6:00221–2019. https://doi.org/10.1183/23120541.00221-2019.

[23] Brekelmans MP, Fens N, Brinkman P, Bos LD, Sterk PJ, Tak PP, et al. Smelling the Diagnosis: The Electronic Nose as Diagnostic Tool in Inflammatory Arthritis. A Case-Reference Study. PLoS One 2016;11:e0151715. https://doi.org/10.1371/journal.pone.0151715.

[24] Team Zorg Enablers. Aeonose Zorg Enablers. Zorg Enablers 2020. https://zorgenablers.nl/aeonose/ (accessed February 22, 2024).

[25] Gerritsen D. Future possibility: the ability to diagnose Rheumatoid Arthritis through breath analysis 2023.

[26] Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: Understanding the properties of diagnostic tests – Part 1. Perspect Clin Res 2018;9:40. https://doi.org/10.4103/picr.PICR_170_17.

[27] Olliaro P, Torreele E. Managing the risks of making the wrong diagnosis: First, do no harm. Int J Infect Dis 2021;106:382–5. https://doi.org/10.1016/j.ijid.2021.04.004.

[28] Chen L, Wu B, Mo L, Chen H, Zhao Y, Tan T, et al. Associations between biological ageing and the risk of, genetic susceptibility to, and life expectancy associated with rheumatoid arthritis: a secondary analysis of two observational studies. Lancet Heal Longev 2024;5:e45–55. https://doi.org/10.1016/S2666-7568(23)00220-9.

[29] Gerosa M, De Angelis V, Riboldi P, Meroni P. Rheumatoid Arthritis: A Female Challenge. Women's Heal 2008;4:195–201. https://doi.org/10.2217/17455057.4.2.195.

[30] Qin B, Yang M, Fu H, Ma N, Wei T, Tang Q, et al. Body mass index and the risk of rheumatoid arthritis: a systematic review and dose-response meta-analysis. Arthritis Res Ther 2015;17:86. https://doi.org/10.1186/s13075-015-0601-x.

[31] Feng J, Chen Q, Yu F, Wang Z, Chen S, Jin Z, et al. Body Mass Index and Risk of Rheumatoid Arthritis. Medicine (Baltimore) 2016;95:e2859. https://doi.org/10.1097/MD.0000000000002859.

[32] Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking Intensity, Duration, and Cessation, and the Risk of Rheumatoid Arthritis in Women. Am J Med 2006;119:503.e1-503.e9. https://doi.org/10.1016/j.amjmed.2005.09.053.

[33] the eNose company. Over ons - The eNose Company. ENose Co n.d. https://www.enose-company.com/over-ons/ (accessed February 22, 2024).

[34] eNose - LINK 2021. https://linkmagazine.nl/enose-heeft-met-opsporen-ziektes-toegevoegde-waarde-voor-patient-en-specialist-maar-marktgang-gehinderd-door-perverse-prikkels/?v=796834e7a283 (accessed March 4, 2024).

Appendix A. aeoNose

The aeoNose is a CE-certified handheld portable device that has been in development for several decades with the objective of becoming the gold standard in disease screening, see figure 4 [33]. An individual must breathe in and out through the mouthpiece of the aeoNose for 5 minutes, using a nose clamp to ensure that the entire breath passes through the mouth. The device is developed by The eNose Company in Zutphen, the Netherlands, the aeoNose analyses exhaled air for the presence of volatile organic compounds (VOCs) [20–23]. Exhaled air contains over 1,500 different VOCs that reflect the body's metabolism [24]. The aeoNose measures these VOCs and, using an algorithm, recognizes and diagnoses various conditions by analysing different breath profiles.

Through breath profile analysis, the aeoNose's algorithm is trained to distinguish between individuals with specific diagnoses and individuals without that diagnoses [20–23]. The eNose Company utilises proprietary software for data analysis, called 'Aethena', which retrieves raw data from a database and manages data compression, analysis, and reporting. During an exhaled breath measurement, 64 x 36 data points are recorded for each sensor, resulting in a data matrix with thousands of records for each individual measurement. The aeoNose stands out from other electronic nose devices by offering the ability to transfer calibration models, thereby enabling large-scale applications [20].

Figure 4: aeoNose from The eNose Company [34].

Appendix B. Informed consent

Medisch Spectrum Twente
een santeon ziekenhuis

Het opsporen van inflammatoire reumatische ziekten met een elektronische neus.

Postadres
Postbus 50 000
7500 KA Enschede

Bezoekadres
Koningsplein 1
7512 KZ Enschede
T (053) 487 20 00
www.mst.nl

Medisch Spectrum △ Twente
Goedgekeurd
Datum 20-01-20
Handtekening

INFORMATIE OVER DEELNAME AAN EEN WETENSCHAPPELIJK ONDERZOEK

TITEL: Het opsporen van inflammatoire reumatische ziekten met een elektronische neus.
ONDERZOEKERS: dr. H.E. Vonkeman, mw. dr. M. Ghiti Moghadam
CENTRUM: Medisch Spectrum Twente

Inleiding
U wordt gevraagd om deel te nemen aan een medisch-wetenschappelijk onderzoek (zie titel). In dit onderzoek wordt onderzocht of stoffen in de uitgeademde lucht kunnen helpen bij het stellen van de diagnose.
U beslist zelf of u mee wilt doen. Voordat u een beslissing neemt, is het belangrijk dat u over de benodigde informatie beschikt om te kunnen beslissen of u wilt deelnemen. Een arts of één van de onderzoekers zal het onderzoek met u bespreken en al uw vragen beantwoorden. U mag ook met familie en vrienden over uw beslissing praten. Neem alstublieft voldoende tijd om te beslissen.
Dit onderzoek zal uw aandoening niet verbeteren, maar andere patiënten kunnen mogelijk in de toekomst voordeel halen uit de informatie die in dit onderzoek wordt verzameld. Hebt u na het lezen van deze informatie nog vragen? Dan kunt u terecht bij de onderzoeker. Op bladzijde 2 vindt u de contactgegevens.

Doel van het onderzoek
Het doel van dit onderzoek is om te onderzoeken of door met een elektronische neus te ruiken aan uitademingslucht op een minder belastende en snellere manier inflammatoire reumatische ziekten aangetoond of uitgesloten kan worden. Deze methode wordt vergeleken met de huidige onderzoeksmethodes.

Onderzoeksprocedure
Als u deelneemt aan dit wetenschappelijk onderzoek wordt bij u naast de geplande onderzoeken een ademanalyse test gedaan, die niet belastend is.
Op dezelfde dag dat bij u op de polikliniek komt, zal u gevraagd worden om maximaal 5 minuten op uw gewone tempo via de mond in een apparaat te ademen, een elektronische neus. De meting vindt plaats bij het afnamelaboratorium in het ziekenhuis. U hoeft dus geen extra bezoek te brengen aan het ziekenhuis.

Het opsporen van inflammatoire reumatische ziekten met een elektronische neus.

Voor- en nadelen van deelname
U heeft zelf geen voordeel van meedoen aan dit onderzoek.
Het onderzoek bestaat enkel uit het opvangen van uitgeademde lucht via een niet belastend apparaat, er worden geen medicijnen toegediend. Er zijn dan ook geen risico's verbonden aan deelname aan dit onderzoek.

Vertrouwelijkheid
Naast bovengenoemde metingen zullen wij mogelijk een aantal andere gegevens van u gebruiken: leeftijd, geslacht, of u rookt of niet, de eventuele aanwezigheid van andere ziekten die invloed kunnen hebben op uw adem, laboratoriumuitslagen, medicijngebruik en de diagnose. Uw persoonlijke gegevens worden enkel in gecodeerde vorm (om ervoor te zorgen dat uw identiteit vertrouwelijk blijft) opgeslagen. Volgens wettelijke bepalingen zullen uw gegevens 15 jaar bewaard worden.

Vrijwillige deelname
Uw deelname aan dit onderzoek is geheel vrijwillig. Als u besluit niet mee te doen, hoeft u verder niets te doen. Indien u wel mee wilt doen, kunt u op elk moment stoppen zonder dat dit gevolgen heeft voor uw verdere behandeling.

Het opsporen van inflammatoire reumatische ziekten met een elektronische neus.

Toestemming

Hierbij verklaar ik dat men mij uitgebreid uitleg gegeven heeft over dit onderzoek. Ik heb dit toestemmingsformulier gelezen en begrepen. Al mijn vragen zijn naar tevredenheid beantwoord. Ik stem vrijwillig in met deelname aan dit onderzoek en verzameling van relevante medische informatie. Ik weet dat ik te allen tijde mag stoppen met dit onderzoek en dat dit geen verdere gevolgen voor mij heeft. Ik ontvang een getekend en gedateerd exemplaar van dit toestemmingsformulier.

Ik wil meedoen aan dit onderzoek.

Naam proefpersoon:

Handtekening: Datum : __ / __ / __

Ik verklaar hierbij dat ik deze proefpersoon volledig heb geïnformeerd over het genoemde onderzoek.

Als er tijdens het onderzoek informatie bekend wordt die de toestemming van de proefpersoon zou kunnen beïnvloeden, dan breng ik hem/haar daarvan tijdig op de hoogte.

Naam onderzoeker (of diens vertegenwoordiger):

Handtekening: Datum: __ / __ / __

Het opsporen van inflammatoire reumatische ziekten met een elektronische neus.

Bijlage A: Contactgegevens voor Medisch Spectrum Twente

Onderzoekscentrum: Medisch Spectrum Twente

Koningsplein 1

7512 KZ Enschede

Hoofdonderzoekers:

dr. H.E. Vonkeman en mw. dr. M. Ghiti Moghadam.

Bereikbaar tijdens kantooruren op 053-4872450. Voor spoedgevallen buiten kantooruren kunt u contact opnemen met het algemene telefoonnummer 053-4872000 en vragen naar de dienstdoende reumatoloog.

Persoonsgegevens:

Voor meer informatie over de naleving van uw rechten bij de verwerking van uw persoonsgegevens kunt u contact opnemen met de Functionaris voor de Gegevensbescherming van MST, mw. P. van Paridon, telefoon: 06-31751387.

Medisch Spectrum ΔTwente	
Goedgekeurd	
Datum	20-01-20
Hand-tekening	