RELATIONSHIP BETWEEN CERTAINTY OF AXIAL SPONDYLOARTHRITIS AND TREATMENT RESPONSE

Master's thesis 2024 – Medicine – Aalborg University
Alexander Andrew Matthew Mills (20175466)
Amalie Birk Nielsen (20166224)
Emma Marie Mills (20165316)

Supervisor
Asta Linauskas, MD PhD, Department of Rheumatology, North Denmark Region Hospital, Hjørring
Department of Clinical Medicine, Aalborg University

Anne Grethe Jurik, MD Professor, Department of Radiology, Aarhus University Hospital

January 2024

ABSTRACT

Objective:

The objective of this study was to investigate, if there was a difference in treatment outcome, when comparing patients with 'certain' axSpA, who fulfil the ASAS classification criteria, and patients with clinical axSpA, who do not meet these criteria.

Methods:

A retrospective cohort study reviewing medical records, where clinical information was collected from electronic journal systems and a clinical online database, DANBIO. Additionally, MRI was reviewed by a specialist in radiology. Clinical and imaging information was collected to determine, if patients received a 'certain' or a clinical axSpA diagnosis. This study employed various statistical methods. Logistic regression analysis was used to evaluate the presence of a reduction in BASDAI ≥ 20, and examined the potential disparities between the two groups. A linear mixed-effects model was utilised to understand the disease burden for each course of treatment, and was calculated as a linear normalised time-weighted BASDAI average. A Kaplan Meier plot was computed for treatment adherence.

Results:

Out of the 129 patients included in this study, 71 were classified as having a 'certain' diagnosis and 58 were classified as having a clinical diagnosis. No statistically significant difference in treatment response was found, when comparing patients in the two groups. Further, it was found that patients with a 'certain' diagnosis adhere to treatment for a longer period of time compared to patients with a clinical diagnosis, though this finding was not significant ($p = 0.078$). Nevertheless, generally, treatment was associated with a decline in BASDAI scores for both groups, but overlapping confidence intervals indicate no statistically significant differences. However, treatment with bDMARDs resulted in a greater reduction in BASDAI score than treatment with NSAIDs, regardless of the axSpA diagnosis ($p < 0.05$).

Conclusion:

This study found no discernible difference in treatment response between patients with 'certain' and clinical axSpA diagnoses. While 'certain' diagnoses appeared to have longer treatment adherence and potentially more effectiveness, this finding also lacks statistical significance. Nevertheless, treatment decreased disease activity for both groups, questioning the relevance of strict adherence to classification criteria.

TABLE OF CONTENT

INTRODUCTION

Axial SpondyloArthritis (axSpA) is a chronic inflammatory disease, which can be classified into radiographic axSpA (r-axSpA) (also known as Ankylosing Spondyloarthritis, AS) and non-radiographic axSpA (nr-axSpA) (1-6). R-axSpA and nr-axSpA fulfil the modified New York criteria (mNY) (Figure 1) and the Assessment of SpondyloArthritis International Society (ASAS) criteria, respectively (Figure 2). However, some patients do not fulfil these classification criteria, and are treated based on a clinical diagnosis.

Modified New York Criteria for Ankylosing Spondylitis
Grading of the classification: • Definitive AS: ○ Fulfills radiological criteria and minimum 1 clinical criterium
Clinical criteria: • Backpain and stiffness for over 3 months, which improve by activity • Impaired range of motion in both lateral and forward bending • Impaired thorax expansion (> 5 cm)
Radiological criteria: • Sacroiliitis grade ≥ 2 bilaterally • Sacroiliitis grade ≥ 3 unilaterally

Figure 1: Modified New York Criteria for r-axSpA/AS

ASAS Classification Criteria for axSpA		
Chronic back pain (> 3 month) and age of onset < 45 years		
Sacroiliitis on X-ray and/or MRI **AND** ≥ 1 SpA characteristic	**OR**	HLA-B27 positive **AND** ≥ 2 SpA characteristic
SpA characteristics: • Inflammatory back pain • Arthritis • Enthesitis • Anterior uveitis • Dactylitis • Psoriasis • Inflammatory bowel disease • Good response to NSAIDs • Family disposition to SpA • HLA-B27 positive • Elevated CRP		

Figure 2: ASAS Classification Criteria for nr-axSpA

AxSpA mainly affects the axial skeleton, and is characterised by onset before age 45, gluteal and/or inflammatory back pain with a duration > 3 months and a good response to Nonsteroidal Anti-Inflammatory Drugs (NSAIDs). However, diagnosing and classifying axSpA can be challenging, as this type of gluteal pain is also commonly observed in other conditions, such as degenerative spinal disease, Modic type 1, HIZ-lesions, osteitis condensans ilii (OCI) and postpartum women (1). Furthermore, biochemical tests are limited to quantification of Human Leukocyte Antigen B27 (HLA-B27) and nonspecific acute phase proteins, such as C-reactive protein (CRP) (1-4,7).

The prevalence of HLA-B27 is approximately 9 % in Denmark, and the prevalence of axSpA among these people is 10-20 %, meaning this test increases the probability of SpA, but is not a definitive diagnostic parameter (8). CRP is frequently used to monitor disease activity, but an elevated value is inadequate to distinguish between, e.g. inflammation and infection. AxSpA is also known to be associated with other inflammatory

diseases, such as IBD, psoriasis, etc. Increased CRP can therefore equally be attributed to the rheumatological disease, as well as any associated diseases (9).

In addition to these biochemical results, Magnetic Resonance Imaging (MRI) can be used to increase the diagnostic certainty of axSpA. However, it is not considered a gold standard. Typical findings of axSpA, such as Bone Marrow Edema (BME), are also seen in other conditions, e.g. mechanical stress, degenerative diseases, pregnancy and OCI. Therefore, it is important to characterise both the location and extent of the BME. Likewise, it is important to evaluate the presence of accompanying structural changes such as erosions, fatty infiltration, fat metaplasia located in an erosion cavity (backfill), widening or narrowing joint space alterations, the presence of fluid in the joint space and ankylosis. These changes and their severity may increase the probability of an axSpA diagnosis (1,7,10,11).

The overlap in clinical symptoms and similarities in MRI findings among these various differential diagnoses, leads to difficulty in establishing a definitive axSpA diagnosis. Such ambiguities not only risk misdiagnosis, but also delay the initiation of relevant treatment. Particularly, the risk of overdiagnosis is concerning. Patients with the clinical manifestations of axSpA, and indefinite MRI findings might not necessarily invariably suffer from axSpA. As patients age, discerning age-related MRI changes from pathological ones becomes even more challenging (1).

Overdiagnosis risks unnecessarily exposing misdiagnosed patients, to the side effects of the prescribed treatment, delaying the diagnosis of the correct disease, and using financial resources on expensive treatment, for which there is no indication. Additionally, an incorrect diagnosis might lead to treatment failure; this raises a pertinent question: Do patients with ‘certain’ diagnoses have better treatment outcomes compared to those with clinical diagnoses, potentially attributable to the initial administration of the correct treatment?

Therefore, due to the limited evidence available in the current literature, this study investigated, if there is a difference in treatment outcome, when comparing patients with axSpA, who fulfil the ASAS classification criteria, and patients with clinical axSpA, who do not meet these criteria.

METHODS

STUDY DESIGN AND SETTING

This retrospective cohort study included patients treated for axSpA at the Department og Rheumatology, North Denmark Regional Hospital, Hjørring, Denmark.

STUDY POPULATION

To ensure the validity and relevance of this study, patients were selected based on several criteria. The study population consisted of patients, who had an active treatment course with either NSAIDs or Disease-Modifying Anti-Rheumatic Drugs (DMARDS) at the Rheumatology Department of The North Denmark Regional Hospital, at the commencement of this project in August 2023. The following International Classification of Diseases 10th revision diagnosis codes were included: M45: Ankylosing spondylitis, M46.1: Sacroiliitis, not elsewhere classified, M46.8: Other specified inflammatory spondylopathies and M46.9: Unspecified inflammatory spondylopathy. Patients were required to have a baseline Bath Ankylosing Spondylitis Disease (BASDAI) score for at least one treatment course, and at least three appointments to evaluate the treatment response.

DATA ENTRY AND STORAGE

DANBIO

DANBIO (12) is an online database established in the year 2000 that constitutes a nationwide patient management system, and clinical quality index for all adult rheumatologic patients with inflammatory musculoskeletal diseases, such as axSpA. The database relies on data reported by patients, and healthcare professionals across various rheumatology departments, and private clinical facilities in Denmark. Patients and physicians are encouraged to enter data during control visits, and when there is an alteration in medication. In this study, DANBIO was utilised to create the initial list of included patients, and to gather the following information: Smoking status, CRP, BASDAI score (0-100), as well as the type, dosage, duration and reason for alteration of NSAIDs treatment and/or DMARDs.

REVIEW OF MEDICAL RECORDS

Electronic medical records were reviewed. Three journal systems (NordEPJ, The National Health Journal and Archive Clinical Suite) were used to collect both baseline data, and data regarding treatment and treatment response. The following information was collected when possible from the medical records: Age, Body Mass Index (BMI), location of back pain (cervical, thoracal, lumbar and/or hip region), onset of back pain, if they had an occupation involving physically demanding work, as defined in "Vejledning om erhvervssygdomme" (13). Furthermore, it was noted, if the patients had a familial predisposition to axSpA, exhibited HLA-B27, or had an elevated CRP. It was also documented, whether associated diseases, such as peripheral arthritis, enthesitis, anterior uveitis, dactylitis, psoriasis in the skin and/or nails and inflammatory bowel disease were present. Regarding evaluations of NSAIDs and DMARDs, the type, dosage, duration and rationale for modifications, were cross-referenced with DANBIO and documented.

REVIEW OF IMAGING EXAMINATION

A radiology specialist reviewed MRI sequences performed within a time frame of 6 months before and after the time of diagnosis, blinded from the clinical information. The MRI identified BME in the subcondral bone on fluid-sensitive scans, including Short Tau Inversion Recovery (STIR) and T2 fat-saturation (FS) sequences of the SIJ and spine. Furthermore, it was noted, if axSpA was the most likely diagnosis.

REDCAP

Data entry and management were carried out employing the Research Electronic Data Capture (REDCap) tool, hosted at Aalborg University Hospital, Denmark. REDCap is a secure, non-commercial online software platform, facilitating data collection for research studies, by providing an interface for data input, data sharing, transaction logging, blinding, etc., and automated export for data transfers to commonly used statistical software (14-17).

EXPOSURE, OUTCOME AND COVARIATES

The exposure variable utilised in this study, was defined as fulfilment of the imaging arm of the ASAS criteria, as the clinical arm is not utilised in Denmark. To assess if patients met the ASAS criteria, it was evaluated if both the MRI results were indicative of axSpA, and if the clinical aspects of the criteria were met. Consequently, the patients were separated into two groups: 1) patients meeting the ASAS classification criteria, prospectively referred to as 'certain' axSpA, and 2) patients who did not fulfil the criteria, henceforth referred to as clinical axSpA.

This study had two primary outcome variables: a binary evaluation of a good treatment response, and a linear normalised time-weighted BASDAI average score. Attainment of a good treatment response was defined as a reduction in BASDAI score ≥ 20, at any point within the first 2 to 52 weeks, after initiation of a new treatment.

The linear outcome variable was calculated for each treatment course, by multiplying the BASDAI score for each appointment by a time-weighted factor. The time-weighted factor was calculated by adding half of the days since the previous appointment, to half of the days until the next appointment. Individual results were afterwards summed across all appointments for each treatment course (Figure 3). Lastly, the summarised value was divided by the individual treatment courses' total duration (in days). This process generated a normalised time-weighted value maintaining the same 0-100 range, as the original BASDAI score.

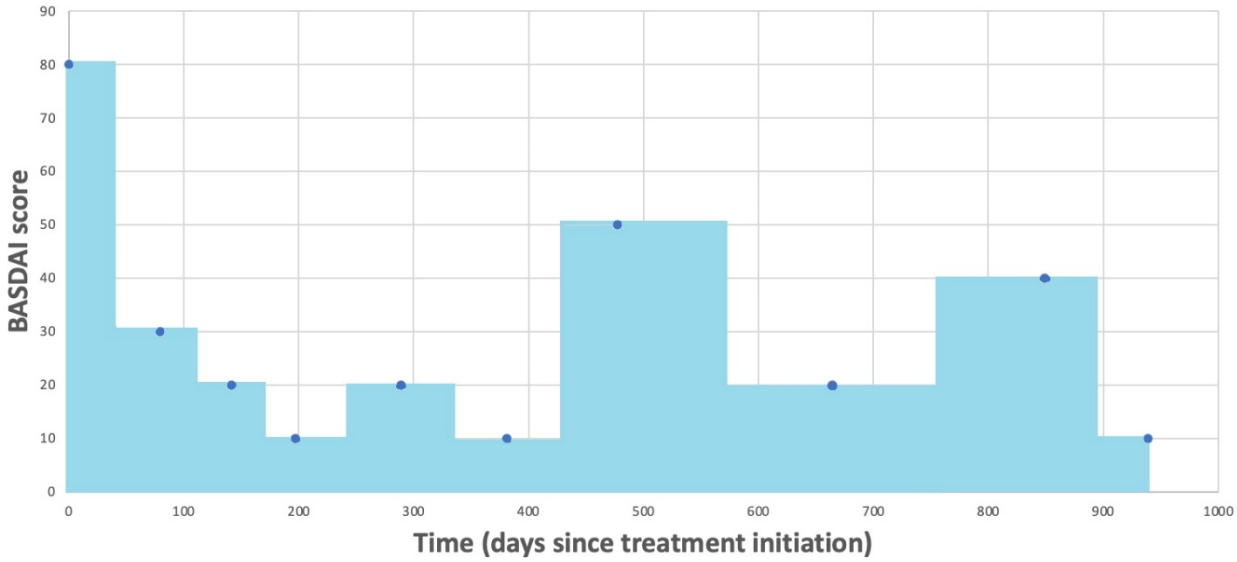

Figure 3: Illustration of how the time-weighted BASDAI score for each appointment was calculated for a fictive treatment course. Each dot represents an appointment. Changes in the area under the curve mark the halfway point between two appointments. The total area under the curve corresponds to each treatment course's summarised time-weighted average BASDAI score. This value is then divided by the entire treatment duration (in days) to normalise the value to a 0-100 range.

Several confounders and a single covariate were included in the statistical analyses. Confounders consisted of age at diagnosis, presence of HLA-B27, gender and smoking status.

Additionally, the covariate treatment type was analysed, which was divided into three categories: NSAIDs, bDMARDs and Sulfasalazine. The bDMARDs category contained the following pharmaceuticals: Adalimumab, Infliximab, Certolizumb, Golimumab, Etanercept, Secukinumab, Ixekizumab, Ustekinumab and Tofacitinib.

Only covariates with no missing data were included in the regression analyses to avoid having to either imputate missing values, and to maintain as large of a dataset as possible. This resulted in several potential confounders, such as BMI, not being included in these analyses.

DEFINEMENT OF TREATMENT RESPONSE

A treatment course was defined based on three criteria: 1) Treatment was initiated when a new prescription started, or a new dosage of the same medication was prescribed. 2) Concerning compliance, patients were required to have adhered to the prescribed treatment until the subsequent appointment. 3) In regard to termination of treatment, it was defined as either the doctor instructing discontinuation, altering dosage or medication, if the patient stopped the treatment, or if it was the final recorded consultation for the patient. In cases where more than one medication was prescribed simultaneously, only the medication marked as the primary axSpA treatment, by the medical record reviewers, was considered. These defined criteria resulted in the identification of 368 treatment courses for analysis.

STATISTICAL ANALYSIS

The baseline demographic and clinical characteristics were documented using descriptive statistics. Confounders and covariates were compared using the Kruskal-Wallis test for continuous variables and chi-squared for categorical variables.

A logistic regression analysis was conducted to evaluate the likelihood of achieving a good treatment response, and log odds were estimated. This model allowed comparison between the two exposure groups, while adjusting for covariates. For the dependent variable, the binary evaluation of a good treatment response outcome variable was utilised. The independent variables consisted of the exposure variable, treatment type, days since treatment start, age at diagnosis, smoking status, presence of HLA-B27, and gender.

Due to the study being composed of longitudinal data, a linear mixed-effects model was employed, to consider both individual variations among patients and changes in outcome over time for each patient. The fixed effects included in the model were the exposure variable, treatment type, smoking status, age at diagnosis, presence of HLA-B27 and gender.

A Kaplan Meier plot was used to visually represent, how the probability of patients adhering to their treatment course changed over time, stratified by the exposure variable. Lastly, a graph visualising the evolution of BASDAI score over time was created.

All statistical analyses were carried out using R version 4.3.1 (18).

ETHICS

This study received authorisation from The Scientific Ethics Committee for the Northern Region of Denmark (case number: 2309175).

RESULTS

STUDY POPULATION AND DEMOGRAPHICS

A total of 129 patients were included in this study. In Table 1, demographic characteristics are shown.

In this study population, there was a notable gender disparity among the patients, as there were markedly more males, who had a 'certain' axSpA diagnosis than females. The gender difference was found to be statistically significant. The median age at diagnosis for patients with a 'certain' axSpA was 31 years, with an interquartile range of 24 to 38 years. Patients with clinical axSpA were generally older, with a median age of 48 years, and an interquartile range of 37.2 to 55.5. This age difference was found to be statistically significant. Regarding HLA-B27 expression, patients with a 'certain' axSpA diagnosis exhibited a significantly higher prevalence at 87.3 %, compared to 53.4 % for patients with a clinical diagnosis. Inflammatory bowel disease and peripheral arthritis had a significantly higher prevalence among clinical axSpA patients, compared to the 'certain' group. Conversely, anterior uveitis was more prevalent among patients with 'certain' axSpA; however, this result was not statistically significant.

Variable	Level	'Certain' axSpA, n=71 (n)	Clinical axSpA, n=58 (n)	Total, n=129 (n)	P-value
ASAS MRI	Yes	100 % (71)	41.4 % (24)	73.6 % (95)	
	No/Clinical	0 % (0)	58.6 % (34)	26.4 % (34)	< 0.001
ASAS clinical	Yes	100 % (71)	36.2 % (21)	71.3 % (92)	
	No	0 % (0)	63.8 % (37)	28.7 % (37)	< 0.001
Gender	Female	26.8 % (19)	51.7 % (30)	38 % (49)	
	Male	73.2 % (52)	48.3 % (28)	62 % (80)	0.006
Age at diagnosis	Median [IQR]	31 [24, 38]	48 [37.2, 55.5]	37 [29, 47]	< 0.001
HLA-B27	Positive	87.3 % (62)	53.4 % (31)	72.1 % (93)	
	Negative	12.7 % (9)	46.6 % (27)	27.9 % (36)	< 0.001
Disposition to axSpA	Yes	24.6 % (15)	14.9 % (7)	20.4 % (22)	
	No	75.4 % (46)	85.1 % (40)	79.6 % (86)	0.318
	Missing	10	11	21	
Back Straining work	Yes	15.1 % (8)	25.6 % (11)	19.8 % (19)	
	No	84.9 % (45)	74.4 % (32)	80.2 % (77)	0.305
	Missing	18	15	33	
BMI	Median [IQR]	26.2 [22.1, 28.6]	25.7 [23.6, 32.2]	26 [22.9, 30]	0.639
	Missing	22	12	34	
Smoking status	Never	53.5 % (38)	55.2 % (32)	54.3 % (70)	
	Current	32.4 % (23)	25.9 % (15)	29.5 % (38)	
	Former	14.1 % (10)	19 % (11)	16.3 % (21)	0.623
Peripheral arthritis	Yes	23.9 % (11)	51.5 % (17)	35.4 % (28)	
	No	76.1 % (35)	48.5 % (16)	64.6 % (51)	0.022
	Missing	25	25	50	
Enthesitis	Yes	22.9 % (8)	19 % (4)	21.4 % (12)	
	No	77.1 % (27)	81 % (17)	78.6 % (44)	1
	Missing	36	37	73	
Anterior uveitis	Yes	25 % (16)	9.3 % (4)	18.7 % (20)	
	No	75 % (48)	90.7 % (39)	81.3 % (87)	0.074
	Missing	7	15	22	
Dactylitis	Yes	9.1 % (3)	11.5 % (3)	10.2 % (6)	
	No	90.9 % (30)	88.5% (23)	89.8 % (53)	1
	Missing	38	32	70	
Psoriasis	Yes	17.3 % (9)	15.6 % (7)	16.5 % (16)	
	No	82.7 % (43)	84.4 % (38)	83.5 % (81)	1
	missing	19	13	32	
Inflammatory bowel disease	Yes	5.9 % (2)	30.3 % (10)	17.9 % (12)	
	No	94.1 % (32)	69.7 % (23)	82.1 % (55)	0.022
	missing	37	25	62	
Reactive arthritis	Yes	30 % (6)	50 % (8)	38.9 % (14)	
	No	70 % (14)	50 % (8)	61.1 % (22)	0.379
	Missing	51	42	93	

Table 1: Baseline demographic characteristics and ASAS classification status for the study population. IQR: interquartile range. n: number of patients.

LOGISTIC REGRESSION ANALYSIS

A logistic regression analysis was conducted to determine, whether there was a disparity in good treatment responses between patients with 'certain' axSpA and those with clinical axSpA. The results revealed no discernible difference in treatment response between these two groups. However, the logistic regression analysis unveiled two statistically significant results: First, treatment with bDMARDs, in contrast to NSAIDs, was associated with a greater likelihood of achieving a favourable treatment outcome. Second, being male, as opposed to female, was associated with a greater likelihood of not achieving a good response. Notably, other covariates examined, such as days since treatment start for each treatment course, HLA-B27 status, age at diagnosis and smoking status, all exhibited no statistically significant association with attainment of a good treatment response.

In order to gauge the model's explanatory power, a Nagelkerke pseudo-R^2-value was calculated (19). This value was calculated to assess the extent to which the included factors contributed to explaining the variance in treatment outcome. The calculation showed a pseudo-R^2 of 0.068, indicating that numerous factors, beyond the covariates incorporated in the model, play a role in explaining the treatment response observed in this model.

Variable	Log odds [95 % CI]	Standard error	Z-value	P-value
(Intercept)	-0.332 [-1.702, 1.024]	0.692	-0.479	0.632
Fulfilling ASAS vs NOT fulfilling ASAS	0.117 [-0.667, 0.888]	0.395	0.297	0.766
Treatment with bDMARDs vs NSAIDs	0.814 [0.229, 1.431]	0.305	2.666	0.008
Treatment with Sulfasalazine vs NSAIDs	0.288 [-1.690, 1.839]	0.858	0.336	0.737
Days since treatment start for each treatment	-0.002 [-0.007, 0.002]	0.002	-1.042	0.297
Age at diagnosis	-0.020 [-0.049, 0.009]	0.015	-1.345	0.179
Current smoker vs never smoker	0.118 [-0.545, 0.761]	0.332	0.356	0.722
Former smoker vs never smoker	0.176 [-0.657, 0.961]	0.409	0.429	0.668
HLA-B27 positive vs negative	0.078 [-0.608, 0.796]	0.356	0.219	0.827
Male vs female gender	-0.666 [-1.251, -0.086]	0.296	-2.248	0.025

Table 2: The results of the logistic regression analysis. Log odds, standard errors, Z-values, p-values and upper and lower limits for 95 % confidence intervals are provided. Significant p-values are highlighted. CI: 95 % confidence intervals

LINEAR MIXED-EFFECTS MODEL

In an effort to gain a better understanding of the disease burden over an entire treatment period, a linear mixed-effects model was performed. This analysis found that patients with 'certain' axSpA, exhibited a normalised time-weighted average of 2 BASDAI scores higher than those with clinical axSpA, although this difference lacks statistical significance. Additionally, only one significant fixed covariate was identified. Treatment with bDMARDs was associated with a 10 lower average BASDAI score, compared to NSAID treatment. Remarkably, for gender, a 6.4 lower BASDAI score was found for males as compared to females, contrary to what the logistic regression analysis found, however this result was not significant. Finally, for the fixed covariates alone, the pseudo-R^2-value was calculated to be 0.091, while the total pseudo-R^2-value for fixed and random covariates was 0.753.

Effect	Variable	Estimate [95 % CI]	Standard error	T-value	df	P-value
Fixed	(Intercept)	48.488 [32.043, 64.932]	8.299	5.843	111.279	<0.001
Fixed	Fulfilling ASAS vs NOT fulfilling ASAS	2.115 [-8.411, 12.640]	5.309	0.398	106.037	0.691
Fixed	Treatment with bDMARDs vs NSAIDs	-10.236 [-13.892, -6.579]	1.858	-5.509	290.195	<0.001
Fixed	Treatment with Sulfasalazine vs NSAIDs	8.687 [-2.200, 19.574]	5.533	1.570	307.882	0.117
Fixed	Age at diagnosis	-0.043 [-0.444, 0.357]	0.202	-0.214	107.096	0.831
Fixed	Current smoker vs never smoker	3.033 [-6.042, 12.107]	4.579	0.662	109.891	0.509
Fixed	Former smoker vs never smoker	-6.230 [-17.657, 5.197]	5.767	-1.080	110.830	0.282
Fixed	HLA-B27 positive vs negative	-4.347 [-14.030, 5.337]	4.886	-0.890	108.709	0.376
Fixed	Male vs female gender	-6.455 [-14.882, 1.971]	4.250	-1.519	105.766	0.132
Random	Patient ID Intercept standard deviation	19.476				
Random	Residual standard deviation	11.891				

Table 3: The results of the linear mixed-effects model. Estimates, standard errors, t-values, df (degrees of freedom), p-values and upper and lower limits for the 95 % confidence intervals are provided. Significant p-values are highlighted. Patient ID intercept standard deviation: The average deviation at baseline between individual patients' BASDAI scores. Residual standard deviation: The variance in the evolution of the BASDAI score over time within each patient. CI: 95 % confidence interval.

ADHERENCE TO TREATMENT

A Kaplan Meier Plot (Figure 4) provided a visual representation, of how the probability, of adherence to treatment, changed over time dependent on the certainty of the axSpA diagnosis. It appears the patients with a 'certain' axSpA diagnosis remain on the same treatment for a longer duration, or at least in the initial 1000 days, compared to those with a clinical diagnosis. However, the observation achieved no statistical significance.

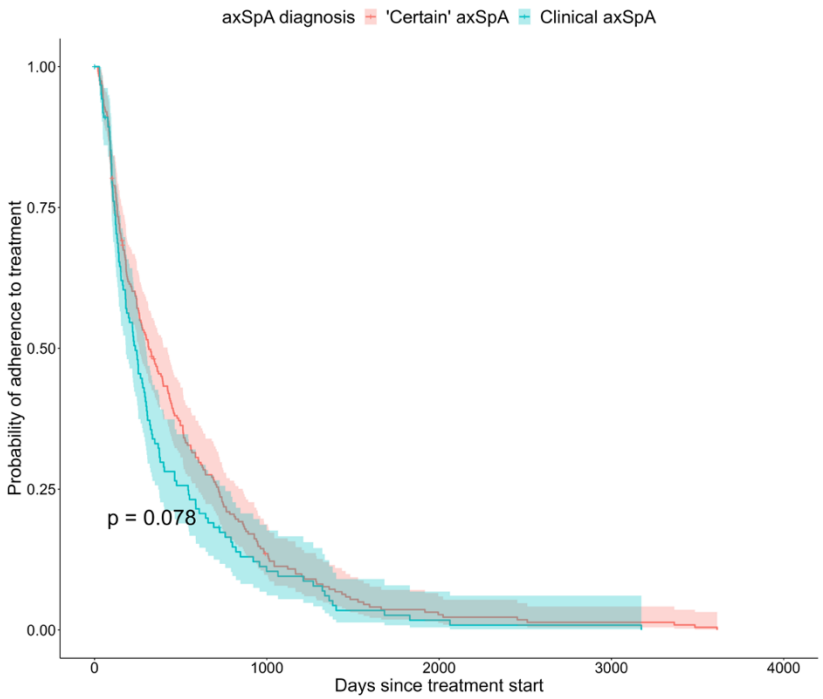

Figure 4: Kaplan Meier Plot. The likelihood of adherence to treatment (NSAIDs, bDMARDs or Sulfasalazine) in days since initiation of treatment. Cox regression was utilised to calculate the p-value.

CHANGE IN BASDAI SCORE OVER TIME

Figure 5 visualises BASDAI scores over the first five years for all treatment courses, subdivided by the certainty of axSpA diagnosis. This showed that treatment generally resulted in a decrease in BASDAI scores for both groups. Further, patients with a 'certain' axSpA diagnosis had a consistently slightly lower score than those with a clinical axSpA diagnosis, and the groups followed similar curves. However, the confidence intervals overlapped throughout the entirety of the graph, indicating no statistically significant differences.

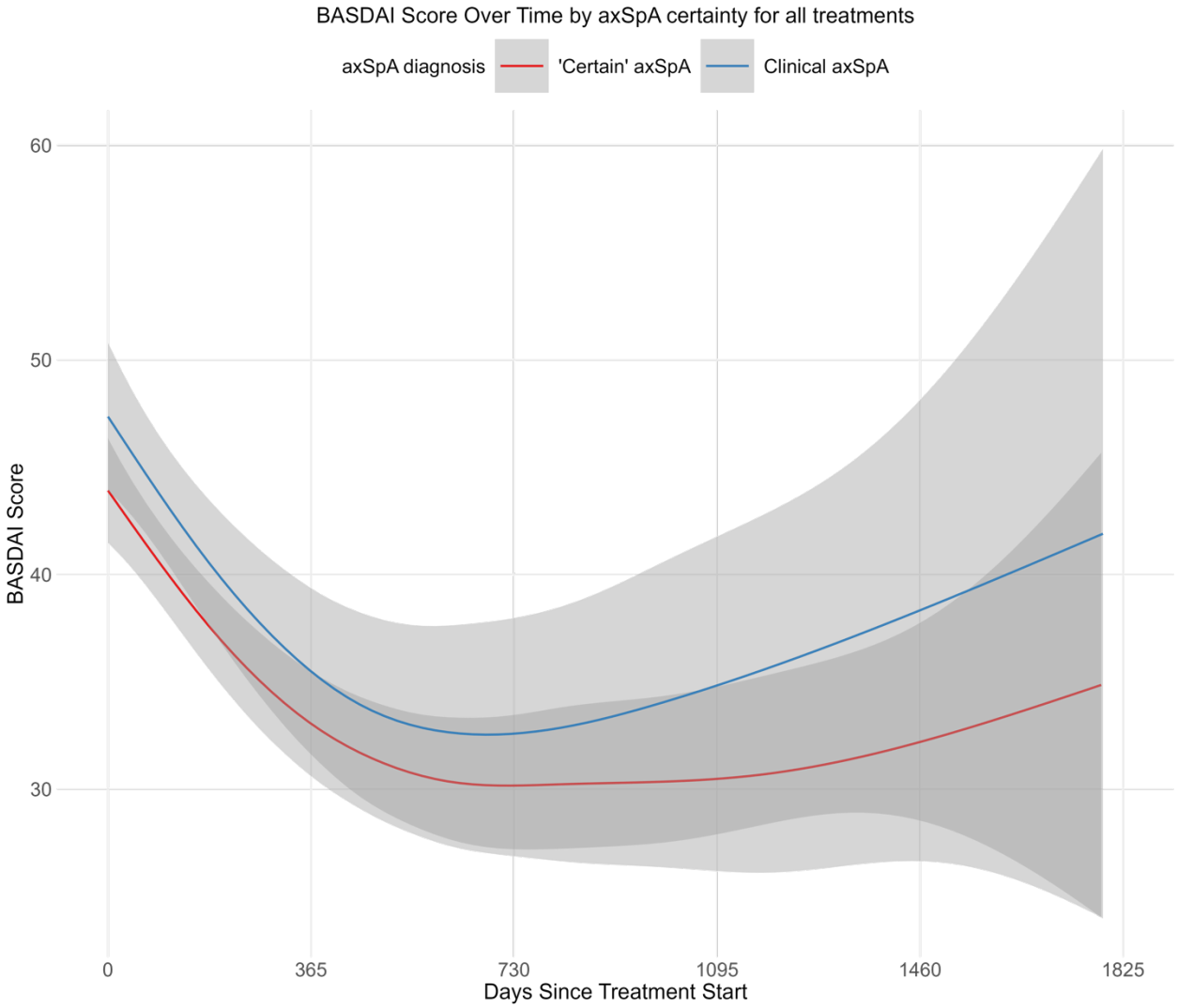

Figure 5: BASDAI scores over the first five years for all types of treatment divided by axSpA certainty. The light grey areas indicate 95 % confidence intervals, whereas the dark grey indicates overlap in confidence intervals.

DISCUSSION

The main objective of this study was to identify, if there existed a disparity in treatment response, when comparing patients, who fulfil the ASAS criteria ('certain' axSpA), with patients who did not meet these criteria (clinical axSpA). The results of this study showed that treatment generally resulted in a decrease in BASDAI scores for both groups, but no significant difference in treatment outcome was observed between them. However, it should be noted that patients with a 'certain' axSpA diagnosis, appeared to adhere to their treatment for a longer duration. This could imply that they had a better treatment response compared to patients, who received the clinical diagnosis, nevertheless, this result was not significant.

DIFFERENTIAL DIAGNOSES

One plausible explanation for this lack of difference could be that a larger than expected proportion of patients, including those with clinical diagnoses, derive therapeutic benefit from treatment. The primary approach for many differential diagnoses of axSpA, such as OCI, degenerative diseases, nonspecific lower back pain, etc., involves the use of NSAIDs (20,21). This commonality in initial treatment, makes identifying differences in treatment outcomes challenging. Additionally, it is well-documented that OCI is a self-limiting condition and therefore, symptoms are anticipated to resolve eventually (20). These factors could account for some of the good treatment responses observed among patients with clinical diagnoses, due to the nature of these differential diagnoses, and not the treatment itself. Furthermore, NSAIDs are known to effectively treat any kind of inflammatory pain (22). Consequently, nonspecific lower back pain might benefit from treatment, resulting in several patients having a good treatment response, independent of the certainty of axSpA diagnosis.

BASDAI SCORE, ASAS20 AND ASAS40

Challenges in identifying improvement in axSpA could be resolved, if there existed an objective and more exact measurement of axSpA disease activity. This study utilised the BASDAI score to evaluate the disease activity, which is an exclusively subjective measurement consisting of patient-reported pain, morning stiffness, fatigue, sleep and overall functional ability (23). One potential bias towards increased values of BASDAI scores, and thereby not achieving a good treatment response, could be negativity bias. If a significant amount of time passed since the last reporting of BASDAI score, patients may find it easier to recall negative experiences associated with bad disease days, as negative experiences often tend to have a stronger impact on emotions and memory (24).

In other studies, the Assessment in SpondyloArthritis International Society 20 % or 40 % (ASAS20 or ASAS40) improvement criteria have been employed, as the primary endpoint in clinical trials, for evaluating disease activity in patients with axSpA (25). To meet the ASAS20 and ASAS40 improvement criteria, patients must achieve at least a 20 % or 40 % improvement in a minimum of three out of four domains: pain, patient's global assessment of disease activity, Bath Ankylosing Spondylitis Functional Index (BASFI), and biomarkers, such as CRP. Regarding ASAS20 and ASAS40 improvement criteria, it is noteworthy that the focus is on percentage and a relative reduction in symptoms and disease activity. Therefore, ASAS20 and ASAS40 improvement criteria are less restrictive and

can capture a broader range of treatment improvements. For further research, this method for observing disease activity could be applied to compare differences in treatment outcomes for 'certain' and clinical axSpA diagnoses.

BDMARDS COMPARED TO NSAIDS

This study's findings revealed only one consistently significant result: patients being treated with bDMARDs, instead of NSAIDs, independent of their fulfilment of the ASAS criteria, achieved significantly better treatment outcomes. bDMARDs are recognised for their ability to target specific proteins involved in inflammation. Consequently, bDMARDs possess properties that potentially alter the course of axSpA, unlike NSAIDs, which primarily provide symptomatic relief by reducing pain and inflammation (26). Similar observations have been seen in prior studies, where the results showed a more prominent reduction in inflammation on MRI, and a better clinical remission, and thereby a better treatment outcome, when the patients received both NSAIDs and bDMARDs, compared to NSAIDs alone (27).

Another pertinent explanation for the better treatment outcomes associated with bDMARDs is patient selection. Patients prescribed bDMARDs are typically selected due to the severity of their condition, potentially predisposing them to a better treatment response, as there is more room for improvement. The current Danish national axSpA guidelines only advocate for considering treatment with bDMARDs for patients who have tried at least two separate NSAIDs, over a four-week period, and continued to have a BASDAI score consistently above 40 (1). This creates a natural selection bias, which could significantly impact the logistic regression analysis, as the outcome is defined as an absolute BASDAI score reduction of at least 20, rather than a relative decrease in scores. Consequently, patients starting with higher BASDAI scores might find it easier to reach this fixed threshold, making the 'good response' criterion more attainable than for patients with lower initial scores.

In contrast, the results of the time-weighted linear mixed model demonstrate, bDMARD treatment results in a lower average BASDAI score than NSAIDs. This contradicts the previous observation, as this model requires bDMARD treatment to maintain consistently lower scores to achieve a lower average, indicating bDMARDs consistently sustain better treatment outcomes, compared to NSAIDs.

An alternative explanation for the better treatment outcomes observed for bDMARDs could be that NSAIDs appear less impactful due to attrition bias. Around two-thirds of patients, who receive NSAIDs experience difficulties in tolerating the maximum NSAID doses (28). Consequently, having to potentially terminate treatment, before a sufficient improvement in disease activity, can be documented.

OTHER FACTORS INFLUENCING THE RESULTS

A large disparity was observed between the pseudo-R^2-values in the logistic regression, and linear mixed effects models. The pseudo-R^2-values for the logistic model, and the pseudo-R^2-value for the fixed portion of the linear mixed effects model were below 10 %, indicating a model with a low explanatory power. However, when accounting for the fixed and random variables, the pseudo-R^2-value for the linear mixed effects model was significantly higher at 75.3 %, indicating a robust model. This disparity indicates

that the fixed effects included in the models only play a minor role, while the individual variation between patient identity, accounts for a much larger portion of the variance.

This emphasises the need for clinicians to consider the complexity of patient-specific factors, beyond the fixed effects. These could include genetic predisposition, the severity of disease at diagnosis, treatment length, concurrent treatment with other medications, potential infections, and the presence of autoimmune diseases, etc.

CLINICAL RELEVANCE

Given this study identified no discernible difference in treatment response between those with 'certain' axSpA, and those with clinical axSpA, it raises a pertinent clinical question: Should treatment be limited to patients with a formal diagnosis, when it appears to be effective regardless of fulfilment of the ASAS criteria?

Previous research (1) has raised concerns about overdiagnosis among the elderly, as they are more susceptible to other differential diagnoses, in particular degenerative diseases. In fact, the prevalence of degenerative diseases, significantly outnumbers that of axSpA, amplifying the potential for misinterpreting age-related changes in MRI of the SIJ (1). However, these potential challenges with classifying patients might not be relevant, if the classification criteria have no impact on treatment outcomes. Focusing on symptomatology and treatment response might be a more relevant clinical consideration, than strict adherence to age-based diagnostic criteria.

Furthermore, it is known that the imaging arm of the ASAS classification criteria utilised in Denmark prioritises specificity (97.3 %) over sensitivity (66.2 %) (29). This naturally leads to a large proportion of patients, who do have axSpA not fulfilling these criteria. Utilising these criteria as a tool for clinical decision-making could, therefore, risk underdiagnosing and potentially denying treatment to patients, who could benefit from it. Underdiagnosis is of particular concern among younger patients, as shorter disease duration is often linked to a more favourable treatment outcome. However, it is important to acknowledge that only a limited number of studies have directly evaluated the impact of treat-to-target strategy (28).

Another topic of clinical importance is that there are currently no validated guidelines for choosing biological treatment. Nevertheless, bDMARDs are high-cost therapies, necessitating a carefully individualised approach for each patient. Therefore, it is advisable to consider other factors, such as the presence of extra-articular manifestations and comorbidities, when adjusting treatment. Cost implications should only be considered, if the treatment outcome is comparable (28). Furthermore, bDMARDs pose a significant risk of severe infections, occasional occurrence of bone marrow suppression, hepatotoxicity and potential exacerbation of congestive heart failure, etc. (30). These factors require clinicians to carefully weigh the potential benefits against the costs and complications for each individual patient and treatment.

STRENGTHS AND LIMITATIONS

This study encountered both strengths and limitations, some of which are already described. A possible limitation is that three individuals reviewed patient records, and could have interpreted journal entries differently, potentially leading to inter-observer variability and biased results. However, before data entry, a few patient records were

reviewed jointly, and a consensus of interpretation was reached. If doubt occurred during data entry, patient records were reviewed jointly.

A strength of this study was that MRIs were assessed by a highly experienced radiologist with up-to-date expertise in MRI interpretation. Nevertheless, only one radiologist assessed MRIs, introducing potential single-observer bias. Two blinded radiologists reviewing MRIs, could be employed for better reliability in future studies.

The external validity of this study was limited by the sample consisting of 129 patients treated at a single centre. However, each patient had multiple treatment courses, which consisted of numerous appointments, resulting in many data points for each patient.

A key strength of the linear mixed-effects model was the use of a normalised time-weighted average BASDAI score as the outcome variable. This provided a better representation of the disease burden over an entire treatment course, as opposed to a momentary insight. However, this could lead to outliers in BASDAI scores having less impact on the overall results.

CONCLUSION

The primary purpose of this study was to identify, if there was a difference in treatment response, when comparing patients who fulfilled the ASAS classification criteria, with patients who did not. The results revealed no significant difference in treatment response between these two groups. Therefore, despite clinical diagnoses potentially being associated with differential diagnoses in terms of treatment outcomes, this study's findings suggest that these diagnostic complexities have no discernible consequences. Patients with ‘certain’ diagnoses did, however, seem to adhere to treatment for a longer period of time, compared to patients with a clinical diagnosis. While this trend may suggest a potential improved treatment effect, no significance was observed. Nevertheless, treatment was associated with a decline in BASDAI scores for both exposure groups. This underscores the importance of clinical decision-making based on individual disease courses and treatment responses, and brings into question the relevancy of strict adherence to classification criteria.

BIBLIOGRAPHY

(1) Hendricks O, Horn HC, Juhl Pedersen S, Pelck R, Juul Sørensen I, Jurik AG, et al. NBV - aksial spondyloartritis. 2021; .

(2) Puhakka KB, Jurik AG, Schiottz-Christensen B, Hansen GVO, Egund N, Christiansen JV, et al. Magnetic resonance imaging of sacroiliitis in early seronegative spondylarthropathy. Abnormalities correlated to clinical and laboratory findings. Rheumatology (UK) 2004;43(2):234-237.

(3) Braun J, Sieper J. Classification, Diagnosis, and Referral of Patients with Axial Spondyloarthritis. Rheum Dis Clin North Am 2012;38(3):477-485.

(4) Hayward RJ, Machado PM. Classification Criteria in Axial Spondyloarthritis: What Have We Learned; Where Are We Going? Rheum Dis Clin North Am 2020;46(2):259-274.

(5) Rudwaleit M, Van Der Heijde D, Khan MA, Braun J, Sieper J. How to diagnose axial spondyloarthritis early. Annals of the Rheumatic Diseases 2004 -05-01;63(5):535.

(6) Rudwaleit M, Landewe R, Van Der Heijde D, Listing J, Brandt J, Braun J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Annals of the Rheumatic Diseases 2009 -03-17;68(6):770.

(7) Maksymowych WP. The role of imaging in the diagnosis and management of axial spondyloarthritis. Nat Rev Rheumatol 2019;15(11):657-672.

(8) Laboratoriemedicinsk vejledning RS. HLA-B-gen (B27);DNA, KIA, Laboratoriemedicnsk vejledning. 2019; Available at: http://lmv.regionsjaelland.dk/dokument.asp?DokID=427548. Accessed 14-11-, 2023.

(9) Pepys MB, Hirschfield GM. C-reactive protein: a critical update. J Clin Invest 2003 -06-15;111(12):1805.

(10) Jurik AG. Diagnostics of Sacroiliac Joint Differentials to Axial Spondyloarthritis Changes by Magnetic Resonance Imaging. J Clin Med 2023;12(3).

(11) Poddubnyy D, Weineck H, Diekhoff T, Redeker I, Gobejishvili N, Llop M, et al. Clinical and imaging characteristics of osteitis condensans ilii as compared with axial spondyloarthritis. Rheumatology 2020;59(12):3798-3806.

(12) Hetland ML. DANBIO--powerful research database and electronic patient record. Rheumatology (Oxford, England) 2011 Jan;50(1):69-77.

(13) Arbejdstilsynet. Vejledning i erhvervssygdomme. 2023; .

(14) Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics 2009 Apr 01,;42(2):377-381.

(15) Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. Journal of biomedical informatics 2019 Jul 01,;95:103208.

(16) Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics 2009 Apr 1,;42(2):377-381.

(17) Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. Journal of biomedical informatics 2019 Jul 1,;95:103208.

(18) R Core Team. R: A Language and Environment for Statistical Computing. 2023;4.3.1.

(19) Nagelkerke NJD. A note on a general definition of the coefficient of determination. Biometrika 1991 Sep 1,;78(3):691-692.

(20) Peter M. Williams1, Doug W. Byerly2. Osteitis Condensans Ilii.

(21) Famili P, N A, N S, L F. Osteoarthritis: Diagnosis and Treatment. Journal of Dental Health, Oral Disorders & Therapy 2018 Jun 13,;9(3).

(22) Peter M. Williams, Doug W. Byerly. Nonsteroidal Anti-Inflammatory Drugs (NSAIDs).

(23) Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: The bath ankylosing spondylitis disease activity index. J Rheumatol 1994;21(12):2286-2291.

(24) Norris CJ. The negativity bias, revisited: Evidence from neuroscience measures and an individual differences approach. Social neuroscience 2021 Feb;16(1):68-82.

(25) Landewé R, van Tubergen A. Clinical Tools to Assess and Monitor Spondyloarthritis. Curr Rheumatol Rep 2015 Jul 1,;17(7):47.

(26) Brandt J, Listing J, Sieper J, Rudwaleit M, Van Der Heijde D, Braun J. Development and preselection of criteria for short term improvement after anti-TNF treatment in ankylosing spondylitis. Annals of the Rheumatic Diseases 2004;63(11):1438.

(27) Zong H, Xu S, Wang J, Chu Y, Chen K, Wang C, et al. Presence of subclinical inflammation in axial spondyloarthritis patients with NSAID/anti-TNF-α drug-induced clinical remission. Clin Rheumatol 2022 May;41(5):1403-1412.

(28) Fragoulis GE, Siebert S. Treatment strategies in axial spondyloarthritis: what, when and how? Rheumatology (Oxford, England) 2020 Oct 01,;59(Suppl4):iv79-iv89.

(29) Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Annals of the rheumatic diseases 2009 Jun 1,;68(6):777-783.

(30) Onecia Benjamin, Amandeep Goyal, Sarah L. Lappin. Disease-Modifying Antirheumatic Drugs (DMARD).