U LISBOA

UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA LISBOA

TRABALHO FINAL
MESTRADO INTEGRADO EM MEDICINA

Clínica Universitária de Reumatologia

Joint counts variability in rheumatoid arthritis patients followed in a real-life setting

Nuno Filipe de Gouveia Kakoo

Julho'2017

TRABALHO FINAL
MESTRADO INTEGRADO EM MEDICINA

Clínica Universitária de Reumatologia

Joint counts variability in rheumatoid arthritis patients followed in a real-life setting

Nuno Filipe de Gouveia Kakoo

Orientado por:

Vasco Romão

Julho'2017

Abstract

In rheumatoid arthritis (RA), the tender and swollen joint counts play a key role in the clinical assessment of disease activity, integrating the disease activity score (DAS), which is determinant in the treatment decision. However, there is considerable variability in these measurements. In this project, we aimed to identify the variables influencing the joint counts, and to determine if there was a significant difference in these measurements when performed in clinical practice by Rheumatology registrars and consultants. For that purpose, we analyzed the possible effect of several variables on the joint counts, by analyzing consecutive visits (n=3127) of 260 patients followed in a real life setting. We compared the differences in the assessment of disease activity parameters between the registrars and consultants using the Student's T-test. To study the impact of each variable in the tender and swollen joint counts, we conducted a multivariate analysis using a linear mixed model. Overall, registrars and consultants counted a similar mean number of tender but a higher number of swollen joints (1.6±2.8 vs. 2.0±3.7; P-value=0.001). However, when adjusting for other variables, neither the training status, nor the years of experience were associated with the joint counts. Importantly, variables such as the patient global visual analogue scale (VAS), physician VAS and disease duration had a significant impact on both the tender and swollen joint counts. Furthermore, the patient's age was significantly associated with the tender joint count. We also found that the inter-observer and inter-patient variabilities were both significant, which can be further investigated in studies that consider simultaneous assessment of patients by several evaluators. We conclude that neither the physician training status nor the years of experience have impact in the joint counts, but the disease activity parameters do.

Keywords: rheumatoid arthritis, joint counts, disease activity, rheumatology training

Resumo

As contagens de articulações dolorosas e tumefactas adquirem um papel essencial na avaliação clínica da artrite reumatóide (AR), dado que integram o *disease activity score* (DAS), que é determinante na decisão terapêutica. Contudo, há uma variabilidade considerável nestas medições. Neste projecto, definimos como objectivos identificar as variáveis que influenciam as contagens articulares, e perceber se há uma diferença significativa nestas medições quando estas são realizadas por internos vs. especialistas. Para tal, analisámos o possível efeito de várias variáveis com base numa amostra de 260 doentes seguidos na prática clínica em consultas consecutivas (n=3127). Começámos por fazer uma análise descritiva para caracterizar a amostra, e, de seguida, usámos o teste T-Student para perceber se havia diferenças significativas entre os internos e os especialistas nos parâmetros de actividade da doença (análise univariada). Posteriormente, realizámos uma análise multivariada usando o modelo linear misto para estudar o impacte de cada variável nas contagens articulares. Os resultados mostraram que os internos e especialistas contaram um número médio de articulações dolorosas igual, mas em relação às tumefactas, estes últimos contaram valores significativamente mais elevados (1,6±2,8 vs. 2,0±3,7, respectivamente;Valor-P=0,001). Todavia, ajustando a análise para as restantes variáveis, nem o grau de especialidade nem a experiência do médico se revelaram significativamente associados às contagens articulares. De notar, tanto a escala visual analógica global do doente como a do medico e a duração da doença tiveram impacte significativo nas contagens articulares. A idade do doente estava significativamente associada às contagens de articulações dolorosas. Por fim, as variabilidades inter-observador e inter-doente foram significativas, sendo necessário um estudo de avaliação simultânea dos doentes por diferentes médicos para investigar este efeito. Concluindo, nem o grau de especialidade do médico nem os anos de experiência se associam às contagens articulares, mas os parâmetros de actividade da doença sim.

Palavras-chave: artrite reumatóide, contagens articulares, actividade da doença, treino em reumatologia.

Index

1 – Background

Rheumatoid Arthritis (RA) is a chronic systemic inflammatory disease mainly characterized by joint involvement and progressive destruction over time. Its main clinical manifestations are joint tenderness (pain at rest) and swelling, indicating an active synovitis process.

In this context, joint count, which is a clinical method of assessment of disease activity, is meant to be an objective parameter, ideally with little inter and intra-assessor variability, as it is known to correlate with RA patients' mortality and prognosis[1]. This parameter, along with the patient global visual analogue scale (VAS), C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), integrates the disease activity score (DAS; most often DAS28, with 28 joints taken into consideration), which is a validated tool to assess disease activity and is crucial in determining the medical course of action: a DAS28 greater than 5.1 implies active disease; lower than 3.2, low disease activity and lower than 2.6 means remission[2]. Considering the current paradigm of the *Treat to Target* approach – adjusting therapy in order to reach and remain in a low disease activity or remission state – joint counts play a central and important role, being an objective parameter that ought to have the less variability possible.

Furthermore, there is a wide range of joint count methods assessing swelling and/or tenderness and including different numbers of joints, for a total count of 36[3], 42[4], 44[5] or even 66/68[6] instead of 28 joints. Comparing the various methods, it was concluded[7] that the 28-joint index was not inferior to the others, and has since then been the most used in clinical practice and research, considering its simplicity and feasibility. However, some issues exist with this approach: most notably the fact the feet joints are excluded from disease activity assessment, contributing to misclassification of a subset patients that might still present active disease at the level of joints not captured by this tool and that are thus at risk of disease progression and disability.

With this in mind, there is still the issue of inter- and intra-observer variability in assessing joint tenderness and swelling, as each medical professional has his/her own set of technical skills and personal experience that will have an impact on determining if a joint is considered active or not[8]. Moreover, the same assessor will not always reach the very same result regarding a certain patient or joint on two consecutive observations,

although this variability has been shown to be acceptable[9]. This reflects the complex nature of the very method itself.

Both bony swelling and joint deformities (which occur essentially with disease progression) often complicate the swollen joint assessment of rheumatoid patients. Neither should be counted as joint swelling, but both can be present along with joint swelling, making this assessment more complicated than it might seem. Also, RA patients with longstanding disease often develop pannus formation, a fibrous tissue that is the result of prolonged inflammation inside the joint but that does not mean active disease *per se* and should not be accounted for in assessing swollen joints, which is sometimes difficult.

Joint tenderness should be assessed by applying a pressure of $4Kg/cm^2$ (usually considered when the nail bed of the assessor becomes white) to each joint and evaluate the patient's reaction for pain. However, this cannot be done in every joint, as in deeper joints like the hip or shoulder, where active and passive movement should help determine tenderness[10]. Importantly, RA patients frequently have other comorbidities, including chronic pain conditions such as fibromyalgia. These patients present a challenge for the assisting rheumatologist, as there will be an overestimation of tender versus swollen joints (on average ≥7 of the former) – determining the so-called "Fibromyalgic Rheumatoid", a subset associated with higher levels of disability and disease activity scores and with a reduction in quality of life, while showing low swollen joint counts and inflammatory markers[11]. This results in misleadingly high DAS28 scores and often in unnecessary therapy escalation[12]. On the opposite end, there is the situation typical of earlier disease in which joint inflammation is sometimes painless and silent to palpation, although there is already histological and imaging proof of synovitis, which is associated with future bone damage and erosions. Thus, the joint count method is actually only a surrogate of inflammation in the joint and might not perform similarly in different phases of the disease[13]. Finally, there is also a "floor effect" for joint counts, where in low disease activity states joint counts are low and the potential for detecting a change is smaller compared to patients with active disease[12].

It is clear thus that joint counts face several challenges. In order to reduce this variability, a study assessed the effectiveness of a training procedure to standardize the 28 joint count across assessors[14]. The authors found a considerable decrease in inter-observer variation, mainly regarding the swollen joints (mean variation coefficient was reduced from 82% to 59%), along with an increase in sensitivity to detect both tender

(41%) and swollen joints (32%). Even though training is effective, the inter-observer variability is still significant, and joint counts are overall not very well reproducible, requiring the same observer to evaluate patients in each serial visit in an optimal setting[15]. This poses a problem as it is usually not feasible to have the same assessor in every evaluation.

In summary, joint counts represent a central tool in assessing disease activity of RA patients that correlates with patient prognosis and mortality[16] and cannot be replaced solely by self-assessment methods, as these have been demonstrated to correlate poorly with physician assessment of synovitis and swollen joints[17,18]. However, they present several challenges for the rheumatologists in clinical practice and do not reflect joint damage progression and disability[19]. In fact, they can be a source of time consumption in clinical visits and they have been shown to be neglected in most patient visits[20]. Moreover, training status and experience of the clinician assessing the joints is an important factor that should be taken into account.

We hypothesize that joint counts are influenced by the training level of the clinician, being less variable and more consistent with other disease activity parameters for experienced consultants. Furthermore, we hypothesize the impact of clinical experience is influenced by other patient and disease variables.

2 – Aims

Our main objective was to identify variables influencing the 28-joint counts in RA patients evaluated on standard clinical practice. In particular, we aimed to assess if there was a significant difference in the assessment of tender/swollen joints according to the physician's experience (Rheumatology consultant vs. registrar) and number of years in training. Moreover, we would like to determine the inter- and intra-observer variation and if the joint counts are influenced by other factors such as: other disease activity-related variables (such as global patient VAS, ESR, CRP, DAS28), gender, age, disease duration, serological status for rheumatoid factor (RF) and anti-citrullinated protein antibodies (ACPA) and concurrent therapy (glucocorticoids, synthetic and biologic disease modifying anti-rheumatic drugs [DMARDs]).

3 – Methods

We used data from the Rheumatic Diseases Portuguese Register, Reuma.pt, concerning the RA patients treated with biologicals at the Rheumatology and Metabolic Bone Diseases Department of Hospital de Santa Maria, Centro Hospitalar Lisboa Norte. In our Department, these patients are followed in the setting of a day hospital, where treatment efficacy and safety are continuously evaluated and decisions on maintaining/changing treatment are taken. This constitutes the ideal situation for a study like the present one, as patients are evaluated by different rheumatologists, consultants or trainees, on different consecutive visits, thus enabling for comparison of joint counts in separate observations of the same patient and also between different patients.

For this purpose, we analyzed data of all RA patients treated with any biologic DMARD at our Department from January 2014 to July 2016. We considered this timeframe expecting to have a significant pool of data on joint counts performed by different physicians with several levels of training. More specifically, we assessed each patient's visit information regarding the following variables: patients and disease characteristics (age, gender, disease duration and seropositivity for RF and/or ACPA status); tender joint count (28 joints – tjc28), swollen joint count (28 joints – sjc28), ESR, patient global VAS (in a scale from 0 to 100 mm), physician VAS; biologic therapy, including anti-tumor necrosis factor (TNF) drugs (i.e., etanercept, adalimumab,

golimumab, infliximab), tocilizumab (interleukin [IL]-6 receptor antagonist), rituximab (anti-CD20, B cell depleting agent) and abatacept (anti-CD80/86 fusion protein that targets T cell activation by inhibiting antigen presenting cell-mediated co-stimulation); current synthetic DMARDs, including methotrexate, sulfasalazine, hydroxychloroquine, leflunomide, azathioprine and cyclosporine; physician's training status (consultant vs. registrar) and years of experience since beginning of Rheumatology training (in groups of 2 years – from less than 2 to over 22 years – as we hypothesize that the first years of practice account for a great amount of variation in joint counts, thus deserving a special attention). We only included the physicians that, during the study period, were always registrars or consultants, thus excluding those that changed physician category during the study period.

In the analysis, our dependent variables are the number of tender and swollen joints out of a 28-joint count and the predictors are all other variables listed above. We set as continuous variables the patient's age, disease duration, tender and swollen joint counts, ESR, patient global VAS, physician VAS, physician's years of experience and biologic treatment duration; as categorical variables we considered gender, autoantibody status, physician's years of experience in categories, monotherapy and biologic drug class. We defined the continuous variables as mean ± standard deviation (S.D.) and the discrete variables as absolute and relative frequencies. We conducted a univariate analysis to compare differences in disease activity parameters (tjc28, sjc28, patient global VAS, physician VAS and ESR) in visits conducted by registrars or consultants, using the Student's T-test. To study the influence of each independent variable in the tender and swollen joint counts, we conducted a multivariate analysis, by applying a linear mixed model. In this model, the variables could either have a so-called "fixed effect" (like age and gender), where we are interested in their specific value (e.g. 54 years and female), or a "random effect" (like the patient and physician), considering that the model assumes that their effect on tjc28 and sjc28 has a normal distribution with mean zero and a standard deviation to be estimated. In addition, the model has an error term, which again is assumed to be normally distributed with mean zero and standard deviation to be estimated. This last term incorporates all effects and contributions that are not captured in the other terms of the model and that are assumed to be unbiased (thus, the mean of the distribution for the error term is zero). Also important to mention is the fact that each variable comes into the model in an additive (or linear) way. As such, by using a simple additive assumption, we are not considering the possible interactions between variables.

Taking the age and gender as examples, the model estimates the same effect of age for both genders, instead of estimating if there is a different effect of age for male and female. What this means is that we're using a regression model that will estimate the tender and swollen joint counts expressed through this equation:

$$tjc28 \text{ } or \text{ } sjc28$$
$$= mean + age \text{ } (fixed \text{ } continuous) + gender \text{ } (fixed \text{ } discrete)$$
$$+ physician \text{ } training \text{ } status \text{ } (fixed \text{ } discrete) + (\dots)$$
$$+ patient \text{ } (random \text{ } effect) + physician \text{ } (random \text{ } effect) + error$$

Equation 1 - Linear model estimating the tender/swollen joint count

We used the SPSS Statistics, version 23.0.0.0 as our statistical analysis software, and the level of significance was considered at less than 0.05 (confidence interval of 95%).

4 – Results

4.1 - Sample characterization and descriptive analysis

We included 260 patients, who were evaluated in a total of 3127 consecutive visits, with a mean number of 12 visits per patient.

The patients' characteristics are described in Table 1. This is a standard population of established RA with the majority of patients (88.1%) being female, with a mean age of 56 years and 7 months, and a mean disease duration of 13 years and 8 months. One hundred and ninety-four patients (79.2%) were positive for RF and/or ACPA.

Table 1 - Patient sample characteristics

Variable	Total (n = 260)
Age, mean (S.D.), years*	56.6 (12.0)
Gender - female, n (%)	229 (88.1)
Disease Duration, mean (S.D.), years*	13.7 (9.4)
RF and/or ACPA positive, n (%)	194 (79.2)
*at time of first visit considered	

The disease activity parameters in each patient's visit (n = 3127) are described in Table 2. Starting with the tender joint count, there was a mean number of 2.9 ± 4.8 tender

joints. We found that in 44.3% and 12.6% of the visits, the patients had 0 or 1 tender joints, respectively (i.e. in over 50% of the patients' visits, there were 1 or less tender joints; Figure 1). In 25.6% of the visits, the number of tender joints ranged from 2 to 5. The mean number of swollen joints per visit was 1.9 ± 3.3. In 53.6% of the patients' visits, there were 0 swollen joints, while in 13.0% there was one (Figure 2). In 22.7% of the visits, the number of swollen joints ranged from 2 to 5. The mean patient global VAS was of 45 mm, and the most frequent referred value was 50 mm (19.3% of the patients' visits). The mean physician's VAS was 31.4 mm, and 20 mm was the most frequent value (given in 15.8% of the visits). At last, regarding the DAS28, a minimum of 0.5 and maximum of 8.7 with a mean value of 3.3 ± 1.4 was seen. By grouping the DAS28 values into categories (Figure 3), we can see that in most of the visits, the patients had moderate RA activity (36.4%). Patients presented in remission in a significant proportion of visits (35.7%). In 16.6% of the visits, patients had low disease activity and only in 11.3% of the visits the disease was highly active.

Table 2 - Disease activity parameters across visits assessed

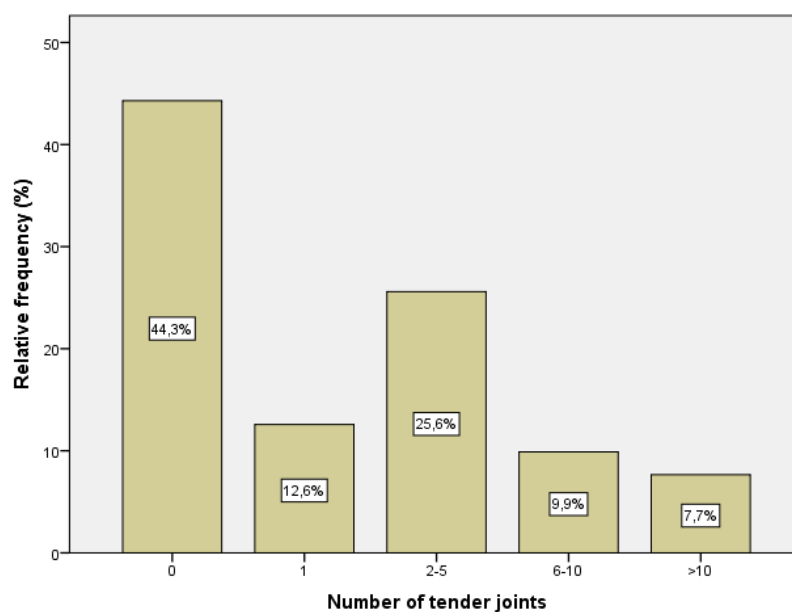| Variable	Total (n = 3127)
Tender joint count (0-28), mean (S.D.)	2.9 (4.8)
Swollen joint count (0-28), mean (S.D.)	1.9 (3.3)
Patient Global VAS, mean (S.D.), mm	45.0 (23.9)
Physician VAS, mean (S.D.), mm	31.4 (21.9)
DAS28, mean (S.D.)	3.3 (1.4)

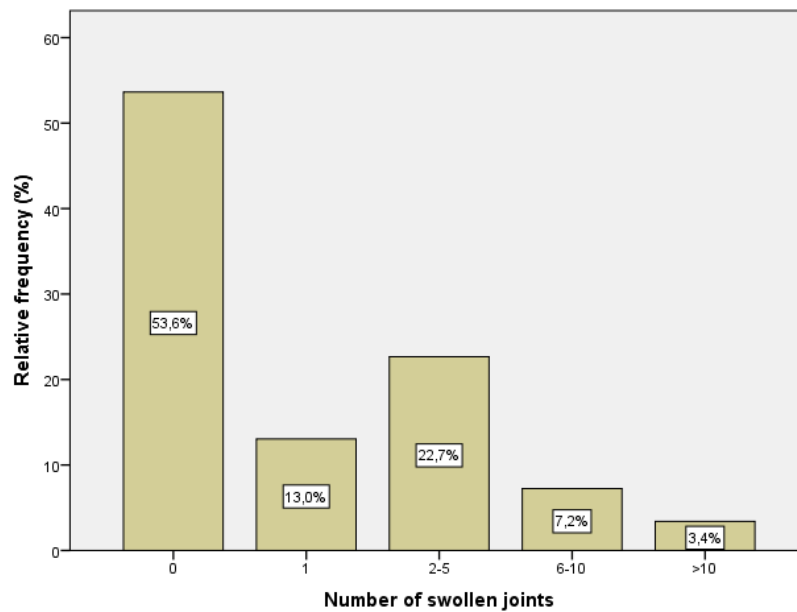*Figure 1 - Proportion of tender joint counts groups across study visits*

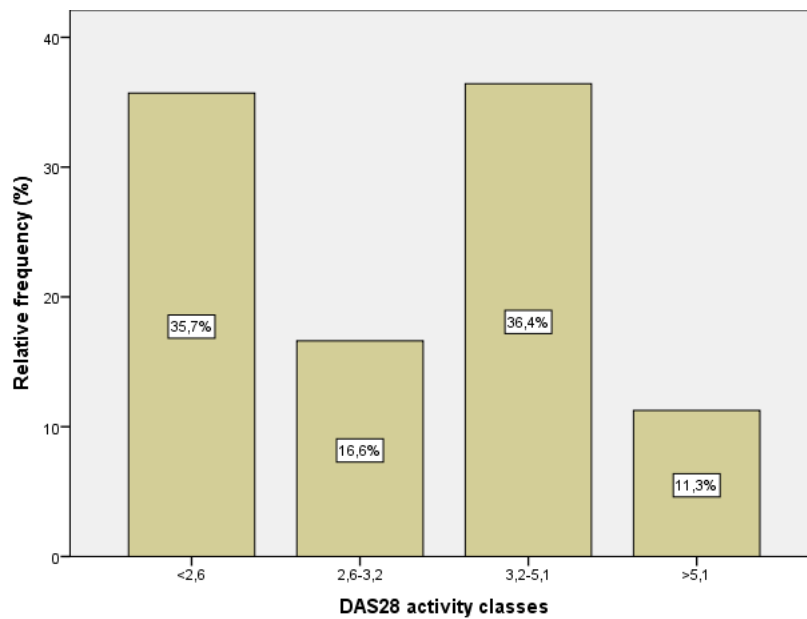*Figure 2 - Proportion of swollen joint counts groups across study visits*

Figure 3 - Distribution of DAS28 activity class across patient visits

Regarding the physicians, this group included 26 observers, out of which 13 were registrars and 13 consultants. Physicians experience at time of visit ranged from 7 days to 22 years of practice. The mean time of experience across visits was 7.6 ± 6.8 years. Figure 4 shows that a significant number of visits (35.5%) was conducted by physicians with only 2 or less years of experience, but there was also a considerable proportion of visits (20.8%) with physicians with 10 to 12 years of experience. The third most prevalent category was the one where physicians had between 2 and 4 years of experience, thus still

as registrars and which accounted for 10.8% of the visits. Mean time since the beginning of training was 1.4 ± 1.0 years for registrars and 13.0 ± 4.8 years for consultants (Figure 5).
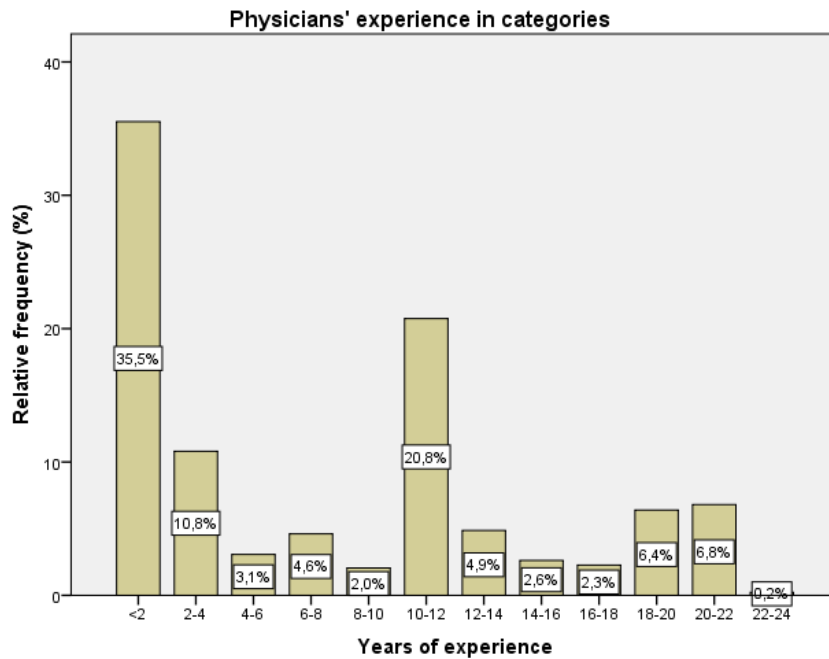

Figure 4 - Physicians' years of experience at the time of visit expressed in categories
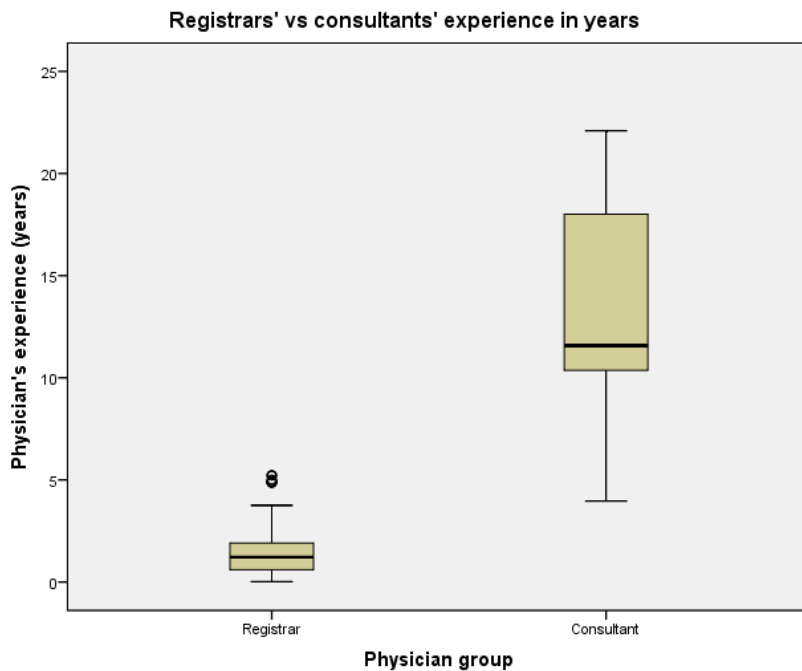

Figure 5 - Physicians' experience at the time of visit comparing Registrars vs Consultants

At last, focusing on the treatment for RA in each patient's visit (Table 3), we can state that in most of them (68.7%), the patients were doing combination therapy with synthetic DMARDs on top of biologics. In 2373 patients' visits (75.9% of the total), they were taking at least one synthetic DMARD. The mean number of active synthetic DMARDs was 0.9 ± 0.7, and it varied between 0 drugs (in 24.1% of the total of patients' visits) and 3 drugs (2.8% of the visits). The most common synthetic DMARD was methotrexate, taken by patients in 2009 visits (64.2%). In 2312 visits (73.9%) patients were taking corticosteroids. Finally, regarding the biologics (Figure 6), in most of the visits, the patients were on anti-TNF therapy (50.4%) and in a significant proportion of visits (39.0%), they were being treated with tocilizumab or rituximab. Only in 9.3% of the visits, patients were not on any biologic therapy.

Table 3 - Rheumatoid arthritis treatment taken by patients in each visit

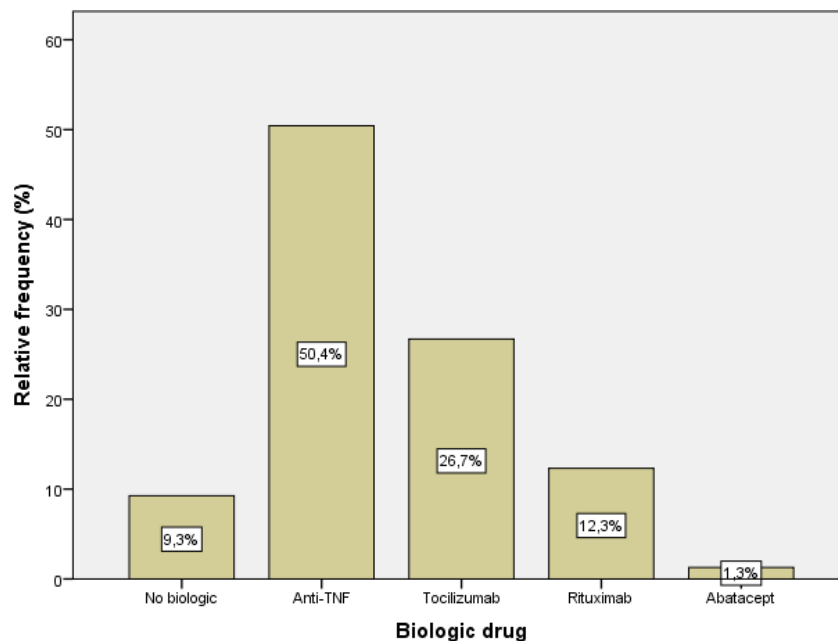| Variable	Total (n = 3127)
Biologic + DMARD , n (%)	2147 (68.7)
Biologic monotherapy, n (%)	690 (22.1)
No biologic, n (%)	290 (9.3)
DMARDS	
Methotrexate, n (%)	2009 (64.2)
Sulfasalazine, n (%)	454 (14.5)
Hydroxychloroquine, n (%)	319 (10.2)
Leflunomide, n (%)	104 (3.3)
Azathioprine, n (%)	7 (0.2)
Cyclosporine, n (%)	16 (0.5)
Corticosteroids, n (%)	2312 (73.9)

Figure 6 - Biologic therapy sorted by biologic drug class

4.2 - Comparison of disease activity parameters assessed by registrars and consultants

The values of disease activity parameters determined in visits performed by registrars and consultants are summarized in Table 4.

We can see that both the swollen joint counts and the physician VAS are significantly different in each group, with the consultants measuring a higher mean count of swollen joints (2.0 ± 3.7 vs 1.6 ± 2.8) but registering lower mean VAS values (28.4 mm vs 34.3 mm) than the registrars. The other three disease activity parameters didn't show any statistically significant difference between the registrars and the consultants.

Table 4 - T-test results regarding the differences between registrars and consultants in the disease activity parameters

Variable	Registrars	Consultants	P-value
Tender joint counts, mean (S.D.)	2.9 (4.3)	3.0 (5.2)	0.559
Swollen joint counts, mean (S.D.)	1.6 (2.8)	2.0 (3.7)	**0.001**
Patient global VAS, mean (S.D.), mm	45.8 (23.5)	44.4 (24.2)	0.140
Physician VAS, mean (S.D.), mm	34.3 (22.0)	28.4 (21.4)	**<0.001**
ESR, mean (S.D.)	21.9 (20.0)	22.3 (20.4)	0.584
P-Values below 0.05 are shown in **bold**			

4.3 - Variables influencing tender and swollen joint counts

We considered the following 12 variables as potentially influencing the joint counts: patient age, gender, disease duration, RF/ACPA status, ESR, patient global VAS, monotherapy, biologic drug class, biologic treatment duration, physician VAS, physician years of experience, and physician training status (registrar vs. consultant).

Regarding the tender joint counts, we found that the model that best explained this dependent variable was the one that included 11 of the variables above mentioned, excluding the physician years of experience (Table 5). The only significant independent variables predicting the tjc28 were age, disease duration, patient global VAS, physician

VAS and biologic treatment duration. The estimated effect of each variable and respective p-value is described in Table 5.

We found that the patient's age has a significant effect in the number of tender joints (thus the latter being higher in older patients), more specifically increasing by 1 tender joint for every additional 25 years. Female patients had a trend towards a higher number of tender joints compared to males, by approximately 1. Longer disease duration, however, was found to negatively influence the tender joint count, decreasing it by 1 for every 16 years and 7 months of disease. Not surprisingly, the patient global VAS had a significant effect in determining the tcj28, increasing it by 1 for every 50 mm increment in VAS. In line with this, the physician's VAS was the most important predictor of tjc28, increasing it by 1 for only 11 mm increment in the scale (P-value < 0.001). Regarding treatment, even though the number of drugs and the biologic drug class weren't significantly associated with tender joint counts, there was a negative association with biologic treatment duration (P-value = 0.039), with an estimated effect of decreasing the tjc28 by 1 for every 8 years and 3 months of treatment. Finally, the training status (registrar/consultant) was not independently associated with the tjc28, when accounting for all the remaining variables (P-value = 0.786). Even though it is not represented in the final model, similar results were found with physician's years of experience as a continuous or categorical variable (P-value > 0.05).

Of note, the inter-observer's variability effect was significant and estimated to be of 1.20 (P-value = 0.003), which means that the tjc28 value can be up to 1.20 higher or lower than the average estimated value (based on all the other variables). It also indicates there are probably unmeasured variables that could explain this effect. Also, the patient's variability effect was highly significant and estimated at 1.79 (P-value < 0.001), which means that in our patient sample, the tjc28 can vary by almost 2 due to a patient "random" effect, which is considered random because the unmeasured variables' effect that could explain this variation is unknown. The model's error variability was 3.33, which means that there can be a variation of up to 3.33 in the tender joint counts that cannot be explained by the model with the variables considered.

Table 5 - Linear mixed model's prediction of 28 tender joint count: independent variables' effect in tender joint counts

Variable	ß-coefficient; p-value
Age	0.04; **0.005**
Gender, female sex	0.91; 0.054
Disease Duration	-0.06; **0.001**
RF and ACPA negative	-0.01; 0.974
ESR	0.01; 0.262
Patient Global VAS	0.02; **0.003**
Physician VAS	0.09; **<0.001**
Combination therapy	0.03; 0.954
Biologic Drug Class	
Anti-TNF	0.00; 1.000
Tocilizumab	-0.30; 0.812
Rituximab	0.53; 0.683
Abatacept	.
Biologic drug duration (months)	-0.01; **0.039**
Physician training status, registrar	0.14; 0.786
P-Values below 0.05 are shown in **bold**	

Regarding the swollen joint counts, there were 4 variables that were neither significant nor contributed to better explain the number of swollen joints – the patient's age, gender, monotherapy/combination therapy and physician's years of experience. As such, only the other 7 variables were included in the model, and out of these, only the patient global VAS, physician's VAS, disease duration and biologic treatment duration were independently associated with the sjc28 (Table 6). How each variable affects the number of swollen joints and their respective level of significance is depicted in Table 6.

The patient global VAS' effect in swollen joint counts was statistically significant (P-value = 0.012) although unlikely to be clinically relevant, as it was estimated to be a decrease of 1 swollen joint per 100 mm reduction in the global VAS. Higher duration of disease and biologic treatment were associated with lower sjc28, with a decrease in 1 swollen joint for every 33 years and 6 months of disease and for every 8 years and 3 months of biologic therapy, respectively. The physician VAS was a significant predictor (P-value < 0.001) of the sjc28, with one additional swollen joint counted for every 12,5 mm increase in the scale. Finally, as for the tjc28, the registrar/consultant status was not significantly associated with sjc28 and neither were the physician's years of experience (P-value > 0.05).

13

Of note, the inter-observer's variability effect was estimated to be of 0.99 (P-value = 0.003), which means that the sjc28 value can be significantly up to 1 joint higher or lower than the average estimated due to a physician "random" effect. Once again, it indicates there are probably unmeasured explanatory variables for this effect. The patient's variability effect was of 1.04 (P-value < 0.001), which means that in our patient sample, the sjc28 can vary by around 1 count due to a patient "random" effect. The model's error variability accounted for 2.33, meaning that there is about a 2.33 variation in the swollen joint count that the model can't explain only with the variables we used.

Table 6 - Linear mixed model's prediction of sj28: independent variables' effect in swollen joint counts

Variable	Effect in sjc28; p-value
Disease Duration	-0.03; **0.014**
RF and ACPA negative	0.37; 0.165
ESR	0.01; 0.059
Patient Global VAS	-0.01; **0.012**
Physician VAS	0.08; **<0.001**
Biologic Drug Class	
Anti-TNF	-0.36; 0.640
Tocilizumab	0.06; 0.938
Rituximab	-0.29; 0.726
Abatacept	.
Biologic drug duration (months)	-0.01; **0.009**
Physician training status, registrar	-0.19; 0.656
P-Values below 0.05 are shown in **bold**	

5 – Discussion

In this study, we hypothesized if the physician training status had an impact on the tender and swollen joint counts, and we found that – on average and in different visits with different patients – consultants counted significantly more swollen joints than registrars, but a similar number of tender joints. However, when we accounted for the effect of other variables in a multivariate model, as well as the patient and physician error, we failed to confirm this finding, as neither tender nor swollen joints were significantly associated with the training status (registrars vs. consultants). Moreover, we also failed to confirm what we had initially proposed: that the years of experience since the beginning of rheumatology training had an impact on joint counts. However, the inter-observer variability values were significant for both the tender and swollen joint counts,

confirming that there are other variables not considered in the model that can explain some variability in joint counts when performed by different assessors. It is also important to mention the patient variability, which in both the tender and swollen joint assessments was significant, indicating that perhaps some patient comorbidities – such as depression and fibromyalgia – could be variables of interest to explain these differences. Unfortunately, we were not able to confirm this due to a significant amount of missing data on these variables that would limit the construction of a multivariate model.

Several variables did influence the tender joint counts though, the most important and significant (P-value < 0.001) of which was the physician VAS, followed by the disease duration, patient age and patient global VAS. This reveals the great influence that other disease parameters (physician VAS, patient global VAS and disease duration) have on the tjc28, which corroborates the fact that this variable is indeed adequate to measure the RA activity. The evaluating physician plays a great role in the joint counts, as his global impression of disease is strongly and positively influenced by the number of tender joints assessed.

Concerning the swollen joints, once again, the most important and significant predictor (P-value < 0.001) was the physician VAS, followed by the patient global VAS, disease duration and biologic treatment duration. However, in opposition to what we had found regarding the tjc28, the patient's age was not significantly associated with swollen joints assessed. This can result from the fact that assessing a swollen joint is actually a less patient-dependent activity compared to tender joint assessment, with the main role being performed by the physician himself. According to this, even though the patient VAS was independently and positively associated with swollen joint counts, its overall impact was low. Also interesting to note is that the biologic treatment duration had the very same impact in both the tender and swollen joint counts. Just like what we had observed with the tjc28, these data confirm the importance of sjc28 in the clinical assessment of RA activity. The physician role seems to be the most important, as the physician VAS is strongly influenced by the swollen joint counts and accounts for the greatest increase in sjc28.

The main limitation of this study is the fact that we did not manage to fully exclude the actual variation in disease activity as a confounding factor accounting for the tender and swollen joint count variability. Even though the study period is not too long (from January 2014 to July 2016), change in disease activity over time is still expected, which may have led to some limitations in the estimated effects of our independent variables.

We tried to address this by using a linear mixed model but we acknowledge that the ideal setting to answer the question we were proposing would be to have a simultaneous evaluation of the same patients by different physicians and then compare the joint counts across groups. Another limitation is that the statistical analysis model we used assumed that the error had a normal distribution and that the dependent variables – tjc28 and sjc28 – were continuous variables (when they actually followed a Poisson distribution). Using a generalized mixed model would have allowed correcting for the latter inadequacy, but it would be too complex and long-taking, making it too hard to actually apply in this context. Yet another limitation concerning the model we used was that we did not consider the interactions between variables, because, unfortunately, there are too many possible interaction terms (even when just thinking of two-way interactions), and it is thus not feasible to estimate them in such a model. Also, we used the Student's T-test to assess the differences in the tender and swollen joint counts in registrars vs. consultants, even though neither the tjc28 nor the sjc28 followed a normal distribution. A non-parametric test could have been used in this case, but considering the large number of visits (over 3000) we considered it appropriate to apply the Student's T-test. Finally, another limitation was related to the fact that, even though there were some variables – such as erosive/non-erosive disease, smoking status, body mass index, CRP, patient pain VAS, health-assessment questionnaire (HAQ) disability index, previous biologic therapy, current glucocorticoids and NSAIDs, comorbidities such as fibromyalgia, depression, anxiety (or concurrent medication for these conditions) – that could be of value to explain the variation in tjc28 and sjc28, we had to exclude them, due to the impracticability of analyzing every possible variable in one study, and also considering that some of them had a great number of missings (namely patient pain VAS, HAQ disability index, and comorbidities such as fibromyalgia).

In conclusion, in a real-life setting, the variables that mostly affect the tender and swollen joint counts are other disease-related parameters (mainly patient global VAS, physician VAS and disease duration) and overall there does not seem to be a significant difference in the assessment of tender and swollen joints between registrars and consultants. Moreover, physician years of experience also do not seem to have an independent effect in joint counts, but there is still a high inter-assessor and inter-patient variability, which can be further investigated in future studies with a different design, considering simultaneous assessment of patients by several evaluators.

6 – Acknowledgments

I would like to thank my advisor, Vasco Romão, and the Biostatistics teacher, Ruy Ribeiro, as this project would not have been possible without all their support.

7 – References

1. Scott DL, Houssien DA. Joint assessment in rheumatoid arthritis. 1996;35:suppl 2.

2. Available at: http://www.nras.org.uk/the-das28-score.

3. Egger MJ, Huth DA, Ward JR, Reading JC WH. Reduced joint count indices in the evaluation of rheumatoid arthritis. *Arthritis Rheum* 1985;28:613-9.

4. Sokka T PT. Eligibility of patients in routine care for major clinical trials of anti-tumor necrosis factor alpha agents in rheumatoid arthritis. *Arthritis Rheum* 2003;48:313-8.

5. Van Der Heijde D, Klareskog L BM et al. Comparison of different definitions to classify remission and sustained remission: one year TEMPO results. *Ann Rheum Dis.* 2005.

6. Association Cccotar. A seven-day variability study of 499 patients with peripheral rheumatoid arthritis. *Arthritis Rheum* 1965;8:302-35.

7. Fuch, HA, Brooks RH, Callahan LF PT. A simplified 28 joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum* 1989;32:531-537.

8. Klinkhoff AV, Bellamy N BC et al. An experiment in reducing interobserver variability of the examination for joint tenderness. *J Rheumatol* 1988;15:492-4.

9. Cheung PP, Gossec L, Mak A ML. Reliability of joint count assessment in rheumatoid arthritis: A systematic literature review. *Semin Arthritis Rheum* 2014;43:721-9.

10. PJ M. Measures of pdoriatic arthritis: Tender and Swollen Joint Assessment, (...), Composite Psoriatic Disease Activity Index (CPDAI). *Arthritis Care Res* 2011;63(11):S64-85.

11. Leeb BF, Andel I, Sautner J, Nothnagl T RB. The DAS28 in rheumatoid arthritis and fibromyalgia patients. *Rheumatology* 2004;43:1504-7.

12. Scott IC, Scott DL. Joint counts in inflammatory arthritis. *Clin. Exp. Rheumatol.* 2014;32(4):S7-S12.

13. Sokka T, Kautiainen H, Mottonen T HP. Erosions develop rarely in joints without clinically detectable inflammation in patients with early rheumatoid arthritis. *J Rheumatol* 2003;30:2580-4.

14. Scott DL, Choy E GA et al. Standardising joint assessment in arthritis. *Clin Rheumatol* 1994;13:182.

15. Pincus T. Limitations of a quantitative swollen and tender joint count to assess and monitor patients with rheumatoid arthritis. *Bull. NYU Hosp. Jt. Dis.* 2008;66(3):216-223.

16. Pincus T, Brooks RH CL. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. *Ann Intern Med* 1994;120:26-34.

17. Choy EH, Khoshaba B, Cooper D, MacGregor A SD. Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. *Arthritis Rheum* 2008;59:192-9.

18. Wolfe F, Pincus T OJ. Evaluation and documentation of rheumatoid arthritis disease status in the clinic: which variables best predict change in therapy? *J Rheumatol* 2001;28:1712-7.

19. Callahan LF, Pincus T, Huston JW III et al. Measures of activity and damage in rheumatoid arthritis: depiction of changes and prediction of mortality over five years. *Arthritis Care Res* 1997;10:381-94.

20. Pincus T SO. Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. *Ann Rheum Dis.* 2006;65:820-2.

8 – Appendix

Resumo do projecto:
Variabilidade das contagens articulares em doentes com Artrite Reumatóide seguidos na prática clínica

A Artrite Reumatóide (AR) é uma doença inflamatória sistémica caracterizada pelo envolvimento articular com destruição progressiva das articulações. As principais manifestações clínicas são a dor e tumefacção articulares, sendo que esta última indica a existência de sinovite.

O método das contagens articulares, que avalia a existência de articulações dolorosas e tumefactas, é utilizado na avaliação clínica dos doentes com AR, uma vez que se correlaciona com a mortalidade e prognóstico da mesma. Como tal, pretende-se que este parâmetro seja o mais objectivo possível, com pouca ou nenhuma variabilidade inter e intra-observador. As contagens articulares, juntamente com a escala visual analógica (EVA) global do doente, EVA do médico, a velocidade de sedimentação (VS), e a proteína C-reactiva (PCR), integram o *disease activity socre 28* (DAS28), que é utilizado na prática clínica não só para avaliar a actividade da doença, mas também para ajudar na decisão terapêutica. Assim, um score < 2,6 indica remissão da doença; < 3,2 baixa actividade e > 5,1 doença activa.

Existem vários métodos de contagens articulares, incluindo o que avalia 36, 44, 46 e até 66/68 articulações, mas o mais usado é sem dúvida o de 28 articulações, uma vez que não só foi comprovado não ser inferior aos outros, mas também é o mais simples e prático. Todavia, este apresenta limitações como o facto de não incluir as articulações dos pés, criando assim uma falha importante relativa aos doentes com envolvimento particular destas articulações, que podem ter doença activa, mas cujo DAS28 não vai permitir detectar.

A não esquecer ainda, temos a própria variabilidade entre médicos, que se justifica com o facto de cada avaliador ter os seus métodos e experiência pessoais que irão ser determinantes na forma como as contagens articulares são feitas. Também o mesmo observador pode avaliar de formas diferentes o mesmo doente em consultas consecutivas, mas esta variabilidade foi comprovada ser aceitável.

A complicar a avaliação das articulações tumefactas, temos o edema ósseo e deformidades articulares, que, embora não devam ser consideradas como tumefacção articular, são facilmente confundíveis como tal. Existe ainda a formação do "pannus" (resultante de inflamação com produção de tecido fibroso) aquando de uma duração prolongada de doença que se assume como factor de confusão na avaliação das articulações tumefactas, pese embora este por si não signifique doença activa.

Na avaliação das articulações dolorosas, a presença de comorbilidades como a dor crónica (p.e. fibromialgia) representam um desafio, pois conduzem a uma sobrestimação (em média de 7 articulações dolorosas a mais, em comparação com as tumefactas) das contagens articulares, representando este subgrupo de doentes um com *scores* de actividade da doença demasiado elevados e baixa qualidade de vida, apesar do reduzido número de articulações tumefactas e de valores dos marcadores de inflamação. Isto culmina em valores de DAS28 irrealisticamente elevados e escalada terapêutica desnecessária. No extremo oposto, temos os doentes em fase inicial de doença, que, como tal, são clinicamente silenciosos, apesar das evidências histológica e imagiológica de doença activa e que poderá evoluir para dano e erosão ósseos. Para concluir, existe uma variedade de factores que complicam o método das contagens articulares e que o tornam num desafio clínico diário, o que é de todo indesejável, dada a elevada importância clínica que assume.

Assim sendo, este projecto teve como objectivo o estudo da variabilidade nas contagens das articulações dolorosas e tumefactas em doentes com AR avaliados na prática clínica habitual. Com esta ideia em mente, pretendemos verificar o efeito dos anos de experiência do médico e respectivo grau de especialidade (interno ou especialista) nas contagens articulares, e identificar as variáveis que as possam influenciar.

Para tal, usámos os dados do Registo Nacional de Doenças Reumáticas – Reuma.pt – que contém as informações relativas aos doentes com AR e que estão a fazer terapêutica biológica no Departamento de Reumatologia e Doenças Ósseas Metabólicas do Hospital de Santa Maria, Centro Hospitalar Lisboa Norte. A janela temporal a que o nosso estudo se refere é de Janeiro de 2014 a Julho de 2016, e foram avaliadas 3127 consultas relativas a 260 doentes diferentes e 26 médicos diferentes. As variáveis que trabalhámos foram as seguintes: sexo e idade do doente, duração da doença, seropositividade para factor reumatóide (FR) e/ou anticorpos anti-péptidos citrulinados (AAPC); contagens de articulações dolorosas (num total de 28), contagens de articulações tumefactas (num total de 28), VS, EVA global do doente, EVA do médico; terapia biológica; DMARDs

(fármacos modificadores da doença para a AR), grau de especialidade do médico (interno vs. especialista) e anos de experiência desde o início do internato em Reumatologia. Começámos por fazer uma análise descritiva das variáveis, definindo as contínuas como média ± desvio padrão, e as discretas como frequência absolutas e relativas. Posteriormente, fizemos uma análise univariada com o teste T-Student para verificar as diferenças nos parâmetros de actividade da doença (articulações dolorosas, articulações tumefactas, EVA doente, EVA médico e VS) das consultas dadas por internos vs. especialistas. Finalmente, fizemos uma análise multivariada para estudar o efeito individual de cada variável independente nas contagens de articulações dolorosas e tumefactas, usando para tal um modelo linear misto.

Dos 260 doentes incluídos, 88,1% eram do sexo feminino com média de idades de 56 anos e 7 meses e duração média da doença de 13 anos e 8 meses. Setenta e nove por cento dos doentes eram seropositivos para FR ou AAPC. O número médio de articulações dolorosas avaliado foi de 2,9 ± 4,8, e em mais de 50% das consultas, os doentes tinham uma ou zero articulações dolorosas. O número médio de articulações tumefactas foi de 1,9 ± 3,3, e em mais de 50% das consultas foram contabilizadas zero articulações tumefactas. O EVA médio dos doentes foi de 45,0 mm, e o dos médicos 31,4 mm. Quanto ao DAS28, o valor médio foi de 3,3 ± 1,4. Agrupando os valores de DAS28 por categorias, percebemos que, em 36,4% das consultas, os doentes tinham actividade moderada da doença e, em 35,7% delas, estavam em remissão. Relativamente aos médicos, houve 13 internos e 13 especialistas, e o tempo médio de experiência dos mesmo ao longo das consultas foi de 7,6 ± 6,8 anos. Em 35,5% das consultas, eles tinham menos de 2 anos de experiência, e em 20,8% delas, tinham 10 a 12. Finalmente, no que toca à terapêutica, na maioria das consultas (68,7%), os doentes estavam a fazer terapêutica biológica combinada com DMARDs. O número de DMARDs médio foi de 0,9 ± 0,7 e na maioria das consultas (75,9%), os doentes estavam a tomar pelo menos um fármaco desta classe, sendo o metotrexato o fármaco mais comum (activo em 64,2% das consultas). Quanto à corticoterapia, os doentes estavam sob corticosteróides em 73,9% das consultas. Por fim, mencionando os biológicos, a terapêutica anti-TNF foi observada em 50,4% das consultas, sendo seguida pelo uso de tocilizumab ou rituximab, o que se verificou em 39,0% dos casos.

Relativamente às diferenças entre internos e especialistas em termos de avaliação de parâmetros de actividade da doença, observámos que os especialistas contabilizam em média um número médio de articulações tumefactas significativamente maior do que os

internos (2,0 ± 3,7 vs. 1,6 ± 2,8, respectivamente; Valor-P = 0,001), ainda que registem valores de EVA mais baixos (28,4 mm vs. 34,3 mm, respectivamente; Valor-P < 0,001). Não houve diferença significativa nos restantes parâmetros da doença, incluindo as contagens de articulações dolorosas (3,0 ± 5,2 nos especialistas vs. 2,9 ± 4,3 nos internos; Valor-P = 0,559)

Por fim, quanto à análise multivariada, foram estudadas as seguintes variáveis: sexo e idade do doente, duração da doença, seropositividade, EVA doente, EVA médico, anos de experiência do médico, grau do médico (interno ou especialista), VS, monoterapia, classe do biológico e duração do biológico. Quanto ao modelo que melhor explicou as contagens de articulações dolorosas, este incluiu todas as variáveis supramencionadas, com excepção dos anos de experiência do médico. Tanto a idade como a EVA global do doente, EVA do médico, a duração da doença e duração do tratamento com biológicos revelaram ser variáveis significativamente associadas às contagens de articulações dolorosas, sendo que as duas últimas variáveis estão associadas a uma diminuição das contagens de articulações dolorosas, enquanto as três primeiras a um aumento. De notar que não houve diferenças significativas entre os internos e especialistas, nem os anos de experiência tiveram influência significativa (Valor-P > 0,05) nas contagens de articulações dolorosas. Todavia, a variabilidade inter-observador foi significativa e teve o valor de 1,20 (Valor-P = 0,003), o que indica que pode haver uma variação do número de articulações dolorosas até 1,2 superior ou inferior ao valor médio estimado com base em todas as restantes variáveis. A variabilidade devido ao doente foi também ela significativa (Valor-P < 0,001) e no valor de 1,79, o que se fica a dever ao efeito de variáveis não consideradas no estudo. Em relação ao modelo das contagens de articulações tumefactas, nem a idade nem sexo do doente, nem as variáveis monoterapia/terapêutica combinada ou os anos de experiência do médico foram incluídas, dado não contribuírem para explicar a variação das contagens de articulações tumefactas. Para além disso, apenas a EVA global do doente, EVA do médico, duração da doença e duração do tratamento com biológicos revelaram ser variáveis estatisticamente significativas. Destas, apenas a EVA do médico estava associada a um aumento da contagem de articulações tumefactas (de 1 articulação por cada aumento de 12,5 mm na escala (Valor-P < 0,001)), enquanto as restantes 3 variáveis estavam associadas a uma diminuição da contagem articular. Mais uma vez, nem o grau de especialidade do médico, nem os seus anos de experiência tiveram qualquer impacto na contagem de articulações tumefactas. Todavia, tanto a variabilidade inter-observador como a relacionada com os

doentes (0,99 e 1,04, respectivamente; Valores-P de 0,003 e < 0,001, respectivamente) foram significativas, o que corrobora a existência de outras variáveis não medidas neste estudo que poderiam explicar este efeito.

A principal limitação deste estudo foi o facto de não termos conseguido excluir a variação da actividade da doença como um factor confundidor na estimação dos efeitos de cada variável independente sobre as contagens articulares. Usámos o modelo linear misto que incluiu os parâmetros da actividade da doença numa tentativa de contornar esta limitação, mas reconhecemos que o ideal teria sido os mesmo doentes terem sido avaliados na mesma data por médicos diferentes. Relativamente ao modelo linear misto que usámos, o facto de este assumir que a distribuição do erro era normal e o facto de não termos considerado as interações entre as variáveis também se assumiram como limitações. Em relação ao teste T-Student, este assume que as distribuições das contagens articulares eram normais (o que não corresponde à realidade), o que, por sua vez, constituiu mais uma limitação. Por fim, tivemos de excluir várias variáveis do nosso estudo (como estado erosivo da doença, Índice de massa corporal do doente, comorbilidades como depressão e fibromialgia, entre outros), o que se ficou a dever não só à existência de vários missings, mas também à inexequibilidade de uma análise tão morosa que integrasse todas as variáveis.

Em conclusão, podemos afirmar que nem o grau de treino do médico nem os anos de experiência têm um impacto significativo nas contagens de articulações dolorosas ou tumefactas, sendo estas especialmente determinadas pelos parâmetros relacionados com a doença – nomeadamente a EVA global do doente, EVA do médico e duração da doença. Esta observação vem reforçar a pertinência das contagens na avaliação da actividade da AR. Por outro lado, é de realçar o facto das variabilidades entre observadores e devida ao doente serem significativas, o que nos leva a supor que talvez outros factores como comorbilidades do doente (depressão e/ou fibromialgia) possam ser variáveis importantes a considerar em estudos futuros em que se dê prioridade à avaliação simultânea dos mesmos doentes por diferentes médicos.