**Resposta aos Revisores**

We thank the reviewers for all suggestions and corrections. A point-by-point reply ensues and all modifications are included in the revised manuscript, accordingly.

***Revisor B Comentário 1:*** Abstract

*.........De acordo com o Disease Activity Score (DAS28), a proporção de doentes* *com atividade alta, moderada, baixa ou em remissão mudou de 50, 45, 0 e 5 (pré-biológico) para 11, 37, 25 e 26 (na altura da re-avaliação) valores absoluto ou* *percentagem?, respectivamente, com melhoria funcional. Setenta e cinco ~~75%~~ dos* *doentes manteve o tratamento original com boa adesão. .......*

Answer to Revisor B Comentário 1: These changes have been introduced in the abstract, both in the Portuguese and the English versions

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Revisor G Comentário 1:***

*For example, in the title, the words “cohort” and “standardized” in the English version, do not have a match in its Portuguese one.*

**Answer to Revisor G Comentário 1:** The Portuguese title now reads “Artrite reumatóide em coorte de doentes submetidos a terapêutica biológica: monitorização padronizada num centro terciário de referenciação em Portugal” and this has been changed in the manuscript.

***Revisor G Comentário 2:***

*In the Portuguese abstract it is missing information that is in the English version.*

**Answer to Revisor G Comentário 2:** corrected

***Revisor G Comentário 3:***

*This study is stated as a retrospective cross-sectional study but also a study taken on a cohort of patients. Maybe this study should be classified cross-sectional cohort study
(Hudson JI, Pope HG Jr, Glynn RJ, The cross-sectional cohort study: an
underutilized design. Epidemiology: 16(3):355-9. 2005.*

**Answer to Revisor G Comentário 3:** We agree and this has been corrected

***Revisor G Comentário 4:***

*ABSTRACT (both in Portuguese and English): when referring to the distribution of patients according to DAS28, proportions are being displayed as percentages and that should be indicated – either substituting “proportion” by “percentage” or including the symbol “%” after the numbers displayed.*

**Answer to Revisor G Comentário 4:** Proportion was substituted by percentage, also as indicated by Revisor B.

***Revisor G Comentário 5:***

*INTRODUCTION, 3rd paragraph: “… progression and remission are robust…”: not sure what robust means here. Is it statistical or medical?*

**Answer to Revisor G Comentário 5:** Robust means medical - in the sense that in the 1990s, cut-off points for the composite measure DAS28 have been developed and are still used in the present time in clinical practice, to classify patients in remission, as well as low, moderate, and severe disease activity. [1] [2]

***Revisor G Comentário 6:***

*6. HEALTH RELATED QUALITY OF LIFE, end of second paragraph: “In the present study SF-36 was standardized according to the norms for the Portuguese population.” -> maybe better to say that it was adapted and validated to the Portuguese population. Not sure what “standardized do the norms of the Portuguese populations” means.*

**Answer to Revisor G Comentário 6:** SF-36 has been adapted and validated for the Portuguese population in the past [3, 4]. This study analysed the SF-36 questionnaire according to the validation, that imposes defined corrections for some of the parameters. The analysis was performed by Dr. Céu Mateus, a Health Economist, using the Portuguese corrections. The text now reads: In the present study, SF-36 was analysed according to the norms for the Portuguese population.

***Revisor G Comentário 7:***

*7. STATISTICAL ANALYSIS: The 1st sentence is not clear. Does it mean that the continuous variables that were tested for normality and passed are summarized in this paper by their means and standard deviations and the others don’t or does it mean that those variables whose sampling distribution look symmetric are summarized in this paper by their means and standard deviations and the others don’t? In either case the beginning of the sentence should be changed to something like “According to the normality (or symmetry) of their sampling distributions, continuous variables were recorded…”*

**Answer to Revisor G Comentário 7:** The beginning of the sentence has been changed as suggested. Only Health-Related Quality of Life variables are considered to have a normal distribution. All other independent parameters were tested for normality and were not normally distributed.

***Revisor G Comentário 8:***

*7. STATISTICAL ANALYSIS: Correlations for nominal variables can only be calculated for those that are ordinal.*

*7. STATISTICAL ANALYSIS: It is mentioned that “Only moderate correlations (Spearman rho values <0.40) were considered for the regression model”. I guess that this means generally that for any regression model that is considered in this paper, only independent variables that are moderately correlated with the response variable are included in the model. Is that it? And by the way 2 things: a) there are several regression models, I guess that you are referring to multiple regression model (response gaussian, linear predictor); b) why not including in the model independent variables (or covariates) that might be strongly correlated to response?*

7. STATISTICAL ANALYSIS: “Nominal two-sided p values of <0.05 were considered statistically significant” -> “It was considered that nominal two-sided p values of <0.05 corresponded to statistically significant covariate (or independent variable) effects.”

**Answer to Revisor G Comentário 8:** The Pearson correlations can only be used for the Quality of Life indices as the other variables are not normally distributed, amongst other factors. The multiple linear regression model fails on several assumptions, particularly because of non-normality and co-linearity. We have withdrawn the correlations between clinical parameters as well as the model.

For the statistical analysis in METHODS the text now reads: “According to the normality of their sampling distributions, continuous variables were recorded as means (± standard deviation) or medians (IQR1 - IQR3). Dichotomous variables were examined by frequency distribution, recorded as percentage, and comparisons were made using the Chi-square test or Fisher’s exact test for independent samples. The T test was used for comparison between normally distributed variables. For continuous data that were not normally distributed, the Wilcoxon Signed Rank test was used to compare paired means and the Mann-Whitney *U* test for comparisons between independent samples. Correlations between continuous variables were analysed with Pearson's correlation coefficient. Overall, only nominal two sided p values of <0.05 were considered statistically significant. Analyses were performed using SPSS version 23 software.”

***Revisor G Comentário 9:***

RESULT 1. INITIAL PATIENT COHORT: DEMOGRAPHIC AND CLINICAL FEATURES:

- “The majority of patients were female and significantly younger than the male counterpart at bDMARD onset (p=0.038).” -> The test used is bilateral so you can only conclude that ages were different, not that female were younger.

**Answer to Revisor G Comentário 9:** Themedian age of the patients is shown in supplementary table I. Females are younger than males. According to the statistical test, this is a statistically significant difference.

***Revisor G Comentário 10:***

- “There were no other differences between the types of bDMARD” -> “other” is not appropriate here;

**Answer to Revisor G Comentário 10:** corrected in the text

***Revisor G Comentário 11:***

- “The first patient started bDMARD in 2002” -> This cannot be seen from the referred table. Maybe include this information somewhere else, when introducing data?

**Answer to Revisor G Comentário 11:** It has now been added in Material and Methods and removed from this section.

***Revisor G Comentário 12:***

- “Treated dyslipidemia and arterial hypertension were frequently documented” -> Value?

**Answer to Revisor G Comentário 12:** These percentages have now been added to the text.

***Revisor G Comentário 13:***

- “The Charlson comorbidity index estimated a median 10-year survival of 90%.” -> “The median of the estimated 10-year survival using the Charlson comorbidity index was 90%.”

**Answer to Revisor G Comentário 13:** corrected in the text

***Revisor G Comentário 14:***

RESULT 2. RECALL PATIENT COHORT: DEMOGRAPHIC AND CLINICAL FEATURES:

- “...the majority of whom between 47 and 66 years of age (table II)” -> I would rather say that “...a major part of whom between 47 and 66 years of age (table II)”. Thus is because many are >=67.

**Answer to Revisor G Comentário 14:** the text now reads: “Recruitment was achieved in 77 patients, a major part of whom were between 47 and 66 years of age (table II)”.

***Revisor G Comentário 15:***

- “However, as regards professional status almost half of the population was in retirement (49%) and 28% were unemployed. “However” -> it doesn’t seem strange since 36% are of age >=67; “population” -> “sample”; “in retirement” -> “retired”; “28%” -> From the table the value is 10.5%; Is being a student or a homemaker considered to be unemployed?

**Answer to Revisor G Comentário 15:** The value 28% was an error. The text now reads “as regards professional status almost half of the sample was retired (49%) and 10.5% were unemployed”.

***Revisor G Comentário 16:***

- “The majority (40.8%) only had 4 years of schooling; 11% had completed 12 years of schooling and 17% underwent a university education, the latter three times more frequent in men. -> this ratio cannot be seen from table II.

**Answer to Revisor G Comentário 16:** this observation has now been added to table II and it states in the text “the latter almost three times more frequent in men”.

***Revisor G Comentário 17:***

- “Over two thirds (79%) of the patients remained on bDMARD (n=61), the majority of whom were on etanercept and tocilizumab” -> You cannot see this from the table (III), can you?;

**Answer to Revisor G Comentário 17:** Data has now been added to table III (it was in supplementary table II).

***Revisor G Comentário 18:***

- “One third of patients (n=22) switched from the original bDMARD, 14 having required further switches” -> What is stated in table III is that the 14 just switched 1 time, right?

**Answer to Revisor G Comentário 18:** Right, this was a mistake. It has been corrected in the text which now reads: “One third of patients (n=22) switched from the original bDMARD, 8 having required further switches.”

***Revisor G Comentário 19:***

- “Likewise, steroid use significantly decreased from 74 to 30%, (p<0.001, Chi square test between frequencies), halving the prednisolone” -> remove comma, correct typo;

**Answer to Revisor G Comentário 19:** comma removed

***Revisor G Comentário 20:***

-“Smoking, obesity, chronic anxiety and depression were only…” -> The information from this sentence on is in Supplementary Table II and this should be written from here for an easier reading.

**Answer to Revisor G Comentário 20:** the information has been added to the text.

***Revisor G Comentário 21:***

- “Unlike before, the Charlson comorbidity index was now higher in man than…” -> “men”

**Answer to Revisor G Comentário 21:** corrected

***Revisor G Comentário 22:***

- “At recall, DAS28-ESR and HAQ were significantly lower when compared to the pre-bDMARD values, both with a p value <0.001,…” -> shouldn’t this comparison be made using Wilcoxon Signed Rank since this corresponds to paired data? - “At recall, DAS28-ESR and HAQ were significantly lower when compared to the pre-bDMARD values, both with a p value <0.001, [CI -2.56 to -1.6 and -1.04 to -0.533) respectively (Wilcoxon Signed Rank test)]” -> The p-value refers to the Wilcoxon Signed Rank tests for paired data; The CI’s are for the differences. But how are these CIs calculated? Assuming normality? Be careful because when you display summary measures related to these variables you always use medians and IQRs, so I’, guessing that normality doesn´t not hold for them…

**Answer to Revisor G Comentário 22:** Confidence intervals have been removed. The text now reads: “At recall, DAS28-ESR and HAQ were significantly lower when compared to the pre-bDMARD values, both with a p value <0.001 (Wilcoxon Signed Rank test). However the median DAS28-ESR of 3.2 (IQR 2.390 – 3.950) was borderline between low and moderate disease activity and the median HAQ-DI index was 1.310 (IQR 0.500 – 1.810), indicative of mild to moderate disability. Additionally, at recall, there was no significant difference in DAS28-ESR or HAQ when comparing patients on bDMARD with patients off bDMARDs (p=0.917 and 0.975, respectively - Mann Whitney-*U* test). Of note, the pre-bDMARD HAQ could only be obtained for 28 of the 77 patients at recall.”

***Revisor G Comentário 23:***

- “However the recall median DAS28-ESR of 3.2 (IQR 2.390 – 3.950) was borderline between low and moderate disease activity, irrespective of current bDMARD therapy (p=0.917 Mann Whitney-U test – MWU) -> What is this testing?

**Answer to Revisor G Comentário 23:** This is the comparison of DAS28-ESR between patients (i) *on* bDMARD and patients (ii) *off* bDMARDs, at recall. Its been explained in the text as seen in above reply.

***Revisor G Comentário 24:***

- “The median HAQ-DI index at recall was 1.310 (IQR 0.500 – 1.810) similarly indicative of mild to moderate disability and best correlated to pre-bDMARD HAQ” -> Why is this?

**Answer to Revisor G Comentário 24:** “and best correlated to pre-bDMARD HAQ” has been removed from the text:

***Revisor G Comentário 25:***

a) - “Recall functional impairment measured by HAQ was only moderately positively correlated with pre-bDMARD HAQ (r=0.563, p=0.002) and negatively moderately correlated to SF-36 HR-QoL indices (PCS and MCS)” -> in table 2 these correlations show a positive value in the last line;

b) -“…and EQ-5D-3L VAS” -> here correlation is strong.

c) - “Spearman correlations between clinical parameters revealed that retention of the original bDMARD” -> this is a yes/no question, right? No order… Same for “pre-DMARD MTX use”.

- Recall disease activity measured by DAS28 was significantly determined by pre-bDMARD DAS28, number of deformities per patient and number of patients on steroids that had stopped bDMARD therapy (adjusted R2 0.9, p< 0. 001), the latter contributing most significantly to recall disease activity.” -> several notes:

-> The results in this paragraph are in Supplementary Table III and this should be stated in the beginning of the paragraph; -> I believe that the p-value presented is for the test that simultaneously all the coefficients of the regression model are null. As such you should state it as a model adequacy measure. ->Not sure what you mean by the last highlighted sentence.

**Answer to Revisor G Comentário 25:** Please see answer to comment 8. The Pearson/Spearman correlations should not have been used for variables are not normally distributed, amongst other factors. The multiple linear regression model fails on several assumptions, particularly because of non-normality and co-linearity. We have withdrawn the correlations between clinical parameters as well as the model.

***Revisor G Comentário 26:***

RESULT 3. RECALL PATIENT COHORT: HEALTH RELATED QUALITY OF LIFE OUTCOMES:

- “Due to missing values, the SF-36 Physical (PCS) and Mental Component Summaries (PCS) could only be calculated for 44 subjects with mean values of 37,7 ± 9,5 and 46,5 ± 11,6, respectively” -> in Figure 1 many more than 44 individuals are mentioned.

**Answer to Revisor G Comentário 26:** The rules for the summary value calculations do not allow for any missing value. In other words, one missing value in a single domain prevents the calculation. This explains the discrepancy between the number of summary componnents and the total number of patients.

***Revisor G Comentário 27:***

- “There were no significant differences between men and women other than in the health transition domain, where men fared better than women as regards feeling somewhat better than the year before (p=0.02)” -> This p-value refers to what? Not clear…

**Answer to Revisor G Comentário 27:** We haveremoved the sentence and the information from methods. We were advised that due to low numbers in each categories it was not possible to perform statistical tests. This was a comparison between males and females with a somewhat improved/much improved opinion of their health status (9/14 and 20/62, respectively) versus the others, referring to supplementary table VI below, which has also been removed.

**Supplementary Figure VI: Health transition according to sex at recall (in comparison with previous year)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Much improved | Somewhat improved | Approximately equal | A little worse | Much worse |  |
| Sex | Male | Count | 2 | 7 | 4 | 0 | 1 | 14 |
|  | Female | Count | 8 | 12 | 24 | 12 | 6 | 62 |
| Total |  | Count | 10 | 19 | 28 | 12 | 7 | 76 |

 |  |  |
|  |  |  |  |  |
|  |  |  |  |  |  |  |

***Revisor G Comentário 28:***

The organization of information in the several tables sometimes make it difficult to follow the paper.

- “With the exceptions of HAQ disability index and other than a strong correlation between the EQ-5D-3L and EQ-5D-VAS and moderate correlations between these and the MCS and PCS, there were no significant correlations between HR-QoL indices and clinical variables (Supplementary Table IV), with respect to the professional situation, schooling or marital status (Supplementary Table VII)” -> This whole sentence is strange and it is not clear what is meant. Besides, there are some bold values in this table whose meaning is not clear and is not explained.

**Answer to Revisor G Comentário 28:** The sentence has been removed

***Revisor G Comentário 29:***

- “Even though it did not reach statistical significance, unemployed patients had a higher DAS28, HAQ, and number of deformities, with the lowest EQ-5D-3L score.” -> How was the significance evaluated?; “a higher” -> “the highest”.

**Answer to Revisor G Comentário 29:** A significantly lower EQ-5D-3L Portuguese tariff was found in the unemployed (T-test) as shown in Supplementary Table IV. Mann Whitney *U* test test between the two revealed no statistical significant difference in age, DAS28-ESR, HAQ or number of deformities. The text now reads: “Full data as regards professional status are shown in Supplementary Table IV. Unemployed patients (n=8) had a significantly lower EQ-5D Portuguese tariff when compared to employed patients (n=21) (p=0.012, T-test), with a similar median age; even though not significantly, unemployed patients had higher DA28, HAQ and number of deformities in comparison to the latter.”

***Revisor G Comentário 30:***

RESULT 4. DEMOGRAPHIC, CLINICAL AND THERAPEUTIC CHARACTERIZATION OF PATIENTS WHO DIED OR WERE LOST TO FOLLOW-UP

- “The deceased patients had a statistically significant lower estimated 10-year survival rate at the time of first biologic therapy (p=0.006) and discontinued bDMARD at a higher frequency (p=0.028) than the recall patients (Supplementary Table VIII).” -> Confirm that you have used the adequate alternative hypothesis.

**Answer to Revisor G Comentário 30:** As far as the 10-year survival, we are comparing a mean recall 80 with mean dead patients of 54.3±25. Given a z value of 2.62 and a critical value of 0.4953, the p value was 1-0.9953=0.0047, allowing for a rejection of the alternative hypothesis. For the frequency of bDMARD discontinuation, the X2 was 17.8, > than 3.84, which is the critical value for 2 degrees of freedom. I confirm to have rejected the alternative hypothesis.

***Revisor G Comentário 31:***

DISCUSSION:

- “Remission or LDA as determined by DAS28 was found in 51% of patients.” -> Do you present this result?

**Answer to Revisor G Comentário 31:** This result is presented in Table II and results from adding the percentage of patients in remission (26%) with those with LDA (25%).

***Revisor G Comentário 32:***

- “DAS28 was also significantly determined by pre-bDMARD DAS28; uncontrolled disease activity being to steroid use in…” -> “DAS28 was also significantly determined by pre-bDMARD DAS28; uncontrolled disease activity being related to steroid use in…”

**Answer to Revisor G Comentário 32:** Please see answer to comment 8. These correlations were withdrawn.

***Revisor G Comentário 33:***

- The median age of the unemployed patients (57.5 years) was nearing retirement time but it should be noted that despite our efforts, they had the highest disease activity, functional impairment and the lowest EQ-D5-3L score.

-> Do not understand “despite our efforts”. For what?

-> Maybe this is why they are unemployed?

**Answer to Revisor G Comentário 33:** The sentence now reads: “The median age of the unemployed patients (57.5 years) was nearing retirement time but it should be noted that despite bDMARD therapy, they had the highest disease activity, functional impairment and the lowest EQ-D5-3L score.”

***Revisor G Comentário 34:***

- “Interestingly, baseline mean EQ-5D-3L and HAQ values for a Greek population” -> reference?

**Answer to Revisor G Comentário 34:** the reference is now included in the text and below. [5]

**References**

[1] P. L. van Riel, "The development of the disease activity score (DAS) and the disease activity score using 28 joint counts (DAS28)," *Clin Exp Rheumatol,* vol. 32, no. 5 Suppl 85, pp. S-65-74, Sep-Oct 2014. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/25365092>.

[2] D. Aletaha, M. M. Ward, K. P. Machold, V. P. Nell, T. Stamm, and J. S. Smolen, "Remission and active disease in rheumatoid arthritis: defining criteria for disease activity states," *Arthritis Rheum,* vol. 52, no. 9, pp. 2625-36, Sep 2005, doi: 10.1002/art.21235.

[3] P. L. Ferreira, "[Development of the Portuguese version of MOS SF-36. Part I. Cultural and linguistic adaptation]," *Acta Med Port,* vol. 13, no. 1-2, pp. 55-66, Jan-Apr 2000. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/11059056>. Criacao da versao portuguesa do MOS SF-36. Parte I--Adaptacao cultural e linguistica.

[4] P. L. Ferreira, "[Development of the Portuguese version of MOS SF-36. Part II --Validation tests]," *Acta Med Port,* vol. 13, no. 3, pp. 119-27, May-Jun 2000. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/11026151>. Criacao da versao Portuguesa do MOS SF-36. Parte II--Testes de validacao.

[5] D. T. Boumpas, P. Sidiropoulos, L. Settas, P. Szczypa, V. Tsekouras, and A. C. Hernandez Daly, "Health outcomes and unmet needs in patients with long-standing rheumatoid arthritis attending tertiary care in Greece: a cohort study," *Health Qual Life Outcomes,* vol. 17, no. 1, p. 73, Apr 29 2019, doi: 10.1186/s12955-019-1127-8.