AIM OF ASSESSMENT

This assessment aim to investigate if there is an association between baseline disease activity (DAS28_Baseline^{o} and pain progression over time in patients with Rheumatoid arthritis.The VAS records the measure of pain over time, across the patient's visit to the clinic 10 times.

Data Inspection.
After careful data inspections, I was able to observe that the dataset contains of 500 patients which visit the clinic for a period of 10 times, so we have ID of 1-500, and Visit_number from 1-10

VAS is a continuous pain score that is recorded at every visit over time for each patient. It range from 19.34 to 100.00 with a mean of 47, which shows that pain is moderate but varies evenly.

The Age_Baseline and DAS28_Baseline are constant for each patient, they were repeated across their 10 visits. The Age_Baseline range from 32 - 94 years, with the mean of 60.6 years which means that the patients were mostly adults. DAS28_Baseline ranges from 1.35 to 8.62, with a mean of 4.1 which explains that the disease progression are from low to very high disease activity.

There are no missing values in the data set.

EXPLORATIVE DATA ANALYSIS

I conducted an exploratory data analysis by grouping baseline columns and checking the trend in pain score across the time of visits. I created a dataset by selecting the observations from visit 1 for all 500 patients.

As earlier stated, at baseline the mean Age of the participants are 60.6 years with a standard deviation of 10.2 and it ranges from 32.3 - 93.7 years. Baseline disease activity ranges from 1.35 to 8.62, with a mean of 4.10 and a standard deviation of 1.19. Te Baseline pain score when extracted, ranged from 19.3 to 61.5 and have a mean value of 35.3 and a standard devotion of 6.7, suggesting moderate levels of pain at their first clinic visit

PAIN SCORES ACROSS OTHER VISITS

When VAS was summarized across the 10 visits, the mean pain moved up to 47.4 with a standard deviation of 11.1, this shows an higher pain score as the patients came to the clinic after their baseline visit and pain score, DAS28_Baseline remains the same across the visits, with further strengthens the integrity of our data set that these co-variates remain constant for individuals across our dataset.

Overall the dataset displays a well structured and balanced study with 500 patients which each contribute ten repeated pan measurement, with patients showing difference in disease activity and pain, with pain level increasing over time, we see the need to employ the mixed_effect model which is specifically designed to handle this type of situations that requires repeated measurements from the same persons over time.

PAIN TREJECTORY PLOT

I selected a sample of 100 patients from the data set with unique IDs, , and plotted a pain trajectory to show how VAS scores changes across the 10 clinic visits. The Blue line shows each patient and the red line shows the average trend. From the plot, we can see that the pain pattern vary for each patients vary but most patient experience an overall increase across different visits, and the average pain trajectory also increases across visits.

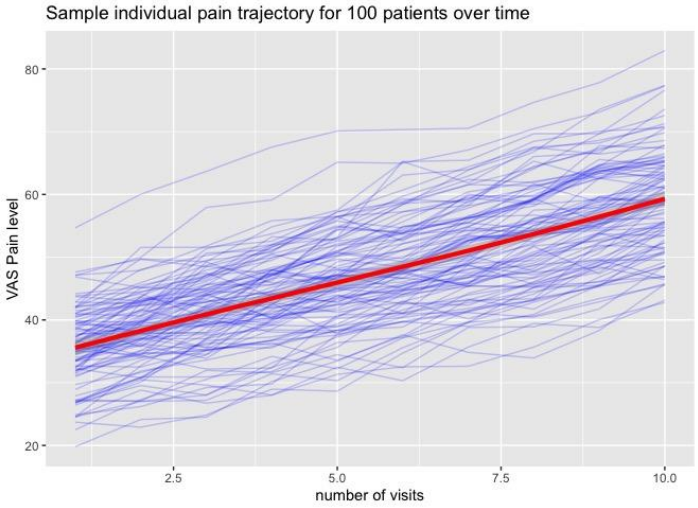


Figure 1

DAS GROUPING EXPLORATION

The DAS28_Baseline was categorised into three groups of (low, medium, high) disease activity based on the tertiles of its distribution. This was achieved by dividing the DAS28_Baseline values into three equal sized groups using the 0, 1/3, 2/3 and 1 points. Each of the patients was assigned to the following categories;

Low_DAS28, Medium_DAS28, High_DAS28.

This categorisation allows for the comparison of pain across different levels of baseline disease activity to help in visual exploration to know if the pain varies by baseline disease activity.

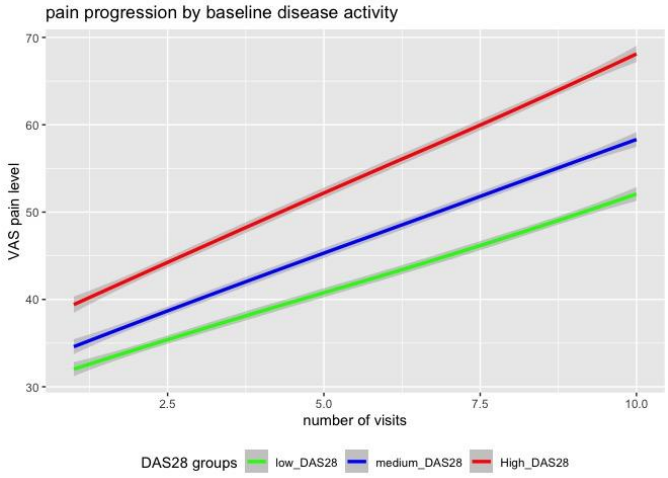


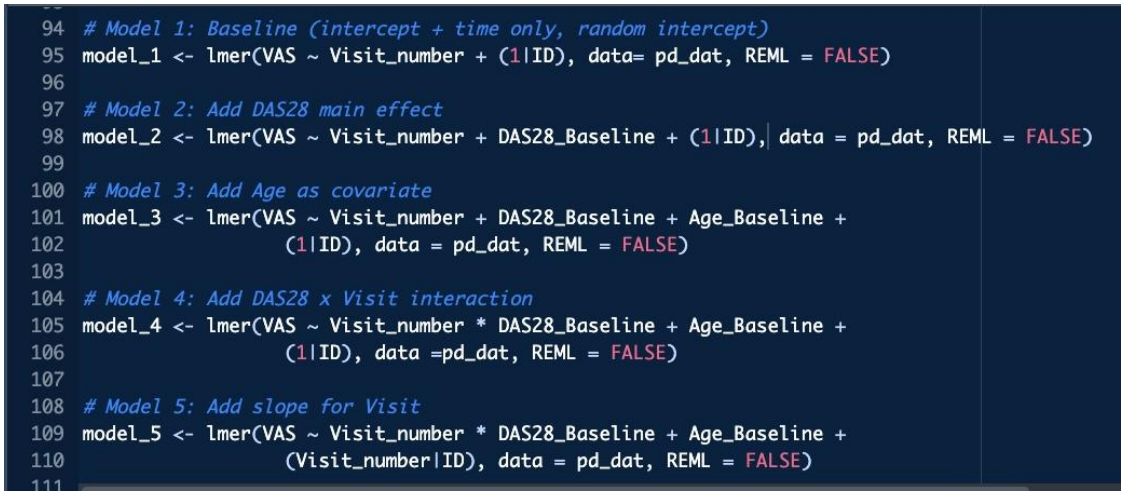
Figure 2

When each grouped baseline was plotted against VAS pain score. Across all groups, although the groups differ consistently in their pain levels, we can still see that the pain level increases steadily over time, which idicate a general rise in reported pain.

The clear separation of the groups help us visually to see that there is clear positive assiciation of DAS28_Baseline and VAS, where we can clearly see that the higher the disease activity, the higher the pain_level and pain progression across time.

STATISTICAL MODELLING

A series of linear mixed-effects models were fitted, using VAS pain score as the outcome, and Visit_number as the time variable, including random effects for patient ID to check for all repeated measurements.



```
# Model 1: Baseline (intercept + time only, random intercept)
model_1 <- lmer(VAS ~ Visit_number + (1|ID), data= pd_dat, REML = FALSE)

# Model 2: Add DAS28 main effect
model_2 <- lmer(VAS ~ Visit_number + DAS28_Baseline + (1|ID), data = pd_dat, REML = FALSE)

# Model 3: Add Age as covariate
model_3 <- lmer(VAS ~ Visit_number + DAS28_Baseline + Age_Baseline +
                  (1|ID), data = pd_dat, REML = FALSE)

# Model 4: Add DAS28 x Visit interaction
model_4 <- lmer(VAS ~ Visit_number * DAS28_Baseline + Age_Baseline +
                  (1|ID), data =pd_dat, REML = FALSE)

# Model 5: Add slope for Visit
model_5 <- lmer(VAS ~ Visit_number * DAS28_Baseline + Age_Baseline +
                  (Visit_number|ID), data = pd_dat, REML = FALSE)
```

Figure 3

Model fit was compared using AIC and BIC, And Model 5 has a very substantial lower AIC (24311.9) compared to Model 4 which has (24681.3) and BIC, of which the Model 5 shows a better fit, so I selected Model 5 with random intercept and slope for Visit_number as the final Model.

RESULT INTERPRETATION FOR FINAL MODEL.

From Model 5

1. Visit_number (estimate ≈ 1.15, p < 0.001):
This means that on a average pain increases over time, the positive coefficient shows that with a given baseline DAS28 and Age, VAS rises as the number of visit increases.

2. DAS28_Baseline (estimate ≈ 2.80, p < 0.001):
This indicates that patients who has higher baseline disease activity, reports higher pain levels, with an estimate of a unit increase in DAS28 level, there will be an increase in VAS by 2.80 points, at the given time and Age.

3. Age_Baseline (estimate ≈ 0.35, p < 0.001):
This findings shows that Older patients get to report higher pain level, Which means that with each additional year of age, the VAS pain level increases by 0.35 points while we maintain every other variable as constant.

4. Visit_number × DAS28_Baseline interaction (estimate ≈ 0.37, p < 0.001):
The positive interaction shows that patients with higher DAS28_Baseline does not just only start with higher pain, but their pain also increases in VAS pain score faster over time.

RANDOM EFFECTS
Model 5 includes
Random intercept variance is approximately 16.6, while standard deviation is 4.1; Which mean that there are meaningful difference between patients in their overall pain levels

Random slope variance for Visit_number is approximately 0.12 and with a standard deviation of 0.35; which means that patients differ in how quickly their pain change over time.

Negative correlation is approximately -0.41 between the intercept and slope; This indicates that patients who has higher initial pain on average tend to have slower rate of increase and vice versa

CONCLUSION ON MODELLING

Higher baseline DAS28 is strongly associated with higher pain with faster progression over time.

MODEL DIAGNOSIS

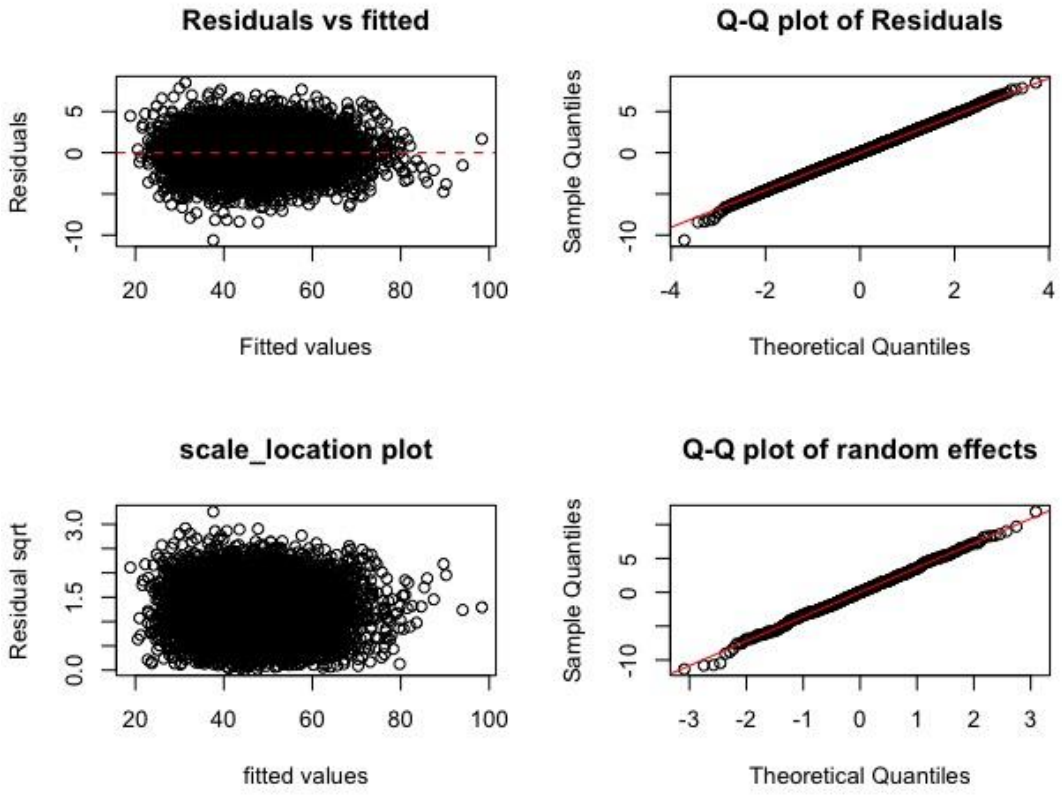


Figure 4 Diagnostic plots for Final Mixed-Effect Model.

To check whether the assumptions of the linear mixed effects model are satisfied, the model was assessed using residual and random effects plots for the final mixed-effects model. The residuals vs fitted plot showed a random cloud of points around zero with no pattern which means that the assumed relationship and constant are reasonable.

The Q-Q plot of residuals shows that the residuals followed the theoretical line of normalcy closely, with just deviations at the extremes, which is still considered reasonable ad consistent with approximately normality.

The scale-location plot showed a constant spread of the square-root residuals through the range of the fitted values. This provides more support for homoscedasticity, And the Q-Q plot of the random intercept for ID shows that points lying close to the line of reference, which indicates that the random effects are also normally distributed.

The diagnostics suggest that the model assumptions are adequately met.

Predictor	Estimate	Std. error	95%CI (lower)	95% CI (higher)	T-value	P-value
intercept	0.05	1.248	-2.39	2.50	0.04	0.967
Visit_number	1.15	0.068	1.02	1.28	16.98	<0.001
DAS28_baseline	2.80	0.163	2.48	3.12	17.13	<0.001
Age_Baseline	0.35	0.017	0.32	0.38	21.12	<0.001
Visit no x DAS28_Baseline	0.37	0.016	0.34	0.40	23.25	<0.001

Table 1 Fixed effects from final mixed-effects model.

DAS28 Group	Start VAS	End VAS	Total Change
Low DAS28	31.0	50.3	19.3
Median DAS28	34.9	58.4	23.5
High DAS28	40.0	68.9	28.8

Table 2 Pain Progression across levels of baseline DAS28.

Patients with higher baseline DAS28 scores start with higher pain levels and experience a greater overall increase in pain over the 10 visits.

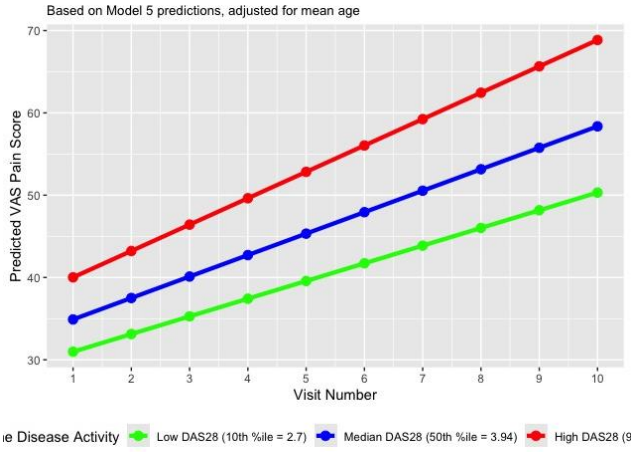


Figure 5 Predicted pain trajectory by DAS28_Baseline groups

This plot further butteress my finfing in the EDA, under figure 2, This plot shows that pain rises faster in the high-DAS28 group. Which strongly suggests that the baseline disease activity strongly influences the trend of pain over time.

CONCLUSION

This analysis aimed to checkwether the baseline disease activity (DAS28) predicts the pain progression over the patient's visits to the clinic over 10 times among patients with rheumatoid arthritis. The exploratory analysis shows that patients with higher DAS28 scores has the tendency to report higher pain levels throughout the consequent follow ups from the baseline clinic visit. The final linear mixed-effect model further reinforce a strong positive association between the baseline DAS28 and both the level and rate of increase in pain scores over time of clinical visits, even after we made accommodation for age, and repeated measurements, the Model diagnostics indicated that all assumptions were adequately met, which supports how reliable are our findings.

Lastly the results proposes that patients who has a higher disease activity at the baseline eventually experience more pain trajectory at a rapid rate, buttressing the importance of early identification and management of high DAS28 in public health and Clinical sessions.