Question 1

A) The brief statistical analysis plan:

1) a consort diagram:

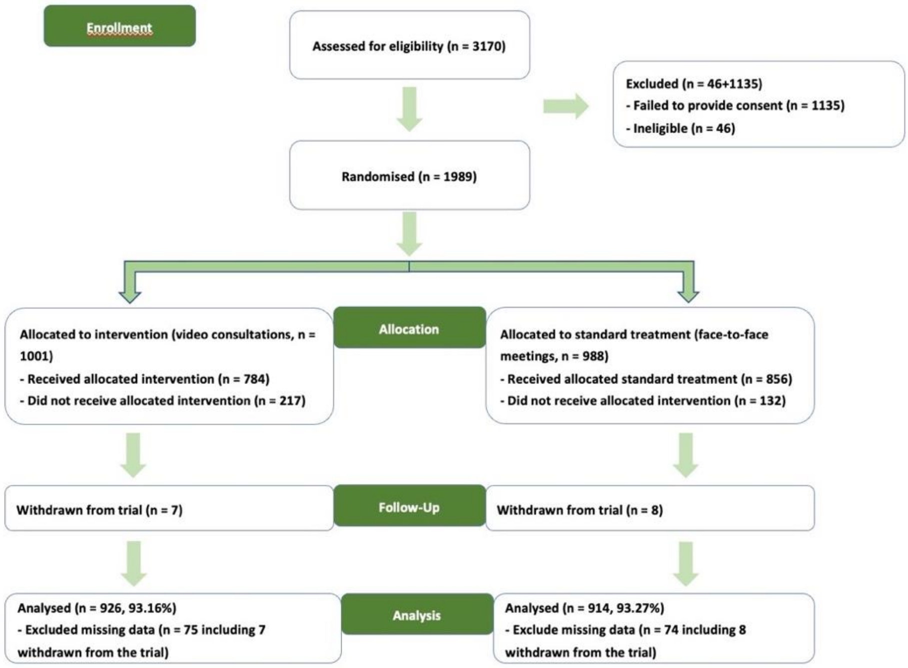

Figure 1. consort diagram for the trial

The consort diagram above describes the flow of the randomised trial, investigating the usefulness of the video consultation compared to usual hospital appointments.

There are four stages of the trial: enrollment of patients, allocation to groups, follow-up and analysis of primary outcome.

Note that there are 149 missing data, with 15 attributed to withdrawals from the trial and the rest with unstated reasons.

2) a table of baseline characteristics:

Table 1. table of baseline characteristics of the trial

Baseline characteristics		**Randomised group***	
variable	**level**	**Standard (N = 988)**	**Video_consult (N = 1001)**
site	London	586(59.31%)	586(58.43%)
	Shrewsbury	402(40.69%)	415(41.6%)
specgrp	Orthopaedics	179(18.12%)	173(17.28%)
	Urology	144(14.57%)	134(13.39%)
	ENT	274(27.73%)	281(28.07%)
	Gastroenterology	206(20.85%)	206(20.58%)
	Other	185(18.72%)	207(20,68%)
gender	Male	486(49.19%)	479(47.85%)
	Female	502(50.81%)	522(52.15%)
ethngrp**	White	790(88.17%)	809(89.89%)

	Black Caribbean/African	33(3.68%)	16(1.78%)
	Indian/Pakistani/Bangladeshi	29(3.24%)	22(2.44%)
	Chinese	4(0.45%)	5(0.56%)
	Other Asian	12(1.34%)	13(1.44%)
	Other	28(3.12%)	35(3.89%)
appointment**	No appointment	540(59.08%)	445(48.06%)
	Appointment	374(40.92%)	481(51.94%)
withdrawn	No	980(99.19%)	994(99.3%)
	Yes	8(0.81%)	7(0.7%)
compliance	Noncompliant	132(13.36%)	217(21.68%)
	Compliant	856(86.64%)	784(78.32%)
age**	(in years)	47.96(20.68)	48.40(20.74)
dusoi**	(in score)	49.54(18.20)	48.22(17.54)
sf12phy0**	(in score)	41.23(11.86)	42.14(11.57)

(to 2 decimal places)
* Data shown are Mean(SD) for categorical variables and Number(%) for continuous variables (Age, dusoi, sf12phy0)
** there are missing data in variables

The above table describes the baseline characteristics of the randomized trial we investigate, including the name of the variable, the level of categorical variables, and their characteristics (mean and standard deviation, or number and percentage) in each randomized group.

We could see that those who used video consultation during the trial tended to make a follow-up appointment, with a percentage of 51.94%; for patients who were allocated to the standard treatment group, the figure is only 40.92%.

3) An unadjusted (complete case, ITT) analysis

Based on the mean difference between the standard treatment group and the video consultation group in Table 2, there is 0.11 more follow-up appointment number in the video consultation group than that in the standard treatment group.

T-test		
Group	**Mean**	**95% Confidence Interval**
Standard Meeting	0.41	0.38 to 0.44
Video Meeting	0.52	0.49 to 0.55
Difference	-0.11	-0.16 to -0.06
P-value: <0.001		

Table 2. a table of statistics of student's t-test between randomised groups

Based on the odds ratio from the logistic regression (see Table 3), the odds of the video consultation group having a follow-up appointment is 1.56 times the odds in the standard consult group.

Logistic Regression Model - Unadjusted			
Appointment	**Odds Ratio**	**95% Confidence Interval**	**P-Value**
Random Group	1.56	1.30 to 1.88	<0.001
Baseline	0.69	0.61 to 0.79	<0.001

Table 3. a table of statistics from an unadjusted logistic regression model

According to Table 4, the risk of further appointment in the video consultation group is 11% more than the risk in the standard treatment group and is 1.27 times the risk in the standard treatment group.

Risk		Risk Difference	Risk Ratio	P-value
Exposed	Unexposed			
0.52	0.41	0.11	1.27	<0.01

Table 4. a risk summary table

Overall, we could conclude from the unadjusted ITT analyses for the complete case that video consultation seems to increase the number of further appointments.

4) An adjusted analysis as required

For variables randgrp, age, and specgrp, the p-values are less than 0.05 and 95% confidence intervals for odds ratio exclude 1, suggesting significant evidence of their effect on the number of follow-up appointments.

The odds of subsequent appointments for the video consultation group is 1.56 times the odds in the standard treatment group, and the odds of subsequent appointments is increased by 0.6% for a one-year increase in age.

The p-values for variables gender and site are larger than 0.05 significance level and the 95% confidence interval for odds ratio includes 1, suggesting that there is no evidence to show that gender and site influence the number of follow-up appointments, though the p-value for site is small.

The specialty group is also a statistically significant variable in the logistic regression model (p-value = 0.029), suggesting that different medical specialties might have an effect on the number of further appointments.

Appointment	Odds Ratio	95% Confidence Interval	P-Value
Random Group	1.56	1.30 to 1.88	< 0.001
Site	0.82	0.67 to 1.00	0.057
Gender	0.99	0.82 to 1.19	0.881
Age	1.00	1.00 to 1.01	0.012
Specialty	1.09	1.00 to 1.17	0.029
Baseline	0.44	0.31 to 0.63	< 0.001

Table 5. a table of statistics from an adjusted logistic regression model

5) An investigation of variable site as required

For subgroup Shrewsbury, the 95% confidence interval for odds ratio (OR = 2.12) excludes 1 and the p-value is small, indicating that there is strong evidence that the follow-up appointment number for the video consult group in Shrewsbury is higher.

For subgroup London, the 95% confidence interval for odds ratio (OR = 1.26) includes 1 and the p-value is larger than 0.5, implying little evidence that the video consult group has an effect on the number of follow-up appointments.

The non-overlap in the 95% confidence intervals for the odds ratio between London and Shrewsbury, along with the differences in odds ratios between the overall cases (OR = 1.56) in both

sites and those in the site subgroups (OR = 1.26 & OR = 2.12), all suggest that there is sufficient evidence of a subgroup difference.
By introducing an interaction term between site and randgrp in the logistic regression model, the p-value for the interaction is small (p = 0.007), indicating strong evidence that the treatment effect for the primary outcome differs between study sites.

Model	Odds Ratio(95% CI)	P-Value
All	1.56 (1.30 to 1.88)	< 0.01
site = London	1.26(0.99 to 1.61)	0.055
Site = Shrewsbury	2.12 (1.59 to 2.84)	< 0.01

Table 6. a table of statistics of regression between appointment and randgrp, stratified by site

Logistic Regression Model – Interaction			
Variables	**Odds Ratio**	**95% Confidence Interval**	**P-Value**
Shrewsbury	0.68	0.52 to 0.89	0.005
Video Consultation	1.26	0.99 to 1.60	0.055
Shrewsbury & Video Consultation	1.68	1.15 to 2.44	0.007
Baseline	0.81	0.68 to 0.96	0.014

Table 7. a table of statistics from an adjusted logistic regression model with interaction

B) See Appendix.

C)

Based on the analysis in part A, it is found site, age and specialty all have an effect on the number of follow-up appointment. There is strong evidence that the video consultation group has a higher number of follow-up appointments, and this is not what we expected for the video consultation intervention. The effect of online intervention is different between London and Shrewsbury, and video consultation in Shrewsbury is less effective than in London, where patients with video consultation are more likely to book a further appointment compared to London.

D)

Problem 1:
Several patients were not compliant with the allocated treatment (video consultation or standard face-to-face consultation). This might wrongly estimate the efficacy of video consultation, though it will provide us with the information on effectiveness of offering treatment.

We could further investigate the problem of non-compliance by carrying out pre-protocol analysis or as-treated analysis to consider treatment efficacy. We could exclude non-compliant patients in each randomised group (PPA analysis), or we could compare them by the actual treatment they take (ATA analysis) to review the usefulness of video consultation.

Problem 2:
There are several patients who withdrew from the trial during the follow-up stage. Also, there are some missing data for unknown reasons. This may lead to a wrong estimation of treatment efficacy due to unknown characteristics of subjects with missing data. For example, all patients with missing

data do not wish to make an appointment, which would greatly influence our analyses and should be taken into consideration.

Sensitivity analyses could be used to investigate and address the problem. We could impute average / worst case / extreme scenarios to provide some insights on the cause of missing data or make estimations of the missing value based on the reason for missingness or patient characteristics.

E)

Alternatively, we could perform a survival analysis given the nature of additional data.

We could calculate the rate of making follow-up appointments to provide insight.
To further investigate, the estimated probabilities of not making further appointments could be calculated via survival analysis. To do this, a new variable called time is considered to represent the time from the exact date of patients' initial consultation until the date of follow-up consultation (if there is any).

The data is censored because there are patients who did not make the follow-up appointment or withdrew from the trial during the follow-up stage. Therefore, we could use Kaplan-Meier estimator of the survival function to calculate the estimated survival probabilities.

Note that for K-M models, we assume that censoring is non-informative, survival times are independent of each other, and there are no competing risks that preclude the occurrence of primary events of interest.

Question 2

A)

To define a variable as a confounder in the association between exposure and outcome, we need to identify three criteria:

1. The variable is a risk factor for the outcome independently of the exposure,
2. It is associated with the exposure,
3. It is not causally affected by the exposure or the outcome.

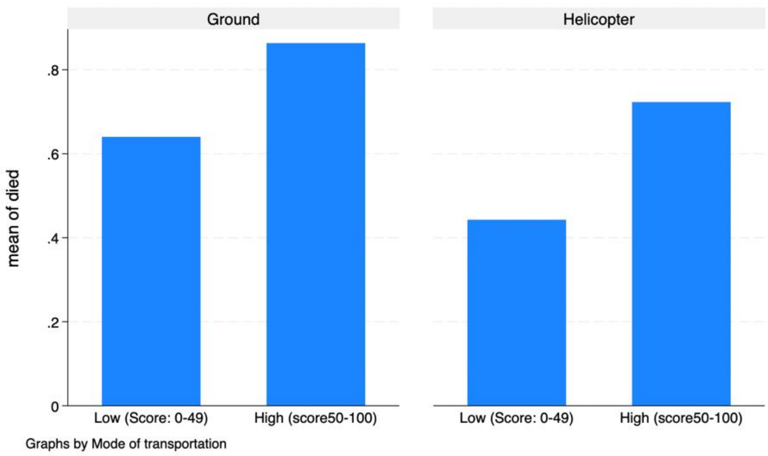

Figure 2. Bar chart: the mean mortality of severity of injury grouped by transportation

Risk of Death				Risk Ratio	
Low		High		Low Severity	High Severity
Helicopter	Ground	Helicopter	Ground	0.69	0.84
0.44	0.64	0.72	0.86		

Table 8. a risk summary table of the trial

By examining the risk ratios for patients with low severity injury and high severity injury on each mode of transportation, it is clear that the group of high severity injury exhibits a higher risk of dying than the one of low severity, regardless transported by ground or helicopter. It means that the severity score is a risk factor for death, independent of the mode of transportation.

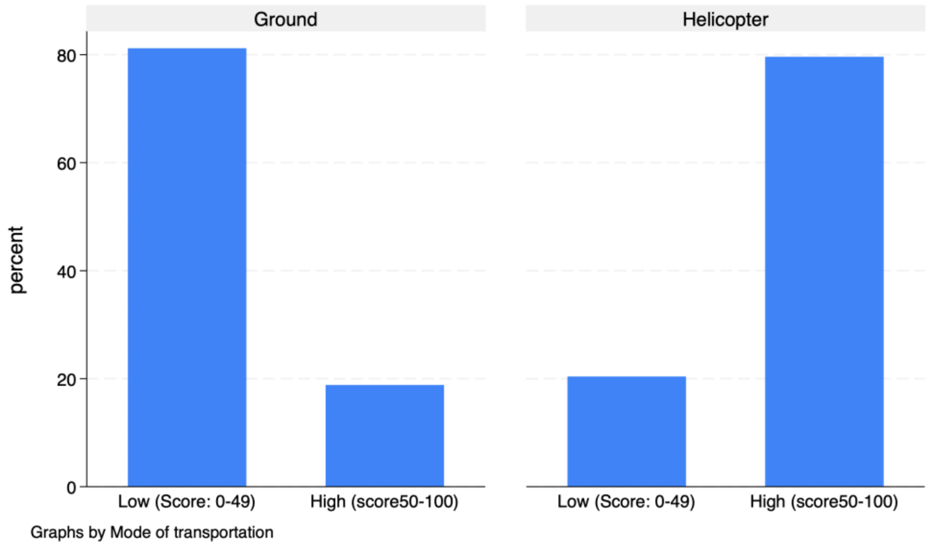

Figure 3. Bar chart: the distribution of patients with different level of severity score on each mode of transportation

The variable severity is not balanced across each mode of transportation (shown in Figure 3). There appears to exist an association between severity and mode of transportation, where patients with higher severity score tend to be carried with helicopters.

Moreover, judging with common sense, the severity score is determined by injury, and it cannot be affected by mode of transportation, or by in-hospital death.

All three criteria for a variable to be a confounder are satisfied. Therefore, it is likely that severity confounds the association between mode of transportation and death.

Next, we perform similar analysis to the variable ‘age’:

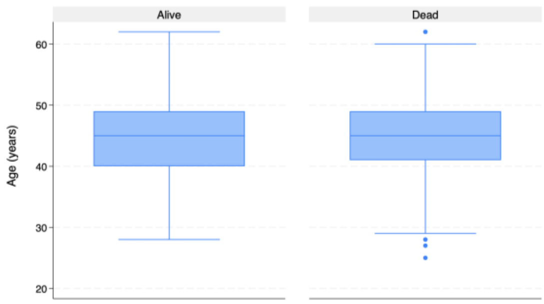

Figure 4. boxplot: age vs died

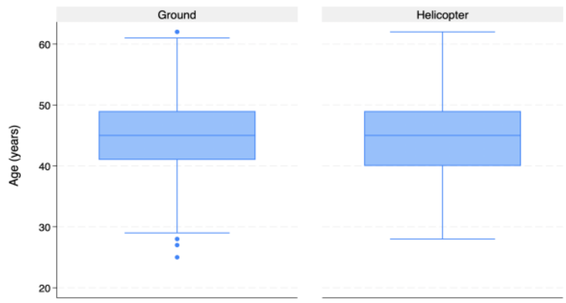

Figure 5. Boxplot: age vs transportation

Looking at Figure 4, the distribution of age is similar among two groups, thus the factor age is not likely associated with in-hospital death.

Likewise, there is no clear difference in age between modes of transportation, as demonstrated in Figure 5. Age is not likely associated with the mode of transportation.

By common sense, age is determined by birth date so cannot be affected by mode of transportation or death, and there is no causal effect between age and the exposure or outcome of the trial. Criterion 3 is met.
To conclude, because the first and second criteria of confounders are not complied with, it's not likely that the variable age confounds the association between mode of transportation and death.

B)

According to the exploratory analyses where we conclude that the severity score of injury is a confounder of the trial, we build a logistic regression model between the outcome (died) and the exposure (mode), adjusting for severity as a confounder.

Logistic regression model including severity		
Died	**Coefficient (95% CI)**	**P-Value**
Mode	-0.848 (-1.20 to -0.49)	< 0.001
Severity	1.24 (0.90 to 1.58)	< 0.001
Baseline	0.578 (0.45 to 0.71)	< 0.001

Table 9. a table of statistics from an adjusted logistic regression model with severity

We could see that the p-values for both variables (mode and severity) are less than 0.01, indicating that there exists strong evidence that there are some relationships between mode and death in the model after adjusting for severity as confounder. It aligns with our conclusion from previous exploratory analysis.

<table>
<tr><th colspan="4">Predicted Risk of Death Formula: $\log\left(\frac{\pi}{1-\pi}\right) = 0.578 - 0.848 \times mode + 1.24 \times severity$</th></tr>
<tr><td colspan="2"></td><td colspan="2">Severity</td></tr>
<tr><td colspan="2"></td><td>Low</td><td>High</td></tr>
<tr><td rowspan="2">Mode of Transportation</td><td>Helicopter</td><td>0.43</td><td>0.73</td></tr>
<tr><td>Ground Ambulance</td><td>0.64</td><td>0.86</td></tr>
</table>

Table 10. a table of predicted risk of death from the logistic model

By looking at table 10, death dedicators are 0.21 and 0.13 lower in helicopter than ground ambulance in low and high severity groups respectively.

We conclude that patients with high severity of injury transported by helicopter have less risk of death than that of ground ambulance.

C) See Appendix.

Question 3

A)

The table below shows that the average scores with and without acupuncture were 16.2 and 21.3 at 12 months respectively. 95% confidence intervals for both groups exclude 0. P-value = 0.002 is significant at a 5% significance level, suggesting that acupuncture reduced the headache severity score.

Group	Mean	95% Confidence Interval	
No Acupuncture	21.3	18.5	24.2
Acupuncture	16.2	14.0	18.3
Difference	5.18	1.67	8.69

P-Value for difference >0: 0.002

P-Value for difference exists: 0.004

Table 11. a table of statistics of student's t-test between intervention groups

The colleague intended to compare the average difference between the scores at randomisation and at 12 months. According to the second table, the headache severity score after acupuncture therapy had reduced by 8.41, which reduced 2.17 more than without acupuncture after 12 months. 95% confidence intervals for both groups include 0, and the p-value for average differences is 0.1119. Hence, the effect of acupuncture on headache severity change is likely insignificant.

Group	Mean	95% Confidence Interval	
No Acupuncture	-6.24	-7.98	-4.51
Acupuncture	-8.41	-10.40	-6.40
Difference	2.17	-0.51	4.85

P-Value for difference exists: 0.1119

Table 12. a table of statistics from student's t-test of average score differences between 2 groups

B)

The table below represents the headache severity scores at randomization in 2 groups. It is observed that the initial average score in the acupuncture group had already been less than the score in the no acupuncture group by 3. This could be the consideration of an overevaluation of the effect of the acupuncture.

	No Acupuncture	Acupuncture
Variable hs0	27.6	24.6

Table 13. Mean headache severity scores at randomisation

The correlation between the baseline hs0 scores and follow-up hs12 scores is 0.71 which is high. Since the baseline scores were considered sufficiently unbalanced, we could carry out an ANCOVA analysis for headache severity scores at randomization to improve the precision of therapy effect analysis.

ANCOVA model				
hs12	**Coefficient**	**95% Confidence Interval**		**P-Value**
hs0	0.71	0.63	0.79	<0.001
group	-3.06	-5.54	-0.57	0.016

Table 14. a table of statistics in ANCOVA model

	Model	**Headache severity score at 12 months**		
Type of analysis		**Treatment effect**	**95% confidence interval**	**p-value**
Unadjusted	$E(Y_{i12\ month}) = \beta_0 + \beta_1 group_i$	-5.17	(-8.69, -1.67)	0.004
Average Differences	$E(Y_{i12\ month} - Y_{i randomisation}) = \beta_0 + \beta_1 group_i$	-2.17	(-4.85, 0.51)	0.112
ANCOVA	$E(Y_{i12\ month}) = \beta_0 + \beta_1 group_i + \beta_2 Y_{i randomisation}$	-3.06	(-5.54, -0.57)	0.016

Table 15. Comparison of treatment effect of acupuncture estimated from different methods.

We can get the conclusion that: there is good evidence that acupuncture therapy is effective in reducing headache.

With unadjusted analysis, though it is statistically significant, it has a comparatively wider confidence interval and does not account for the imbalance between baseline groups.

Regarding the analysis of change, the 95% confidence interval includes 0 and the p-value is 0.11, so it shows little evidence that acupuncture is effective. Also, since the baseline and the change in scores are correlated, the average difference analysis combines the therapy effect and the effect of regression to the mean. Therefore, it underestimates the acupuncture therapy effect. Additionally, the baseline imbalance could not be accounted for in the average difference analysis.

A narrower confidence interval could be observed under ANCOVA analysis and its average reduction of pain score after 12 months is smaller than that in the unadjusted analysis. This is resulted from the imbalance in baseline pain score (acupuncture group had already had a lower pain score before any treatment), ANCOVA provides a treatment effect adjusted for baseline imbalance which leads to a smaller reduction of pain.

In summary, ANCOVA analysis could adjust for the baseline imbalance and result in a narrower confidence interval with a statistically significant p-value, so ANCOVA is the preferred analysis approach.

C) See Appendix.

Appendix

```
* Question 1

* import dataset
use dataset_q1_J.dta, clear

* A.

* 1)
* count for missing data to be used in consort diagram
count if missing(appointment)

* 2). tables for baseline characteristics
* 2-way contigency tables for categorical variables
tab randgrp
tab withdrawn randgrp, col
tab compliance randgrp, col
tab appointment randgrp, col
tab site randgrp, col
tab gender randgrp, col
tab ethngrp randgrp, col
tab specgrp randgrp, col
* summary tables for continuous variables
bysort randgrp: sum age
bysort randgrp: sum dusoi
bysort randgrp: sum sf12phy0


* 3) an unadjusted, complete case ITT analysis
* a student's t-test between randgrp
ttest appointment, by(randgrp)
* an unadjusted logistic regression
logistic appointment randgrp
* table providing risk difference and risk ratio
cs appointment randgrp


* 4) adjusted analysis (logistic regression model)
logistic appointment randgrp site gender age specgrp


* 5) investigation of the variable site
* logistic regression for London
logistic appointment randgrp if site == 0
* logistic regression for Shrewsbury
logistic appointment randgrp if site == 1
* logistic regression with interaction term
logistic appointment site##randgrp
```

```
* Question 2
* import data
use dataset_q2_J.dta, clear

* A.

* the table for mode and died, each for severity
bysort severity: tab died mode
* figures in calculations of risk below are all extracted from the table above
* calculated risk of death for low as severity, ground ambulance as transportation mode
di 585/914
* calculated risk of death for high as severity, ground ambulance as transportation mode
di 183/212
* calculated risk of death for low as severity, helicopter as transportation mode
di 27/61
* the calculated risk of death for high as severity, helicopter as transportation mode
di 172/238
* show the risk ratio in each stratum
cs died mode, by(severity)

* confounder test for criteria 1: is severity a risk factor for died, independently of mode?
graph bar died, over(severity) by(mode)

* confounder test for criteria 2: is severity associated with mode of transportation?
graph bar, over(severity) by(mode)

* confounder test for criteria 1: is age a risk factor for died, independently of mode?
graph box age, by(died)

* confounder test for criteria 2: is age associated with mode of transportation?
graph box age, by(mode)

* B.

* adjusted logistic regression model with severity as confounder
logistic died mode severity, coef

* figures in calculations of predicted risk below are from logistic regression model above
* the predicted risk of death for high as severity, helicopter as transportation mode
di 1/(1+exp(-(0.578-0.848+1.24)))
* the predicted risk of death for high as severity, ground ambulence as transportation mode
di 1/(1+exp(-(0.578+1.24)))
* the predicted risk of death for low as severity, ground helicopter as transportation mode
di 1/(1+exp(-(0.578-0.848)))
* the predicted risk of death for low as severity, ground ambulence as transportation mode
di 1/(1+exp(-(0.578)))
```

```
* Question 3
* A.
* Introducing data
use dataset_q3_J.dta, clear

* Applying t-test for the average headache severity scores at 12 months between 2 groups
ttest hs12, by(group)

* Differences between headache severity scores at randomization and at 12 months in 2 groups
generate hs_diff = hs12 - hs0

* t-test for average differences between 2 groups
ttest hs_diff, by(group)

* B.
* Determining whether baseline data in 2 groups are balanced or not
bysort group: sum hs0

* Determining the efficiency of regression models
corr hs0 hs12

* ANCOVA analysis adjusted for scores at randomisation
regress hs12 hs0 group
```

Declaration of ownership:

We agree that we contribute to this assessment equally.

Sign : 21073274 20024555 21073015

We are aware of the UCL Statistical Science Department's regulations on plagiarism for assessed coursework. We have read the guidelines in the student handbook and understand what constitutes plagiarism. By submitting the ICA report as detailed in the text of the ICA, we hereby affirm that the work has entirely been carried out by us.