A clinical trial was done of a new drug that is intended to reduce blood pressure in patients with moderately high blood pressure. All subjects involved in the study had average systolic blood pressure between 135 and 150. Subjects were eligible for the study only if they were between the ages of 18 and 70, and were in good health at the start of the study. 600 subjects meeting these criteria were recruited, of whom 330 were men and 270 were women. The body mass index (BMI, equal to weight in kilograms divided height in metres squared) of each subject was also measured at the start of the study. The distribution of sex, age, and BMI among the subjects is shown by the scatterplot below, with sex given by the type of dot (black=male).

The 600 subjects were randomly divided into two groups of 300, with one group (the treatment group) taking the new drug, while the other group (the control group) took an inert placebo. Both groups also received standard advice on how to control blood pressure through diet. The trial was double-blind, with neither the subject nor the physicians measuring blood pressure aware of which subjects were in the control group and which were in the treatment group. The drug has very few side effects, so it is unlikely that a subject would be able to tell whether they were receiving the drug on that basis.

Average systolic blood pressure was measured for each subject at the start of the trial, and at
the end of the trial, a year later. The change in blood pressure from the start of the trial to the end was considered to be the response variable of interest.

The investigators for this trial first analysed the results using a two-sample $t$ test. Variances were not assumed to be the same for the two groups. The R output for this test was as follows:

```r
> t.test(bp.change[treat==1], bp.change[treat==0], var.equal=FALSE)

Welch Two Sample t-test

data: bp.change[treat == 1] and bp.change[treat == 0]
t = -6.7772, df = 468.415, p-value = 3.694e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.5882275 -0.8742412
sample estimates:
mean of x mean of y
-2.790328 -1.559094
```

The histograms of change in blood pressure for the control and treatment groups are as follows:

```
bp.change[treat == 0]
Frequency
-15 20 0 5

bp.change[treat == 1]
Frequency
-15 20 40
```

The investigators then performed a regression analysis, in which the covariates were an indicator of whether the subject was in the treatment group, blood pressure at the start of the study, sex, age, and BMI. The R output for this regression was as follows:

```r
> summary(lm(bp.change ~ treat + bp.start + sex + age + BMI))

Residuals:
Min 1Q Median 3Q Max
-10.14558 -1.21033 0.08628 1.33864 6.35751

Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.619579 3.144273 -1.151 0.25013
 treat -1.232977 0.181798 -6.782 2.87e-11 ***
 bp.start 0.020732 0.021303 0.973 0.33085
 sex -0.577492 0.181664 -3.179 0.00156 **
 age -0.010344 0.007070 -1.463 0.14397
 BMI -0.003705 0.025783 -0.144 0.88577

---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.206 on 594 degrees of freedom
Multiple R-Squared: 0.09351, Adjusted R-squared: 0.08588
F-statistic: 12.26 on 5 and 594 DF, p-value: 2.525e-11
```
When the investigators (Jones, Smith, and Zhang) reported the results of the study, their conclusions were as follows:

This study shows that the drug reduces blood pressure in subjects with moderately high blood pressure by more than a placebo ($t$ test, $p < 0.0001$). A regression analysis confirms this, and also reveals that the drug is more effective in males than in females ($p = 0.0016$), with the estimated average reduction in blood pressure attributable to the drug being 1.8 for males and 1.2 for females. The drug appears to be equally effective for subjects with any initial blood pressure, within the range of 135 to 150 that was examined in this study. The effect of the drug appears to be the same for subjects of any age (in the range 18 to 70 examined in this study) and for subjects with any body mass index (BMI).

Answer the following questions:

a) Are the above conclusions of Jones, Smith, and Zhang justified by the analysis they performed? To the extent that they are, indicate why; to the extent they are not justified, indicate why not. Are there any other conclusions of importance that can be or might be drawn from this data? Describe any additional tests that you think should be done, either to provide justification for these conclusions, or to look at other aspects of the data.

b) Shortly after the study described above was released, the result of another study that involved only men age 20 to 30 was reported. This study tested the same drug, and also recruited only subjects with blood pressure at the start of the study between 135 and 150. There were 40 subjects, who were randomly assigned to treatment and control groups (20 each). A two-sample $t$ test was done, with the output being as follows:

```
Welch Two Sample t-test

data: bp.change[treat == 1] and bp.change[treat == 0]
t = -1.1689, df = 33.646, p-value = 0.2506
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:  
-1.7252513 0.4655764
sample estimates: 
mean of x mean of y 
-2.547342 -1.917504
```

The conclusions of the authors of this study were as follows:

Our results contradict the recent clinical trial by Jones, Smith, and Zhang, which reported an average reduction in blood pressure of 1.8 for male subjects. In contrast, we see no statistically significant effect of the drug. Accordingly this drug should not be routinely used at present, pending further studies to resolve this contradiction.

Comment on the validity of these conclusions, and suggest what further analysis (if any) of data from either study would be desirable.

c) A third study was done later. This study involved only women age 20 to 30. It tested the same drug, and also recruited only subjects with blood pressure at the start of the study between 135 and 150. There were 200 subjects, who were randomly assigned to treatment and control
groups (100 each). A two-sample $t$ test was done, with the output being as follows:

Welch Two Sample t-test

data: bp.change[treat == 1] and bp.change[treat == 0]
t = 0.6576, df = 197.709, p-value = 0.5116
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.3000643 0.6003030
sample estimates:
mean of x mean of y
-1.633816 -1.783936

Write a paragraph drawing conclusions from this study (in the context of the two previous studies), and suggesting any further analysis or experimentation that you believe would be useful.