Renal disease: the goal of the Swiss Analgesic Study was to assess the effect of taking phenacetin-

containing analgesics on kidney function and other health parameters. A group of 624 women were

identified near workplaces from Basel, Switzerland, with high intake of phenacetin-containing

analgesics. This constituted the “study” group. In addition, a control group of 626 women were

identified, from the same workplaces and with normal N-acetyl-P-aminophenyl (NAPAP) levels, who

were presumed to have low or no phenacetin intake. The urine NAPAP level was used as a marker of

recent phenacetin intake. The study group was then subdivided into high-NAPAP (group 1) and low-

NAPAP (group 2) subgroups according to the absolute NAPAP level. However, both subgroups had

higher NAPAP levels than the control group (group 3). The women were examined at baseline during

1967 and 1968, and also in 1969, 1970, 1971, 1972, 1975, and 1978, during which their kidney function

was evaluated by several objective laboratory tests. In particular, increased levels of serum creatinine

may indicate poor kidney function.

1. (10 points) Create summary statistics for the data. Describe how study subjects are distributed

across NAPAP groups, as well as characterizing the groups in terms of age. In particular,

demonstrate with graphs and numerical summaries how/if the serum-creatinine levels change

for women in the different NAPAP groups (you do not need to calculate any tests at this point).

Mention the number of patients lost to follow-up over time. Make sure to label your graphs and

write a few descriptive sentences to describe your findings. You may use any statistical software

for this step.

2. (10 points) Compare the mean baseline serum-creatinine level among the NAPAP groups.

Describe and characterize all pairwise differences (including statistical significance). Write a few

sentences to summarize your findings. Please include an ANOVA table in your report. Check the

ANOVA assumptions and comment on the validity of the model.

3. (8 points) Repeat problem 2 with the serum-creatinine levels in 1978. How do your answers

differ from the answers to problem 2, if at all? Write a few sentences to summarize your

findings. Please include an ANOVA table in your report.

4. (8 points) Compare the mean change in serum-creatinine levels from the baseline visit to the

1978 visit among the NAPAP groups. Describe and characterize all pairwise differences. Write

a few sentences to summarize your findings. Please include an ANOVA table in your report.

5. (10 points) Analyze the differences in 1978 serum-creatinine levels among groups while

controlling for age. Include an ANOVA table and any other relevant SAS output. What are the

average serum-creatinine levels among each group, controlling for other variables? Are these

values different than the summary statistics you generated in problem 1? Why or why not?

What effect do other predictive variables have on serum-creatinine, and are the effects

statistically significant?

6. (10 points) Repeat problem 5, but include baseline serum-creatinine level as a predictive

variable. After answering the questions in problem 5, address how the analysis changed with

the inclusion of the new variable.

7. (10 points) Serum-creatinine levels over 1.3 are considered high and are associated with renal

insufficiency. What proportion of women in each group have high creatinine levels? Is the

proportion of women with high creatinine levels in 1978 significantly associated with NAPAP

group? How do you know? Include a frequency table in your analysis.

8. (4 points) What is the relative risk of high serum-creatinine levels in 1978 among women in the

high NAPAP group versus the low NAPAP group? You may calculate this by hand. Write a

sentence explaining your result.

9. (4 points) What is the odds ratio of high serum-creatinine levels in 1978 for women in the high

NAPAP group versus the control group? You may calculate this by hand. Interpret your answer.

10. (18 points) Run a logistic regression modelling the presence of high creatinine levels in 1978 as

the response variable and age and the presence of high creatinine levels in 1968 as the predictor

variables. Interpret the coefficients and discuss their significance (you may want to adjust the

reference groups and/or event status in your SAS code to make these have nicer

interpretations). What is the odds ratio in favor of high creatinine levels in 1978 for women

with high creatinine in 1968 versus those who do not have high creatinine in 1968?

SAS code that can be copy/pasted into SAS to run: 10 points

BONUS (5 points): What is the estimated attributable risk of high creatinine in 1978 for women in the

high-NAPAP group versus the control group? Interpret your answer.

BONUS (5 points): This analysis essentially ignored all of the data gathered between 1968 and 1978.

How might the intervening data points be included in the analysis?