Question 2 Part 1:

In determining what factors are the best predictors for cancer, and thus for doctors to look for when making their recall decision, or first goal was to check and see if any of the radiologists in the data set were perhaps more conservative than others when making their decision. If perhaps one doctor was more likely to recall a patient regardless of actual significant risk factors, it could bias the small data set we have towards one radiologist’s decisions.

We began by looking at the raw percentage of recalls in the data set and noticed that only 14.99% of all patients are recalled. Due to such a low percentage of recall occurrences, recall can therefore be termed a “rare outcome”, making accurate predictions more difficult. Next we looked at the raw number of percentages of breast cancer outcomes in out data set and noticed that there only 37 of the 987 positive cancer outcomes in the data, meaning that only 3.74% of all patients actually have cancer. Finally, we did the same process in order to view the raw number and percentage of patients who were seen by each radiologist in order to determine whether any one radiologist had seen a larger proportion of patients than others. Each of the 5 radiologists is associated with roughly 20% of the observations in the data set meaning that they each have a near equal weight on the total number of patients.

Our next step was to create a table that compared each radiologist with the raw percentage of recall outcomes that they are associated with. From the raw data, it looks as though radiologist 89 and 66 have the highest proportion of recalls with 3.85% and 3.74%, respectively, of the total 15% of recalls that occur (or roughly 25.67% and 24.99% of the total number of recall outcomes. The lowest percentages of recall outcomes were those conducted by radiologist 34, whose recall decisions were only 1.77% of the original 15% (roughly 11.48% of the total recall outcomes).

Although it is clear that some radiologists have much higher recall percentages than others, because this is only based on the raw data, we do not know whether some radiologists see a larger number of patients who, due to their specific risk factors, are at a greater risk for breast cancer. Therefore in order to find whether some of the radiologists are perhaps more clinically conservative than others, we must find the proportion of recalls associated with each radiologist while holding all patient risk factors fixed.

In order to do this, we created four generalized linear models to find out which radiologists have a higher recall rate, regardless of the set of patients that they have actually seen. The first model regresses cancer on recall in order to find what the odds are that someone is recalled given that they actually have cancer. This regression came back statistically significant, as one would hope, and it is clear that if a patient does have cancer, the chances that they are recalled by a doctor are incredibly high. However, due to the low percentage of recall outcomes, it has a very high deviance score. The next model regresses all variables in the data set on recall and also comes back with a high deviance score, however when looking at the odds rations for each radiologist, we can see that radiologists 66 and 89 are 1.49 and 1.61 times more likely to recall a patient than radiologist 13 who is the omitted category. It should also be noted that each radiologist did come back as statistically significant at a .05% level, and thus each radiologist individually affects a recall decision, and not just the patient risk factors. Our next model included and interaction between radiologist and cancer in order to control for cancers affect on recall in association with each radiologist. Something to note here is that radiologist34 actually came back as statistically significant on recall outcomes when the interaction variables were included even though the odds of them recalling a patient are only .053 times as likely to recall a patient as radiologist13 (compared to radiologist89 who is 1.5 times more likely to recall a patient than radiologist13).

From these models, we can see that even when we hold patient factors constant for all radiologists, radiologist89 and radiologist66 are more clinically conservative than their colleagues due to having much larger odds for recalling patients than radiologists 13, 34, and 95.