Update BIOSCAD data set analysis

Update

Analysis:

The same analysis was performed on the updated (14/09/17) BIOSCAD data set as was performed on the original. This includes both building regression models using elastic net and general regression, and the building classification models using logistic regression. There was little to no improvement in the models. This can be mostly attributed to two reasons:

1. The majority of the data remained the same. Only 28% of the data was modified.
2. Most of the new data was still either below the fit curve range or the detection limit. As such, the problems with overfitting and garbage in, garbage out still occur.

Visualisation of updated data set:

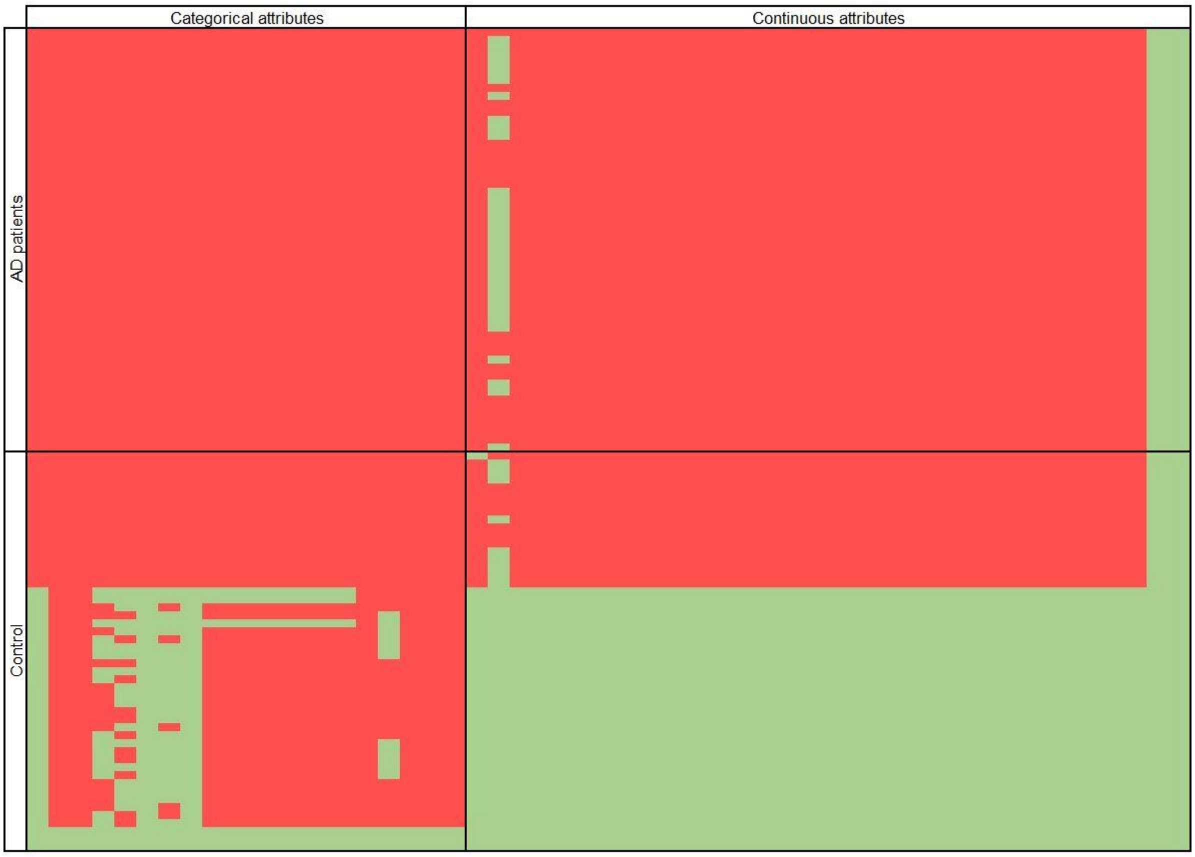

Figure 1

Figure 1 illustrates how the data set was modified from the original. Each square represents a value – if the square is red the data was unchanged, if the square was green the data was changed.