**Appendix 1** – Methods and results for sample size sensitivity analysis

There were significantly more adults aged 15-54 years in the NASS database compared to older occupants. There was a possibility that the general model for all adults was more accurate for younger adults due to their overrepresentation in the data. In order to determine the sensitivity of our results to the effect these differences in sample size, the analysis was repeated with up-sampling and down-sampling of age groups.

In the down-sampled analysis, occupants aged 15-54 were randomly removed so there were equal numbers of occupants 15-54 years and 55+ years. For the up-sampled analysis, occupants aged 55+ years were randomly duplicated so there were equal numbers of occupants 15-54 years and 55+ years. The same procedure from main analysis of using a split-sample and multiple iterations was employed to determine the AUC. The results are shown below:

|  |  |  |
| --- | --- | --- |
| Sampling Type | Age Group | AUC over 1000 iterations |
| Down-sampling (n = 6852) | 15-54 (n = 3426)55+ (n=3426)65+ (n=1819) | 0.867 (0.821-0.909)0.820 (0.773-0.862)0.809 (0.748-0.863) |
| Up-sampling (n = 27422) | 15-54 (n = 13711)55+ (n=13711)65+ (n=7366) | 0.874 (0.851-0.896)0.835 (0.814-0.853)0.820 (0.791-0.847) |

The up-sampled and down-sampled analyses both show a decreasing AUC with older occupants. All AUCs were higher in the up-sampled group but the trend remained the same. This indicates that higher accurate for younger occupants is not due to there over representation in the data.