Background There have been studies suggesting that various measures of health, health behaviours or health seeking cluster within households. Previous research on the impact of omitting small clusters from analyses have suggested little impact on parameter estimates or standard errors but have typically assumed small intraclass correlations – not usually the case for the household. Since there are typically few individuals per household, and many households contain just one person, several studies have excluded household from the analysis. We examine what difference it makes in practice when the level of household is excluded for a variety of outcomes.

Methods 7901 adults from 5063 households in 556 small areas were interviewed in the 2003 Scottish Health Survey, in 2512 (49.6%) of households just one adult participated. We analysed systolic blood pressure (SBP), BMI, current smoking status, eating 5+ portions of fruit and vegetables per day, eating oily fish at least once per week, and having seen a GP within the past 2 weeks. All results were adjusted for age and sex; in addition, we examined the effects of education, social class and area deprivation. Interactions were fitted between sex and all terms in the fixed and random parts of the model. Multiple imputation was used to account for item non-response. Multilevel linear and logistic regression models including and excluding household were fitted by Markov chain Monte Carlo and the results compared, with population average estimates obtained for logistic regression models.

Results The extent of clustering varied by outcome and gender; the proportion of the total unexplained variance at the household level ranged from 8.6% (95% Credible interval 0.3–24.5) for systolic blood pressure for women to 87.5% (83.7–90.7) for eating oily fish for men. Bias was evident in both the fixed parameter estimates and their standard errors. The bias in the fixed parameters tended to be higher for outcomes for which the household variance was higher.

Conclusion For all outcomes the inclusion of household provided an improvement in model fit. Excluding the level of household led to biases not just in the standard errors but also in the parameter estimates themselves. This has implications not only regarding the need to account for clustering within households but also regarding the use of “sandwich estimators” that adjust the standard error but not the regression coefficient.