**Table S2**. Comparison of two models, one that assumes that individuals in the same household are independent (logistic) and the second that assumes intra household clustering (random effects or generalized linear mixed model).

| Group | Dependant Variable | Adjusted odds ratios (Period-1 vs. Period-2) | | ICC |
| --- | --- | --- | --- | --- |
| | | Logistic model | Random effects model | |
| <2 years | Overall colonization | 0.58(0.42-0.78) | 0.53(0.37-0.77) | 0.23 |
| | VT | 0.35(0.27-0.46) | 0.32(0.23-0.44) | 0.11 |
| | NVT | 1.67(1.30-2.14) | 1.73(1.32-2.27) | 0.08 |
| 2-5years | Overall colonization | 0.66(0.45-0.96) | 0.61(0.38-0.96) | 0.21 |
| | VT | 0.71(0.51-0.99) | 0.68(0.47-0.99) | 0.08 |
| | NVT | 0.97(0.71-1.33) | 0.98(0.66-1.45) | 0.24 |
| 6-12years | Overall colonization | 0.65(0.48-0.88) | 0.63(0.44-0.89) | 0.13 |
| | VT | 0.60(0.40-0.90) | 0.58(0.13-2.65) | 0.98 |
| | NVT | 0.79(0.58-1.07) | 0.76(0.53-1.10) | 0.18 |
| 13-18years | Overall colonization | 0.46(0.27-0.79) | 0.46(0.26-0.81) | 0.01 |
| | VT | 0.45(0.15-1.30) | 0.31(0.13-0.72) | 0.02 |
| | NVT | 0.46(0.25-0.83) | 0.40(0.18-0.88) | 0.23 |
| 19-45years | Overall colonization | 0.45(0.31-0.64) | 0.42(0.27-0.64) | 0.21 |
| | VT | 0.36(0.17-0.75) | 0.23(0.13-0.40) | 0.01 |
| | NVT | 0.43(0.28-0.66) | 0.35(0.23-0.55) | 0.03 |
| >45years | Overall colonization | 1.11(0.53-2.35) | 0.63(0.32-1.24) | 0.01 |
| | VT | 0.70(0.17-2.88) | 0.31(0.15-0.64) | 0.01 |
| | NVT | 1.14(0.46-2.81) | 0.40(0.20-0.79) | 0.02 |
| Overall | Overall colonization | 0.68(0.59-0.78) | 0.66(0.56-0.78) | 0.13 |
| | VT | 0.52(0.44-0.62) | 0.46(0.37-0.57) | 0.24 |
| | NVT | 1.01(0.88-1.16) | 1.02(0.86-1.22) | 0.17 |

ICC – intra class correlation coefficient

The logistic model gives the same adjusted relative risks given in Table 2 in the manuscript (if the log binomial is used) – it assumes that subjects from the same household are independent. The hierarchical model aka the random effects model or Generalized Linear Mixed Model assumes that subjects from the same household are correlated. The model fitted here is $Y\_{ij} = b\_i + Beta0 + Beta1*Period\_{ij}$ + other factors such as age, gender, day care attendance, where $b\_i \sim N(0, sigma^2)$. It is the value of sigma that tells us how much clustering is in the data. The Intra-Class Correlation (ICC where class is the household) tells us the amount of clustering in the data $ICC=sigma^2/(sigma^2+(pi^2)/3)$.