

Supplementary Text S1: Exclusions and representativeness of the final study sample

Some children aged 2-4 years on the 31st August 2014 and registered in a THIN practice on the 31st January 2015 were not included in the final study sample due to missing data or because they were not registered with a THIN practice meeting quality criteria since before their first birthday. Also, only one child per family was included – one child per family was therefore randomly selected.

Figure S1 outlines the number of children excluded at each stage.

Children not excluded at random ($n=31,240$) were similar in terms of gender (48.5% of children excluded were girls compared to 48.8% of children in the study ($\chi^2 p = 0.39$)). Excluded children were significantly older, although the absolute difference was small; the mean age of children in the study was 3.0 years compared to 3.1 years among children who were excluded t -test $p < 0.001$. Excluded children were less likely to be vaccinated, yet the absolute difference was small: 39% in the cohort cf. 34% among children who were excluded. The slightly lower uptake among children who were excluded is partly due to excluding children who were registered in a THIN practice during the autumn of 2014. Hence they may have been vaccinated in another GP practice, but this is not captured in THIN.

Children aged 2-4 years on the 31st August 2014 registered with a THIN practice on the 31st January 2015
n=93,724

Exclude: children registered after their first birthday
n=24,559

Exclude: children with missing information on number of children in household
n=2,668

Exclude (at random): more than one child per family
n=4,939

Exclude: children with missing Townsend score
n=3,887

Exclude: children with missing rurality indicator
n=126

Final sample
n= 57,545 children