

## **SUPPLEMENTARY FILE 3 – DETAILED METHODOLOGY**

This Supplementary File details the methods used to calculate the cumulative incidence of referrals, assessments, child in need status and child protection plans.

### **Method for cohort 1**

1. Cohort 1 was defined as children born in 2012/13. These were followed up to the end of age 5.
2. The numerators for cohort 1 were the number of new referrals observed at each age.
3. The denominators for cohort 1 were ONS mid-year population estimates for each year of age as the cohorts aged.
4. Hazards of referrals at each age were then calculated by dividing the number of new referrals at each age by the number at risk at each (the previous year's population estimate minus the number of observed referrals). The cumulative hazard (CH) was then calculated as the cumulative sum of the age-specific hazards, and the cumulative incidence was calculated as  $1 - \exp(-CH)$ .

### **Method for cohort 2**

5. Cohort 2 was defined as children born in 2005/6, and who therefore turn 6 in 2011/12. These were followed up to age 12.
6. The numerators for cohort 2 were calculated in three steps:
  - a. As above, the numerators for cohort 2 were the number of new referrals observed at each age. However, because of left-censoring, a number of "first" referrals as observed in the data were in fact recurrent referrals. It was therefore necessary to estimate the number of expected recurrent referrals.
  - b. The numbers of expected recurrent referrals at age 6, 7, 8 and so on were based on, firstly, the recurrence rates calculated and documented in Supplementary File 4. Secondly, we used the actual numbers of referrals at ages 0-5 observed in cohort 1 to calculate the expected number of recurrent events across childhood.

This method accounts for the fact that the risk of re-referral (or the other events) declines with time, hence why we estimated the number of recurrent events by single year of age at event and for each year that follows (Supplementary File 4). That is, a child referred at age zero has a risk of re-referral that is highest at age 1 but then declines with time; by the time such a child is age 6 and in our observable window, their risk of re-referral is substantially lower. By contrast, a child first

## **SUPPLEMENTARY FILE 3 – DETAILED METHODOLOGY**

referred at age, say, 5, has the highest risk of re-referral at age 6, which is in the observable window.

This method assumes that the risk of recurrence of events from ages 0-5 into the 6-12 age period was the same for cohort 2 as it was for cohort 1.

- c. The number of expected recurrent referrals was subtracted from the number of observed referrals. This gave corrected yearly numerators.
7. The denominators (ONS mid-year population estimates) were additionally discounted to remove the estimated number of children no longer at risk, because they would have been referred in the left-censored period. We used the cumulative incidence of referrals calculated in step 4 to do so.
8. The cumulative incidence was then calculated as for cohort 1.

### **Method for cohort 3**

9. Cohort 3 was defined as children born in 2000/1, and who therefore turn 13 in 2013/14. These were followed up to age 17.
10. The numerators for cohort 3 were calculated in a similar way as in cohort 2 (step 6). In estimating the recurrent events from the 0-5 years into the 13-17 years, the observed numbers of referrals from cohort 1 was used. For recurrence from the 6-12 years into 13-17 years, the corrected numbers of referrals in cohort 2 were used.
11. The denominators (ONS mid-year population estimates) for cohort 3 were also discounted in the same way as for cohort 2 (step 7), additionally discounting for the cumulative incidence of referrals in the 6-12 period.
12. The cumulative incidence was then calculated as for cohort 1.

### **Calculating the total cumulative incidence of referrals, assessments, child in need status and child protection plans**

13. The estimates derived for each cohort at steps 4, 8 and 12, which are conditional probabilities, were then combined as follows to produce the final cumulative incidence estimates:

$$CI = 100 * (1 - (1 - a)(1 - b)(1 - c))$$

Where CI is the overall cumulative incidence to age 18 and a, b and c are the cumulative incidence estimates in cohorts 1, 2 and 3, respectively.

14. In order to calculate the cumulative proportion of children ever assessed and recognised as being CiN, the same method as in steps 1 to 13 was followed, *mutatis mutandis*.