**Guidance document for using CAFCASS CMS code**

The following code combines the Cafcass CMS tables into the format where every row is a child on an application in a case.

*Recommended file structure*

A folder which contains all the do files, a folder for raw Stata datasets which contains three sub-folders called “Final files”, “Interim files” and “Lookups”. The do files will save all interim files (cleaned datasets) into the interim files folder, all lookups which have been imported from DB2 into the Lookups folder and the final files from the end of each do file into the Final files folder.

There are several do files and below is an explanation of what each one does. This code has been tested using the 2024/01/11 refresh from Cafcass.

**Master\_cms** – will run all of the following do files in order to create the final dataset. At the start fill in the globals with your username and password, the refresh year of the dataset you are using (note, this has been tested on the most recent Jan 2024 refresh (2024/01/11), manual checks to ensure no changes to the data are highly recommended if you change this date), and your project number within SAIL. Also, change the command directories to your own file structure.

**01create\_applications\_cms** – this cleans the applications file. It contains information by the application ID and case ID. There may be multiple applications within a case. The dataset records the application type, application date (date received, data completed etc.), court levels and court IDs. For the court levels and IDs it merges the lookup files from the CAFEREFV table. You have multiple numbers of court references: court name, court ID, region, DFJ area, circuit reference and the law type. *Note, it does not clean the lookup files themselves. If you are using for any analysis you will need to check these variables carefully because for example there looks like multiple instances of the courts recorded in different ways.*

**02create\_people\_cms** – this cleans the people file which contains a person ID in the data (for anyone recorded as involved in a case). This file already contains people characteristics including ethnicity, disability, religion, language, gender, and week of birth (measured at the first Monday of the week of birth) and the person\_ID as well as the ALF IDs (for linkage to family court/census data).

**03application\_members\_cms** – this cleans the applications members file which links the person\_id with the application\_id and identifies the role of the person on the application (e.g. subject, respondent, applicant). There should not be people with multiple roles on an application so the code looks for duplicates and removes any and recodes those with multiple roles as unclear. Then merge in the people file to get the characteristics of people on applications. The sample is restricted to children (defined as anyone under the age of 18 and the subject on the application).

**04create\_relationships\_cms** – this cleans the person relationships file which details relationships between people in the dataset. Depends on the relationships of interest but this code identifies the mother and father of children on applications along with their own characteristics (merged from the people file). This is structured differently to the ECMS in that it identifies the child and their parents/relations rather than person and related person. First, the people file is merged with the related person ID to obtain the characteristics of the related person. Then, the file is reshaped so that there is one row per individual with all the related people and their characteristics in wide format. After that it’s possible to choose which relationships are of interest and run a loop with the related person and their characteristics. This has been done using the mother and father characteristics but it’s possible to change the relationship type to identify others (though parental relationships seem to be the only ones captured in the CMS in contrast to the ECMS where more extended family are included). The end result is a dataset of applications relating to a person attached to the relationships and characteristics of that related person. *Note: for around 0.5% of the sample there are multiple mothers and fathers, nothing has been done to resolve this, the code just takes the first instance of a “mother” for her characteristics.*

**05create\_legaloutputs\_cms** – this cleans the legal outputs file. The legal outputs are recorded per case (this differs from the ECMS where legal orders are recorded per person and application). A flag is created for each legal output recorded. There is a large number of legal outputs so they have been categorised into legal orders and outputs that were other measures of the case. Then further put into categories to make a more manageable dataset. This is entirely subjective. The legal outputs cleaned file is then merged into the main dataset.

**06create\_cases\_cms** – this cleans the cases file. It adds the court level IDs and the as well as the court form, law type, local authority and case type. The code adds the lookups and saves a new case file which is then merged into the main dataset.