



Using administrative data to quantify overlaps between public and private children law in England

Report for the Ministry of Justice on the Children in Family Justice Data
Share pilot

Authors:

Matthew A Jay
Rachel Pearson
Linda Wijlaars
Sofia Olhede
Ruth Gilbert

UCL Legal Epidemiology Group
UCL GOS Institute of Child Health
Population, Policy & Practice Programme
30 Guilford Street
London WC1N 1EH

matthew.jay.15@ucl.ac.uk

April 2019

Contents

Abstract 3

Summary of recommendations..... 3

Glossary 6

Abbreviations..... 8

Author contributions 8

Acknowledgements 8

Chapter 1: Background..... 9

Chapter 2: Sources of data 11

Chapter 3: Results and discussion 15

Chapter 4: Feedback and recommendations 25

References 32

Appendix I – data cleaning..... 33

Appendix II – ethnicity data 37

Abstract

The research

A conceptual distinction is often drawn between public and private children law under the Children Act 1989. Yet it has been suggested that, in practice, there is considerable movement of cases and individuals between the public and private systems. Currently, there is limited quantitative evidence of this suspected overlap. We aimed to take advantage of the Children in Family Justice Data Share (CFJDS)—linked data from the family courts, Cafcass, the National Pupil Database and children’s social care—to identify children involved in private family law cases, to describe their index cases and to quantify the proportion of these children who are also involved in public family law or who otherwise return to the private family courts. We observed roughly 32,000 to 45,000 section 8 cases per year between 2011 and 2016, involving between 48,000 and 66,000 children. Overall, 18% of those children with section 8 proceedings in calendar year 2011 had further private law proceedings by the end of March 2015. Moreover, 3.4% appeared in the public family courts during this period and at least 1.9% became looked after under section 20 of the 1989 Act. The presence of welfare concerns (i.e. section 7 report, rule 16.4 appointment or section 37 report) at the private index case did not predict return to court or accommodation under section 20.

Feedback on the data share

These analyses are based on the CFJDS pilot, which the Ministry of Justice recently invited applications to be tested. We had access to the secure Micro Data Lab, where the CFJDS data were stored, in December 2018 and January 2019. Routine linkage of family justice datasets, including those from the Ministry of Justice, Cafcass and Department for Education, was a key recommendation of the 2011 Family Justice Review and the Nuffield Family Justice Observatory data scoping report in order to improve the generation of research evidence in this area. Provision of access to the data share, and FamilyMan in particular, is therefore welcome. We make a series of recommendations based on our experience with the data to further improve their utility for researchers. These are discussed in detail in chapter 4 of this report and are summarised below.

Summary of recommendations

- **Recommendation 1:** To scale up access to the data share by external researchers, standard operating procedures, including access within a restricted area, could be considered.
- **Recommendation 2:** To ensure that software issues are dealt with promptly, a formal mechanism for external researchers to log software issues could be established.
- **Recommendation 3:** To improve analytical capability within the data analysis environment, a mechanism by which third-party packages for R and other supported statistical software can be downloaded on demand could be established. We understand, for example, that the Office for National Statistics

are mirroring the Comprehensive R Archive Network repository for use in their Secure Research Service.

- **Recommendation 4:** To improve the evidence base within family justice, data could be included on the caregivers that are collected in Cafcass and FamilyMan and that can be linked to children in order to build family cohorts.
- **Recommendation 5:** To ensure transparency and clarity around the data linkage undertaken to create the data share, the MoJ could provide documentation for researchers in line with the Guidance for Information about Linking Data Sets (Gilbert et al, 2018).
- **Recommendation 6:** To improve research output quality, mechanisms could be put in place to allow for external, publically available and anonymous data (such as deprivation indices) to be imported into the data analysis environment.
- **Recommendation 7:** Continued dialogue between researchers and each of the data providers involved in the CFJDS could be supported and encouraged for the mutual benefit of the data provides and researchers. Examples of such interchange could include the Children’s Social Care Data User Group, seminars, webinars and user support groups.
- **Recommendation 8:** The following minor improvements could be made to the metadata:
 - Further information on known issues with the data should be highlighted within the metadata document.
 - The different ID variables should be more clearly identified. The “id_supplied” variable in DfE, in particular, is ambiguous.
 - Mismatches between the metadata and data supplied should be corrected. For example, the SEN variable supplied (primary SEN type) is not the variable documented in the metadata (SEN provision).
- **Recommendation 9:** Our recommendations concerning the datasets are outlined in the following table.

Limitation	Dataset(s) affected	Recommendation
<p>9(i) Not all children and cases in FamilyMan and Cafcass appear to be in the CFJDS, especially given the shortfall of cases reported in the Results chapter compared with the routinely published figures. This is possibly due to the fact that only children who linked between FamilyMan and Cafcass are included. However, all children could be included as researchers may want to use each database separately. Those interested in linkages would also benefit from having the entirety of each database as information on children and cases who do not link assists in evaluating linkage and in understanding the groups of children under study.</p>	FamilyMan and Cafcass	All children and cases in FamilyMan could be included in the CFJDS, regardless of linkage status.
<p>9(ii) Age is a crucial variable for any analysis involving individuals. However, the way in which age is recorded in the CFJDS hinders analysis. The age variable is stored in the table “APP_DISP_JOIN” and cannot be linked to the other data tables for analysis because it uses a different child-level identifier. Further, age is categorised (e.g. <1 year old, 1-4 years old, etc.) whereas approximate date of birth is often more useful in longitudinal analyses (Note: exact date of birth is rarely needed other than for linkage).</p>	All	For each child, month and year of birth (but not day) could be provided. This will give analysts the flexibility to compute approximate ages as necessary at different events (e.g. case start or case end) or to form their own categorised age variable. Providing birth information in this way is less disclosive than date of birth but

Limitation	Dataset(s) affected	Recommendation
		still enables full multivariable analysis with age. An alternative would be to compute age in integer years at application date.
9(iii) There are several different child-level identifiers across the FamilyMan files.	FamilyMan	To ensure complete linkage, a consistent child-level identifier could be used across files from the same database.
9(iv) Not all court events were included in the CFJDS: only earliest and latest hearings within a case. Our original plan was to examine the complexity of cases by reference to number and timing of events in cases as this information is not routinely recorded for all private family law cases in existing data resources available to researchers (e.g. Cafcass). Access to this data would enable researchers to produce their own indicators of, for example, different combinations of applications and orders on a per-child and/or per-case basis.	FamilyMan and Cafcass	Complete data from FamilyMan and Cafcass (i.e. all records by individual and court event) could be provided in long format.
9(v) A serious limitation of the CLA data is that only one episode per year is included. This undermines the utility of the linked CLA data in analyses using this data share. Research has shown an enormous degree of variability between children in the number of episodes that they have within a period of care (e.g. some children will have multiple episodes in different placements or on different legal statuses). The current format of the data precludes accurate longitudinal analysis, which stands to be a key feature of this dataset. For example, the proportion of children who have a section 20 episode or who enter care under police protection is likely to be underestimated (see Results chapter).	CLA	All episodes in the CLA dataset could be included.
9(vi) The CLA data is missing key dates (such as episode start date), therefore precluding longitudinal analysis. Available only are episode end date and period of care length. It should be noted that a “period of care” is a period of one or more temporally contiguous episodes; a new episode begins when there is a new legal status or new placement. Therefore, the episode end date minus the period of care does not reliably give either the episode start date or the period of care start date. In addition, we know from our own analyses with the complete CLA data that the length of period of care variable is erroneous and this can be demonstrated also to be the case within the CFJDS CLA data.	CLA	Episode start and end date could be included so that periods of care, and their duration, can accurately be determined by the analyst.
9(vii) There is a large amount of missing data in the NPD censuses. For example, primary SEN type is completely missing from 2013 onwards.	NPD censuses	The missing data in the NPD censuses could be investigated and rectified.
9(viii) The NPD censuses and CLA data do not have the same temporal coverage as the other datasets.	NPD censuses and CLA	The censuses and CLA data could cover the same period so as to enable full longitudinal analysis.
9(ix) Important census variables, notably the income domain affecting children index and FSM eligibility are absent, as is the SEN provision variable.	NPD censuses	The income domain affecting children index, FSM eligibility and SEN provision variables could be requested and included in the CFJDS.

Glossary

Cafcass. The Child and Family Court and Advisory Support Service: Cafcass are a national social work organisation who, through appointed guardians, advocate for children in public family proceedings and carry out safeguarding checks in private family law proceedings.

Care order. A court order authorising the local authority's care plan and enabling the authority to receive the child into care.

Emergency protection. A child may become looked after through an emergency protection order granted by the court or through police protection (no court order required). Where a local authority wishes to assess a child but the parents refuse it can also apply for a child assessment order but in practice this rarely occurs. These orders and powers are strictly time-limited.

Private family law. The branch of family law dealing with disputes between private individuals (such as parents on relationship breakdown) about the upbringing of children. Such proceedings are characterised mainly by section 8 orders.

Public family law. The branch of family law dealing with state intervention in the upbringing of children. Such proceedings are characterised mainly by applications for care and supervision orders by local authorities.

Rule 16.4 appointment. A situation where, in private family law proceedings, a guardian is appointed to advocate for a child where to do so is necessary for the child's welfare.

Section 7 report. The court may order either Cafcass or the local authority to prepare a report for the court on such matters relating to the welfare of the child as the court requires.

Section 8 order. A group of court orders used to determine disputes between private parties concerning the upbringing of children. Included are child arrangements orders (formerly contact orders and residence orders), specific issue orders and prohibited steps orders.

Section 20 accommodation. Local authorities have a set of duties and powers under section 20 to accommodate children. In some circumstances (e.g. there are no parents) the authority must do so. Where a person with parental responsibility is willing and able to provide accommodation, the authority may only look after a child under section 20 where that person does not object.

Section 37 report. A court may order a local authority to undertake an investigation of a child's circumstances with a view to consider applying for a care or supervision order (though the court may not compel the authority to apply for such an order).

Supervision order. A court order placing a child under the supervision of the local authority.

Welfare concerns. In this report, 'welfare concerns' are identified in private family law cases by reference to whether a section 7 or section 37 report was ordered or a rule 16.4 appointment was made.

Abbreviations

ADJ	APP_DISP_JOIN (one of the FamilyMan tables, discussed in chapter 2)
Cafcass	Children and Family Court Advisory and Support Service
CFJDS	Children in Family Justice Data Share
CLA	Children Looked After
DfE	Department for Education
FJO	Family Justice Observatory
FSM	Free School Meals
IMD	Index of Multiple Deprivation
LA	Local authority
Moj	Ministry of Justice
NPD	National Pupil Database
SEN	Special Educational Needs

Author contributions

The authors designed the study and analysis plan and applied to the Ministry of Justice for access to the data. MAJ performed all statistical analyses. All authors contributed to the preparation of this report. Neither the Ministry of Justice nor the funder played any role in study design, analysis, interpretation or the preparation of this report. The views expressed are those of the authors.

Acknowledgements

We would like to thank the Ministry of Justice for providing access to the Children in Family Justice Data Share. In particular, Alexy Buck, Amy Summerfield, Sarah French and Lisa Robinson all helped to facilitate access. We also thank Charlie Smuliman, Nick Newcomen and James Wickenden for escorting MAJ while at the Ministry. Finally, we thank Ross Black for providing assistance with understanding the data and linkage process.

The Medical Research Council via a UCL-Birkbeck Doctoral Training Partnership Flexible Training Supplement funded MAJ for the duration of this project.

Chapter 1: Background

Structure of this report

This chapter contains a brief background to the area of investigation and the Children in Family Justice Data Share (CFJDS). The data are outlined in the next chapter (chapter 2) and results are described and discussed briefly in chapter 3. Chapter 4 discusses our experience of working with the CFJDS.

Throughout this document 'private law' and 'public law' refer specifically to private and public *family law*.

Public and private family law

Within the context of the Children Act 1989, a distinction is frequently drawn between public family law (such as local authority [LA] accommodation under section 20 or applications for care orders under section 31) and private family law (such as disputes between parents experiencing relationship breakdown about where and with whom a child should live, cases which are brought primarily under section 8). However, this neat distinction between the two 'branches' of children law is, it has been suggested, too simplistic (e.g. Bainham, 2013). It has been argued, for example, that many private cases are in fact driven by LA impetus by supporting and encouraging family members to apply for the 'private' orders rather than initiating court action themselves (Bainham, 2013). Secondly, the LA will already be involved in the lives of many children incidentally to the private law proceedings or may become involved if welfare concerns are raised. Finally, public cases can result in so-called private orders, even if no such order was applied for, or a combination of private and public orders in the form of special guardianship plus supervision.

These issues concern each court case and the factors immediately surrounding it. It should of course also be noted that children are looked after by LAs through section 20 accommodation (i.e. the child's being received into care either where there is no known person with parental responsibility or where such a person does not object to the accommodation) and through police emergency protection powers. Neither of these require, in the first instance at least, recourse to the court.

The true extent of public/private overlap is unquantified and, crucially, the above discussion focuses on overlap concerning cases contemporaneously. It ignores the longitudinal aspect, which is important in understanding the true scale at which children are affected by the Act. For example, approximately 0.6% of children are in care at any given time but of all children born between 1992 and 1994, 3.3% entered out-of-home care at least once before their 18th birthdays (Mc Grath-Lone et al, 2016). In other words, when accounting for this longitudinal aspect, it is likely that the overlap between children's experience of the public and private law systems is more substantial than would first appear by inquiring only whether a child at any given moment is involved with both systems. It is, however, not known how many children involved in private law disputes are involved with the public family law system (including through care orders

and other court orders, accommodation under section 20 and police emergency protection) or will go on to be.

Recently, the Ministry of Justice (MoJ) has invited applications to access and analyse data held in its Children in Family Justice Data Share (CFJDS) as part of a time-limited pilot and feasibility study of the dataset and its secure Micro Data Lab. The CFJDS contains linked data from the MoJ's family courts administrative database named FamilyMan, Children and Family Court Advisory and Support Service (Cafcass) data, Children Looked After (CLA) data and National Pupil Database (NPD) data—these latter two being held by the Department for Education (DfE). The provision of access to such data, especially FamilyMan, is welcome and this was one of several recommendations in the Family Justice Review (MoJ, 2011) and the Nuffield Family Justice Observatory (FJO) data scoping report to improve the quality and availability of research evidence in the Family Justice System (Jay et al, 2017).

Objectives

The study had three objectives:

- 1) Identify yearly cohorts of children subject to section 8 applications and describe these children and the cases.
- 2) For children subject to private law proceedings in 2011, quantify the proportion who return to court (public or private) or become looked after through extra-judicial legal routes within five years.
- 3) Following completion of the pilot, to provide feedback to the MoJ based on our experience of the CFJDS and its associated secure Micro Data Lab. This feedback (which is provided in the present report) will be considered as part of an internal MoJ review of the pilot.

Chapter 2: Sources of data

A summary of some principal data considerations are given here. A schematic representation of the data sources is given in Figure 1, overleaf. Methodological issues are elaborated upon throughout in chapters 3 and 4.

Data linkage

FamilyMan and Cafcass data were linked by MoJ according to a multi-wave deterministic algorithm. Essentially, children were linked using identifiers common to both datasets such as names, genders, dates of birth and ethnicities. Match rates were in the region of 75% though we do not at the time of writing have precise details of this. Linkage between FamilyMan and DfE was conducted according to a similar method by DfE. The CFJDS contains a linked IDs table which we used to link children's records.

Time period

Cafcass data are available from March 2007. Prospectively collected FamilyMan data are available from 2003, though prior to 2007 data for the Family Proceedings Courts were weighted estimates based on a subset of courts. An administrative system upgrade was completed by December 2010, meaning the most reliable data are available from January 2011 (Jay et al, 2017). Linked NPD and whole-population CLA data are available from 2005/6. We chose to analyse cases starting in calendar years 2011 to 2016 as this gave complete coverage across FamilyMan and Cafcass. Issues around coverage of the CLA and NPD data are considered below.

Individual datasets: FamilyMan

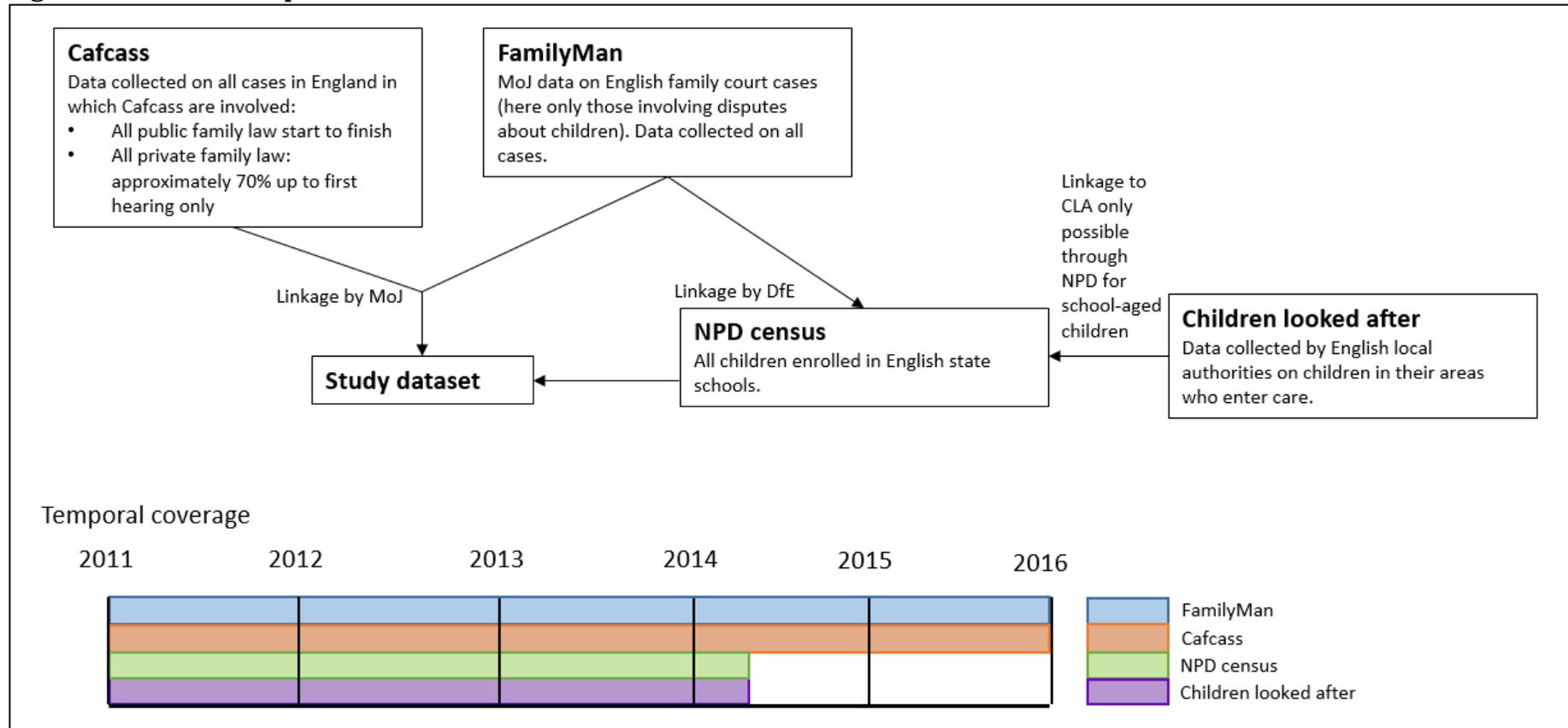
The FamilyMan dataset consisted of a table called APP_DISP_JOIN (ADJ), which was a long-format dataset with applications and orders and, among other things, the child's age in pre-defined brackets. However, it did not use the same child ID as the other FamilyMan tables and therefore age could not be directly linked to them for analysis.

Besides ADJ, there are two wide tables: "Wide" and "No Orders". These are identical but for the fact that "Wide" contains a series of binary indicators indicating the presence of certain applications and disposal types. "No Orders" is therefore superfluous and was discarded. "Wide" is actually something of a hybrid of a wide and long format dataset, depending on the unit of analysis. There is one row per child per case.

There are also two long datasets. "Orders" contains simply a list of applications and disposals but no dates and there is no method to link this information precisely to the other tables due to a lack of identifiers. The dataset labelled "Long" contains only disposals and no applications so I discarded it.

In these FamilyMan tables, all completion dates are non-missing indicating that only disposed-of cases were included (i.e. cases started but continuing were not included).

Figure 1. Schematic representation of the data sources



Note that years are given only for the years used in the present analysis. NPD National Pupil Database.

Individual datasets: Cafcass

There are two Cafcass tables, labelled A and E included in the CFJDS. A is a child-level “spine” containing some demographic data (principally sex and ethnicity, though other variables, including age, are available in the complete Cafcass records not part of the CFJDS) and E is a long-form table containing applications and orders. Cafcass has missing case completion dates indicating that, unlike the FamilyMan tables, it does include unfinished cases

Individual datasets: Children Looked After

The CLA is long-format and is missing key variables and data. Notably, it does not contain episode start date, only episode end date and period of care length (which, excepting the fact that a period of care is different to an episode, is inaccurate). Only one episode per financial year is given, meaning that any changes of episode within year are missing. The CFJDS contains CLA data up to 31 March 2015.

Individual datasets: National Pupil Database

The NPD consists of a range of modules including, for example, the annual census, attainment, absences and exclusions. We only examined the census files. The NPD census is a rich data source on school-age children and includes demographic data and other key information such as special educational needs (SEN) and educational attainment. However, few variables from the censuses were included in the data share. There are two principal SEN variables in the censuses: Primary SEN type and SEN provision. The latter indicates what level of SEN support the child is receiving (Nothing; or Action/Action Plus/Support; or a Statement/Education, Health and Social Care Plan) whereas the former indicates the type of SEN the child has (e.g. autism). The CFJDS has the primary SEN type variable but not SEN provision. For our purposes, SEN provision would have been more informative as we were more interested in the level of support required, not the type of SEN the child has. We therefore had to dichotomise children into having SEN vs not.

Another key issue with the census data is that there appear to be large swathes of missing data in the census file. We did not have time to investigate this in depth but our results show that the prevalence of SEN appears to drop to 0% in latter years, which cannot be correct. Further, the censuses in the CFJDS only cover up to 2014/15.

Cleaning

We undertook data cleaning as documented in Appendix I.

Analysis

We first identified children undergoing s 8 applications within each calendar year 2011 to 2016. Only the first application per child is counted (i.e. the index case). For these children, we then described, using appropriate summary statistics as outlined in the

results, the (a) index cases; and (b) the children themselves. Finally, we computed the proportions involved in further proceedings and in non-judicial care entry, up to March 2015, disaggregated by whether welfare needs were identified in the index case.

Analysis was conducted in R with R Studio in December 2018 and January 2019.

Chapter 3: Results and discussion

Objective 1: Identify yearly cohorts of children subject to section 8 applications and describe these children and the cases.

Case characteristics

Tables 1 and 2 (overleaf) show the characteristics of the index cases. In Table 1, the characteristics obtained from FamilyMan are given; in Table 2, those from Cafcass. Many more cases were identified each year in Cafcass than in FamilyMan. The Cafcass figures were closer to the routinely published figures available on the Cafcass website (Cafcass, 2018) than were the FamilyMan figures relative to the MoJ Family Court Statistics Quarterly reports (MoJ, 2018). For example, in calendar year 2016, according to the Cafcass website, there were 36,699 private law cases starting compared to 35,357 (96%) in our Table 2. According to MoJ figures, there were 48,246 cases starting that year, compared to 28,208 (58%) in our Table 1.

The distribution of children per case is roughly equal in both datasets: in just over 60% of cases there was only one child, a little under 30% had two children, about 7.5% had three and the remainder had four or more (Tables 1 and 2).

FamilyMan and Cafcass both provide case duration data. However, the case duration variable within Cafcass records durations as the date a case is opened on their system to the date that it is closed and not when the case actually finishes. For most private law cases this will be some time shortly after the first hearing; the Cafcass duration data are therefore omitted. Using the FamilyMan case duration variable, a decline in the median case duration is observed over time (Table 1). However, especially for 2016, this is likely due to truncation of cases that start in that year but whose end dates have not yet been observed. See Figure 2 (after Table 2, below) for histograms of the end dates of cases starting each year. We observed a large spike in 2014 for cases starting in 2011 and 2012. This related to the summer months of 2014 (Figure 3) and indicates a possible administrative exercise of 'filling in' missing end dates.

Finally, data on welfare need identified up to the first hearing are available in Cafcass (Table 2). Forty-six per cent to 47% of cases had some kind of welfare need identified up to the first hearing, the majority of which was accounted for by section 7. Small but rising proportions of cases had a rule 16.4 appointment or section 37 report.

Table 1. Characteristics of index cases (private family law) per FamilyMan

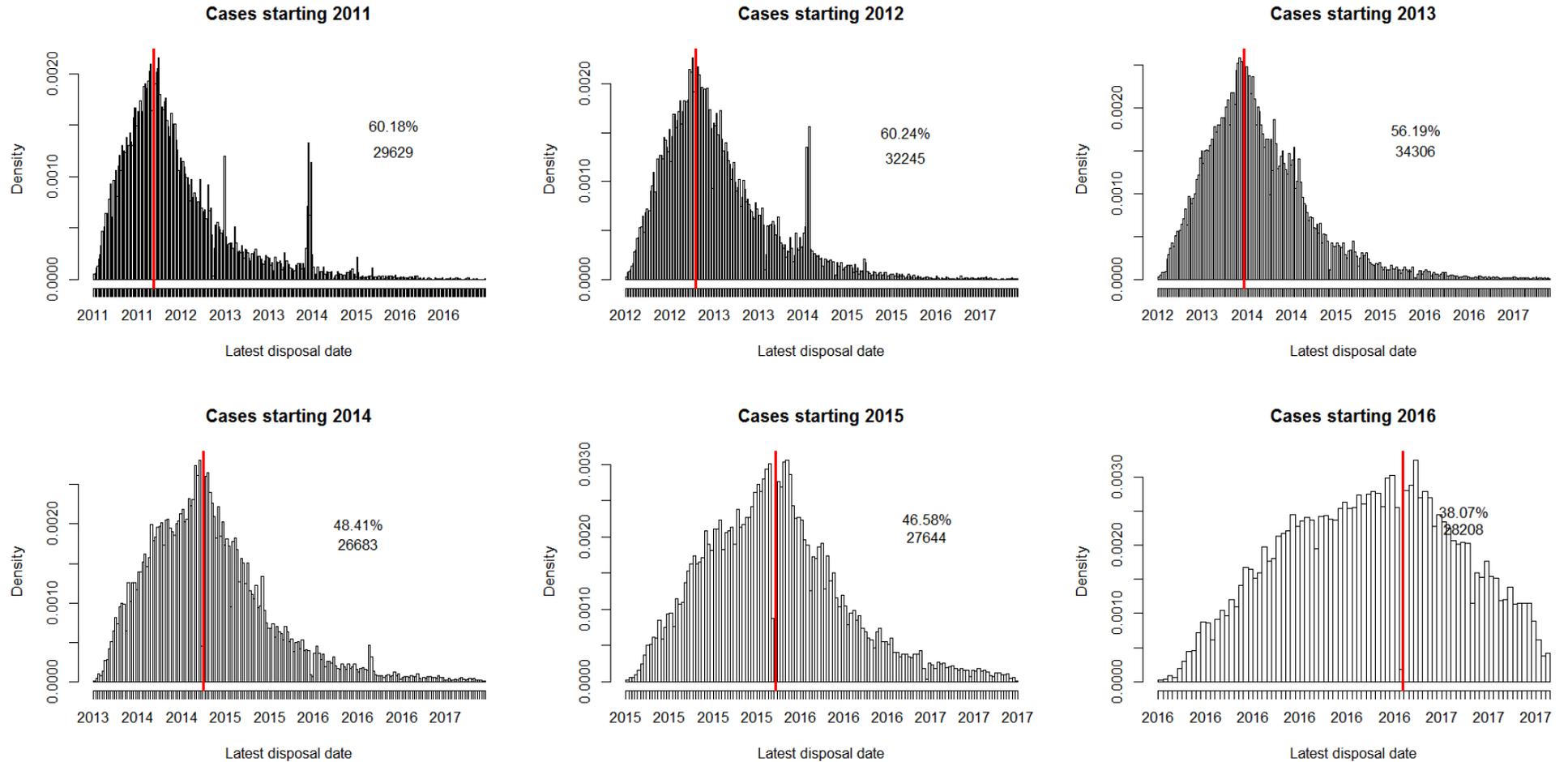
		2011	2012	2013	2014	2015	2016
N cases		29629	32245	34306	26683	27644	28208
Children per case	1	18753 (63.3%)	20654 (64.1%)	22139 (64.5%)	16737 (62.7%)	17367 (62.8%)	17602 (62.4%)
	2	7918 (26.7%)	8491 (26.3%)	8970 (26.1%)	7401 (27.7%)	7659 (27.7%)	8034 (28.5%)
	3	2306 (7.8%)	2487 (7.7%)	2479 (7.2%)	1936 (7.3%)	2050 (7.4%)	2043 (7.2%)
	4	486 (1.6%)	466 (1.4%)	535 (1.6%)	482 (1.8%)	448 (1.6%)	432 (1.5%)
	5	115 (0.4%)	109 (0.3%)	126 (0.4%)	100 (0.4%)	88 (0.3%)	75 (0.3%)
	6+	51 (0.2%)	38 (0.1%)	57 (0.2%)	27 (0.1%)	32 (0.1%)	22 (0.1%)
Duration	Median (IQR)	31.14 (14.43, 61.14)	31.00 (15.14, 58.43)	28.00 (14.29, 49.71)	22.14 (11.14, 38.86)	20.43 (10.71, 34.29)	18.00 (9.00, 28.00)
	(5 th , 95 th centiles)	(3.57, 152.43)	(4.43, 120.43)	(5.00, 98.29)	(4.29, 81.86)	(4.57, 66.14)	(4.29, 47.14)
	>26 weeks	16817 (56.8%)	18370 (57.0%)	18257 (53.2%)	11327 (42.5%)	10512 (38.0%)	8183 (29.0%)

Table 2. Characteristics of index cases (private family law) per Cafcass

		2011	2012	2013	2014	2015	2016
N cases		39840	43168	44838	32473	32467	35357
Children per case	1	24535 (61.6%)	26975 (62.5%)	28182 (62.9%)	19922 (61.3%)	19873 (61.2%)	21518 (60.9%)
	2	11114 (27.9%)	11807 (27.4%)	12171 (27.1%)	9234 (28.4%)	9341 (28.8%)	10270 (29%)
	3	3153 (7.9%)	3381 (7.8%)	3394 (7.6%)	2516 (7.7%)	2492 (7.7%)	2745 (7.8%)
	4	784 (2.0%)	749 (1.7%)	813 (1.8%)	626 (1.9%)	578 (1.8%)	663 (1.9%)
	5	174 (0.4%)	191 (0.4%)	198 (0.4%)	141 (0.4%)	131 (0.4%)	129 (0.4%)
	6+	80 (0.2%)	65 (0.2%)	80 (0.2%)	34 (0.1%)	52 (0.2%)	32 (0.1%)
Welfare need	Any*	18346 (46.1%)	18779 (43.5%)	20267 (45.2%)	15110 (46.5%)	15091 (46.5%)	16565 (46.9%)
	<i>Section 7</i>	17651 (44.3%)	17970 (41.6%)	19242 (42.9%)	14022 (43.2%)	13948 (43.0%)	15178 (42.9%)
	<i>Rule 16.4</i>	388 (1.0%)	697 (1.6%)	1288 (2.9%)	1286 (4.0%)	1353 (4.2%)	1347 (3.8%)
	<i>Section 37</i>	755 (1.9%)	749 (1.7%)	769 (1.7%)	734 (2.3%)	668 (2.1%)	923 (2.6%)

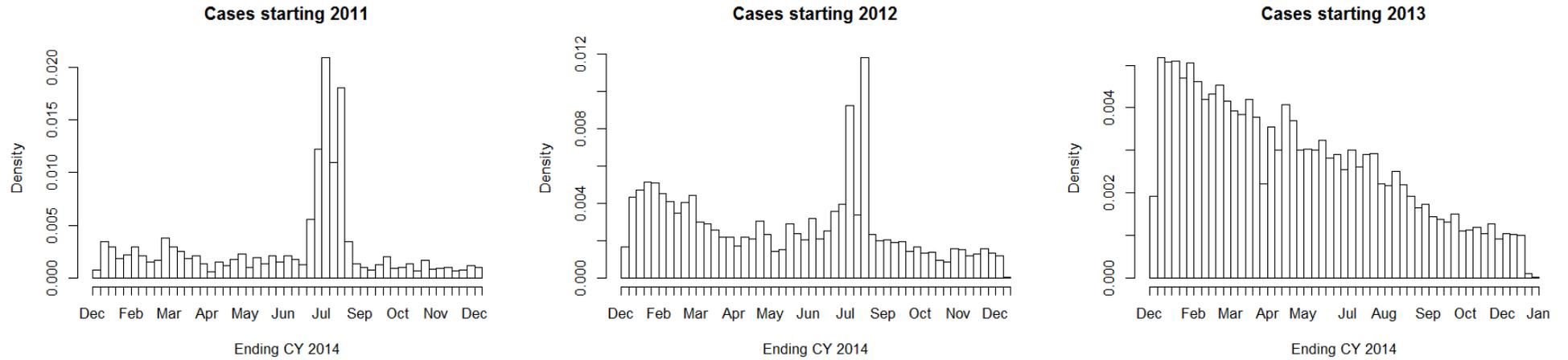
* I.e. any one (or more) of section 7, rule 16.4 appointment or section 37.

Figure 2. Histograms of case end dates for index cases (private family law) starting each calendar year.



Each bar represents the number of cases ending in each week. The vertical, red line is the last day of the given year. The percentage given in each panel is the percentage of cases ending after the last day of the year.

Figure 3. End dates in 2014 for index cases (private family law) starting in 2011, 2012 and 2013.



Each bar represents the number of cases ending in each week.

Child characteristics

Both Table 3 and Table 4 (overleaf) describe the children involved in private law cases within each calendar year between 2011 and 2016; Table 3 is derived from FamilyMan data (which gives age) while Table 4 is derived from Cafcass data (which gives gender and ethnicity) with linked SEN data (derived from linkage to NPD). Note that the age data are in the FamilyMan ADJ table, which is not linkable to the FamilyMan Wide table, from which the data on cases, above, are drawn. The age data are therefore for a slightly different group of children due to a slight discrepancy between the two data tables. The age distribution pattern, however, is similar to that observed in Halliday et al (2017) when they examined private law applications in Cafcass data.

The gender and ethnicity distribution are shown in Table 4. The very high proportion of missing ethnicity data, across all years, renders the ethnicity estimates unreliable. Proportions by different ethnic group are given in Appendix II on the assumption that data are missing completely at random. However, it is possible that this assumption does not hold and that therefore the proportions presented in Appendix II may be biased.

Whether a child had SEN was determined by reference to the two NPD census years surrounding the calendar year of the index case. For example, children with index cases in 2011 were determined to have SEN if they had an SEN code in the censuses for 2010/11 or 2011/12. Because the census data supplied stop in 2014/15, this means that children with index cases in 2015 can only have SEN recorded in 2014/15, and those in 2016 cannot have any. However, we found that the prevalence of SEN drops from 14% in 2013 to 0% in 2014: this is an error likely induced by the missing data identified in the NPD census data and listed in the methods, above.

Table 3. Characteristics of children with index cases (private family law) in each calendar year per FamilyMan.

	2011	2012	2013	2014	2015	2016
N children (Wide)*	43863	47251	49968	39398	40647	41048
N children (ADJ)*	44917	48272	51116	40402	41710	42138
Age						
<1	2882 (6.4%)	3138 (6.5%)	3169 (6.2%)	2372 (5.9%)	2202 (5.3%)	2052 (4.9%)
1-4	15905 (35.4%)	17447 (36.1%)	18031 (35.3%)	14031 (34.7%)	14112 (33.8%)	13460 (31.9%)
5-9	16230 (36.1%)	17416 (36.1%)	19308 (37.8%)	15518 (38.4%)	16461 (39.5%)	17004 (40.4%)
10-14	8505 (18.9%)	8880 (18.4%)	9238 (18.1%)	7466 (18.5%)	7834 (18.8%)	8482 (20.1%)
15-17	916 (2.0%)	989 (2.0%)	1020 (2.0%)	771 (1.9%)	895 (2.1%)	948 (2.2%)
Other†	479 (1.1%)	402 (0.8%)	350 (0.7%)	244 (0.6%)	206 (0.5%)	192 (0.5%)

* Age data are contained only in the ADJ file, which cannot be linked to the Wide file, and therefore the age data presented are for a slightly different group of children than otherwise presented in this manuscript.

† The 'other' age category is defined by MoJ as those who are aged 18 by the time the order was made or where age was missing.

Table 4. Characteristics of children with index cases (private family law) in each calendar year per Cafcass

	2011	2012	2013	2014	2015	2016
N children	60008	64093	66240	48766	48691	53100
Gender						
Female	28914 (48.2%)	30874 (48.2%)	31858 (48.1%)	23041 (47.2%)	23297 (47.8%)	25457 (47.9%)
Male	30324 (50.5%)	32494 (50.7%)	33563 (50.7%)	24349 (49.9%)	24269 (49.8%)	26303 (49.5%)
Unknown	770 (1.3%)	725 (1.1%)	819 (1.2%)	1376 (2.8%)	1125 (2.3%)	1340 (2.5%)
Ethnicity						
White	9287 (15.5%)	8351 (13.0%)	7731 (11.7%)	7135 (14.6%)	17238 (35.4%)	28355 (53.4%)
Black	316 (0.5%)	287 (0.4%)	377 (0.6%)	432 (0.9%)	890 (1.8%)	1425 (2.7%)
Asian	645 (1.1%)	694 (1.1%)	714 (1.1%)	790 (1.6%)	1876 (3.9%)	2748 (5.2%)
Mixed	952 (1.6%)	989 (1.5%)	1118 (1.7%)	1101 (2.3%)	2515 (5.2%)	3891 (7.3%)
Other	152 (0.3%)	145 (0.2%)	149 (0.2%)	172 (0.4%)	369 (0.8%)	586 (1.1%)
Unknown	48656 (81.1%)	53627 (83.7%)	56151 (84.8%)	39136 (80.3%)	25803 (53%)	16095 (30.3%)
N link to NPD	46451	49247	49281	34326	31097	10042
Has any SEN	5907 (12.56%)	8239 (16.49%)	6925 (13.79%)	0 (0.00%)	0 (0.00%)	0 (0.00%)

NPD National Pupil Database; SEN Special Educational Needs.

Objective 2: For children subject to private law proceedings in 2011, quantify the proportion who are otherwise made subject to various returns to the system.

Table 5 (overleaf) provides information on whether children with an index case in 2011 returned to court or otherwise into the care system by the end of March 2015. Data are stratified by whether there were welfare concerns (i.e. a section 7 or 37 report or a rule 16.4 appointment) identified by the first hearing within the case. Overall (final column), 21% of children had a return to court. The majority of these were returns in private applications. Into public law 3.4% children returned (mostly on application for a care order [3.1% of all children]). Only 0.2% of children returned for placement or adoptions proceedings but this is likely to be an underestimate as Cafcass are not always involved in these cases.

In order to identify whether a child is received into care under section 20 or police powers, it is necessary to link to the CLA data which occurs through the NPD censuses. As only school-aged children are in the NPD censuses, not all children will link to it, and therefore to the CLA data. Table 5 gives the number of children with a link to NPD data. Of these, overall, 21% either returned to court (public or private) or were received into care under section 20 or police protection, as identified in the CFJDS CLA data. However, as noted above, the CLA data are incomplete and therefore these figures are likely to be an underestimate, especially for police protection.

Finally, stratification by welfare concerns indicates that the return rate to court was similar in those who had welfare concerns identified at the index case vs those who had not (19% with welfare concerns at the index case vs 22% without). The difference in the return rate to court or section 20 or police protection was similar (20% in those with welfare concerns at the index case vs 22% without).

Table 5. Children with an index case (private family law) in 2011 returning by the end of March 2015 to the system by type of return application and whether welfare needs were identified by the first hearing

	No welfare concerns	Welfare concerns	Overall
N children	31290	28718	60008
Ever return to court?	7023 (22.44%)	5418 (18.87%)	12441 (20.73%)
Ever return to private court?	5919 (18.92%)	4726 (16.46%)	10645 (17.74%)
<i>Section 8</i>	5971 (19.08%)	4761 (16.58%)	10732 (17.88%)
<i>Special guardianship order</i>	26 (0.08%)	20 (0.07%)	46 (0.08%)
Ever return to public court?	1218 (3.89%)	828 (2.88%)	2046 (3.41%)
<i>Care order</i>	1102 (3.52%)	758 (2.64%)	1860 (3.10%)
<i>Supervision order</i>	133 (0.43%)	84 (0.29%)	217 (0.36%)
<i>Emergencies*</i>	60 (0.19%)	60 (0.21%)	120 (0.20%)
<i>Secure accommodation</i>	†	†	9 (0.01%)
Ever return for placement/adoption?	86 (0.27%)	33 (0.11%)	119 (0.20%)
N children with NPD link	24365 (77.87%)	22671 (78.94%)	47036 (78.38%)
Ever return to court or section 20?	5419 (22.24%)	4390 (19.36%)	9809 (20.85%)
<i>Section 20</i>	474 (1.95%)	415 (1.83%)	889 (1.89%)
<i>Police protection</i>	11 (0.05%)	15 (0.07%)	26 (0.06%)

Data are from Cafcass for court returns and children looked after data for section 20 and police protection. Welfare needs are “any welfare needs” as defined in Table 2. * Includes applications for emergency protection orders or child assessment orders. † Suppressed due to small cell counts. NPD National Pupil Database.

Chapter 4: Feedback and recommendations

In this chapter, we discuss our experience using the Micro Data Lab (i.e. the infrastructure for data access and output release) and the CFJDS in general. Lisa Robinson asked us a number of specific questions, which are reproduced and answered below. Finally, we have also taken the opportunity to suggest some improvements that would significantly enhance the utility of this dataset for future analysis and research; for this we have partly drawn on the FJO data scoping report (Jay et al, 2017). The recommendations are drawn together at the end of this chapter.

Where we use the word 'I,' this refers particularly to MAJ in his capacity as data analyst.

Micro Data Lab

The Data Lab consisted of a laptop with no internet access. Data were stored on an MoJ server and I had access only to a folder containing the data that form the CFJDS. To access the building, I was given an 'escorted day pass' and was required to be accompanied from reception to my desk, and vice versa. Statistical disclosure control was performed by MoJ staff, after which outputs were e-mailed to me.

There were two software issues. The first was that, due to having no internet access, Microsoft Office was disabled before analysis was completed. Although we were able to work around this (coding output tables in R and writing directly to CSV files and taking written notes in Notepad), we may not have been able to do so in other projects. Analysts would benefit from a mechanism by which such software problems can readily be fixed, especially as analysts' time with the Micro Data Lab is circumscribed.

The second was that R packages could not be installed (or at least not in the time in which we had access to the data). Although I did not use other statistics programmes, the same is therefore likely to be true of Stata. Again, although we were able to engineer work-arounds, we can easily envisage analyses we would wish to perform or outputs we would wish to produce for which third-party packages are essential.

We therefore have four recommendations concerning the Micro Data Lab.

- **Recommendation 1:** To scale up access to the data share by external researchers, standard operating procedures, including access within a restricted area, could be considered.
- **Recommendation 2:** To ensure that software issues are dealt with promptly, a formal mechanism for external researchers to log software issues could be established.
- **Recommendation 3:** To improve analytical capability within the data analysis environment, a mechanism by which third-party packages for R and other supported statistical software can be downloaded on demand could be established. We understand, for example, that the Office for National Statistics are mirroring the Comprehensive R Archive Network repository for use in their Secure Research Service.

The CFJDS

The MoJ should be applauded in its efforts to exploit FamilyMan and to carry out the ambitious and difficult task of linking FamilyMan data to Cafcass and DfE data. The FJO data scoping report (Jay et al, 2017) recommended the provision of access to FamilyMan and routine linkage across family justice system datasets in order that policy- and practice-relevant research questions can be rigorously evaluated. The CFJDS is unique in providing information on children in public and private family law (and across the entirety of these cases) as well as those who enter care via non-judicial routes (i.e. through section 20 accommodation and police protection). It also, crucially, provides data on substantive outcomes—education—which, although these were beyond the scope of our investigation, are of central importance to children and young people across their entire life courses.

There are, however, a number of areas where the CFJDS could be enhanced, rendering it an extremely powerful data source for understanding the Family Justice System more holistically. Some of these relate to the CFJDS as a whole, some to the individual datasets that comprise it.

The CFJDS as a whole

Parents. A major omission from the CFJDS are data on the parents. Although the CFJDS is principally about children, the lives of children cannot fully be understood without reference to their caregivers, especially their parents: this was emphasised both in the FJO data report (Jay et al, 2017, recommendation 3) and the Family Justice Review Interim Report (MoJ, 2011). This is especially true in the context of child protection and parental disputes about children (i.e. the vast bulk of children law) where, for example, at least 24% of mothers return to court within seven years of their first care case, usually with a different child (Broadhurst et al 2017). A focus only on the child therefore ignores crucial drivers of service use: the circumstances of the caregivers. Data on adult caregivers (principally mothers and fathers, but also others) are available in Cafcass with sufficient quality for linkage to create family cohorts and research.

Information about linkage. As the CFJDS is a linked dataset, certain information, such as linkage steps and rates, is required in order to fully understand the data and evaluate risks of bias that the linkage process might have introduced into analysis. Collaboration between academia and data providers was recommended in the FJO data report in order to improve reporting of analyses relying on linked data (Jay et al, 2017, recommendation 6). In particular, data providers and researchers need to ensure that the Guidance for Information about Linking Data Sets (Gilbert et al, 2018) be followed and this requires the data providers in particular to share information about linkage.

External data. A question arose as to the possibility to link in external, publically available data such as school-level Index of Multiple Deprivation (IMD) scores or school-level rates of Free School Meals (FSM). Socioeconomic circumstances would have been of particular interest to our analyses but the FSM and IMD variables from the NPD were not included in the CFJDS. As the school-level Unique Reference Number was available, a

solution would have been for us to have obtained the school-level rate of FSM and to have created school-level indices of deprivation using publically available data from the DfE (DfE, no date) and the Ministry of Housing, Communities and Local Government (Ministry of Housing, Communities and Local Government, 2015). These indices, which pose no identification risk for children, could then have been imported into the Data Lab to significantly enhance analyses at no extra cost. Other analysts may wish to include other publically available data at aggregated levels that do not risk disclosure of identifiable information, such as LA spending, area-based health indices or geographical indicators. However, this was not possible and the inability to include this information impoverishes analyses.

On-going dialogue. Analysis using administrative data from public bodies requires continuous dialogue between researcher and data provider as the data can rarely be understood fully on their own, even with high quality metadata. In particular, FamilyMan is not routinely accessed by external analysts. Therefore, there is little user knowledge of this data set outside of the MoJ. The CFJDS, of course, is in a unique position in that it consists of data from three separate providers. Nonetheless, researchers need input from data providers to understand how the data are generated and their context. Such dialogue between researchers and data providers is also of direct benefit to data providers and the services they represent as potential improvements to data collection and quality can be highlighted and built upon. We therefore strongly encourage continued dialogue between researchers and each of the data providers involved in the CFJDS. Examples of such interchange could include the Children’s Social Care Data User Group, seminars, webinars and user support groups.

Metadata. The metadata provided give a basic description of the variables included. Some minor comments on the metadata:

- Further information on known issues with the data (for example the issue with end dates in FamilyMan identified in Figure 2, above) should be highlighted within the metadata document.
- The different ID variables should be more clearly identified. The “id_supplied” variable in DfE, in particular, is ambiguous (is it, e.g., the MOJ_ID or something else?).
- As noted in the methods chapter, the SEN variable supplied (primary SEN type) is not the variable documented in the metadata (SEN provision).

Our recommendations on the CFJDS as a whole are therefore:

- **Recommendation 4:** To improve the evidence base within family justice, data could be included on the caregivers that are collected in Cafcass and FamilyMan and that can be linked to children in order to build family cohorts.
- **Recommendation 5:** To ensure transparency and clarity around the data linkage undertaken to create the data share, the MoJ could provide documentation for researchers in line with the Guidance for Information about Linking Data Sets (Gilbert et al, 2018).

- **Recommendation 6:** To improve research output quality, mechanisms could be put in place to allow for external, publically available and anonymous data (such as deprivation indices) to be imported into the data analysis environment.
- **Recommendation 7:** Continued dialogue between researchers and each of the data providers involved in the CFJDS could be supported and encouraged for the mutual benefit of the data provides and researchers. Examples of such interchange could include the Children’s Social Care Data User Group, seminars, webinars and user support groups.
- **Recommendation 8:** The following minor improvements could be made to the metadata:
 - Further information on known issues with the data should be highlighted within the metadata document.
 - The different ID variables should be more clearly identified. The “id_supplied” variable in DfE, in particular, is ambiguous.
 - Mismatches between the metadata and data supplied should be corrected. For example, the SEN variable supplied (primary SEN type) is not the variable documented in the metadata (SEN provision).

Individual datasets

Summary. Given that FamilyMan has not been previously shared for external research, its inclusion in the CFJDS renders the latter particularly attractive as a research data source. It is the only dataset, as far as we are aware, that provides information on both private and family court cases from start to finish; the Cafcass dataset, by contrast, only records cases in which they are involved and, for the majority of private cases, this involvement ends after the first hearing. However, the structure of the datasets as provided, as well as certain omissions, which are outlined below, mean that the full potential of FamilyMan cannot be realised.

Unfortunately we did not examine any NPD datasets other than the censuses as the information contained in them was beyond the scope of our study. We cannot therefore comment on them.

- **Recommendation 9:** Our recommendations concerning the datasets are outlined in the following table.

Limitation	Dataset(s) affected	Recommendation
<p>9(i) Not all children and cases in FamilyMan and Cafcass appear to be in the CFJDS, especially given the shortfall of cases reported in the Results chapter compared with the routinely published figures. This is possibly due to the fact that only children who linked between FamilyMan and Cafcass are included. However, all children could be included as researchers may want to use each database separately. Those interested in linkages would also benefit from having the entirety of each database as information on children and cases who do not link assists in evaluating linkage and in understanding the groups of children under study.</p>	FamilyMan and Cafcass	All children and cases in FamilyMan could be included in the CFJDS, regardless of linkage status.
<p>9(ii) Age is a crucial variable for any analysis involving individuals. However, the way in which age is recorded in the CFJDS hinders analysis. The age variable is stored in the table “APP_DISP_JOIN” and cannot be linked to the other data tables for analysis because it uses a different child-level identifier. Further, age is categorised (e.g. <1 year old, 1-4 years old, etc.) whereas approximate date of birth is often more useful in longitudinal analyses (Note: exact date of birth is rarely needed other than for linkage).</p>	All	For each child, month and year of birth (but not day) could be provided. This will give analysts the flexibility to compute approximate ages as necessary at different events (e.g. case start or case end) or to form their own categorised age variable. Providing birth information in this way is less disclosive than date of birth but still enables full multivariable analysis with age. An alternative would be to compute age in integer years at application date.
<p>9(iii) There are several different child-level identifiers across the FamilyMan files.</p>	FamilyMan	To ensure complete linkage, a consistent child-level identifier could be used across files from the same database.
<p>9(iv) Not all court events were included in the CFJDS: only earliest and latest hearings within a case. Our original plan was to examine the complexity of cases by reference to number and timing of events in cases as this information is not routinely recorded for all private family law cases in existing data resources available to researchers (e.g. Cafcass). Access to this data would enable researchers to produce their own indicators of, for example, different combinations of applications and orders on a per-child and/or per-case basis.</p>	FamilyMan and Cafcass	Complete data from FamilyMan and Cafcass (i.e. all records by individual and court event) could be provided in long format.
<p>9(v) A serious limitation of the CLA data is that only one episode per year is included. This undermines the utility of the linked CLA data in analyses using this data share. Research has shown an enormous degree of variability between children in the number of episodes that they have within a period of care (e.g. some children will have multiple episodes in different placements or on different legal statuses). The current format of the data precludes accurate longitudinal analysis, which stands to be a key feature of this dataset. For example, the proportion of children who have a section 20 episode or who enter care under police protection is likely to be underestimated (see Results chapter).</p>	CLA	All episodes in the CLA dataset could be included.
<p>9(vi) The CLA data is missing key dates (such as episode start date), therefore precluding longitudinal analysis. Available only are episode end date and period of care length. It should be noted that a “period of care” is a period of one or more temporally contiguous</p>	CLA	Episode start and end date could be included so that periods of care, and their

Limitation	Dataset(s) affected	Recommendation
episodes; a new episode begins when there is a new legal status or new placement. Therefore, the episode end date minus the period of care does not reliably give either the episode start date or the period of care start date. In addition, we know from our own analyses with the complete CLA data that the length of period of care variable is erroneous and this can be demonstrated also to be the case within the CFJDS CLA data.		duration, can accurately be determined by the analyst.
9(vii) There is a large amount of missing data in the NPD censuses. For example, primary SEN type is completely missing from 2013 onwards.	NPD censuses	The missing data in the NPD censuses could be investigated and rectified.
9(viii) The NPD censuses and CLA data do not have the same temporal coverage as the other datasets.	NPD censuses and CLA	The censuses and CLA data could cover the same period so as to enable full longitudinal analysis.
9(ix) Important census variables, notably the income domain affecting children index and FSM eligibility are absent, as is the SEN provision variable.	NPD censuses	The income domain affecting children index, FSM eligibility and SEN provision variables could be requested and included in the CFJDS.

Feedback questions

Below are the specific questions asked by Lisa Robinson, with our answers in italics.

I would appreciate any comments on:

- i) the communications pre arrival to the MoJ?

Applying for access to data was very straightforward, as was organising days to visit. We would imagine that most researchers would appreciate booking significant amounts of time in the Data Lab in advance. I was able to proceed flexibly owing to our particular funding arrangements but this may not always be the case.

- ii) their experience whilst at the MoJ?

I felt very welcome. All staff were helpful and the working environment was pleasant.

- iii) In terms of data availability, did you think that the variables provided sufficient information for your project? Are there any variables that were not included that would have been useful either for this project or potential future MDL projects?

We consider that there are some serious limitations in terms of variables and population coverage. See above.

- iv) How useful did you find the metadata that accompanied the datasets? Was there any information missing that would have been useful to have at the start/before the start of your project?

It was very useful to have access to a variable list before accessing the Data Lab. However, we found some discrepancies between the metadata and the CFJDS and a missed opportunity to include information on limitations within the data. See above for details.

- v) The time period for the CFJDS data is 2010-2015 – how useful would an annual refresh of this information be? Would a less frequent update still make this data fit for purpose? If so, how frequent (for example, would an update every 3 years suffice)

Annual refreshes and routine linkages would be extremely useful. This time period would be satisfactory for some projects but too limited for others.

- vi) receipt of the work post visit?

This was very straightforward and quick.

References

- Bainham A (2013) 'Private and public children law: an under-explored relationship.' 25(2) *Child and Family Law Quarterly* 138.
- Broadhurst K, Alrouh B, Yeend E, Harwin J, Shaw M, Pilling M, Mason C and Kershaw S (2015) 'Connecting events in time to identify a hidden population: birth mothers and their children in recurrent care proceedings in England.' 45(8) *British Journal of Social Work* 2241.
- Cafcass (2018) <https://www.cafcass.gov.uk/about-cafcass/research-and-data/private-law-data/> (accessed 4 Jan 2019).
- DfE (no date) 'Get information about schools.' Available from <https://get-information-schools.service.gov.uk/> (accessed 7 Jan 2019).
- Gilbert R, Lafferty R, Hagger-Johnson G, Harron K, Zhang L, Smith P, Dibben C and Goldstein H (2018) 'GUILD: GUIDance for Information about Linking Data sets.' 40(1) *Journal of Public Health* 191.
- Halliday E, Green R and Marsh B (2017) 'Private law cases that return to court: a Cafcass study.' Available from <https://www.cafcass.gov.uk/about-cafcass/research-and-data/cafcass-research/> (accessed 4 Jan 2019).
- Jay MA, Woodman J, Broadhurst K, Gilbert R (2017) 'Who cares for children? Population data for family justice research.' Available from <http://wp.lancs.ac.uk/observatory-scoping-study/> (accessed 7 Jan 2019).
- Mc Grath-Lone L, Dearden L, Nasim B, Harron K and Gilbert R (2016) 'Changes in first entry to out-of-home care from 1992 to 2012 among children in England.' 51 *Child Abuse & Neglect* 163.
- Ministry of Housing, Communities and Local Government (2015) 'English indices of deprivation 2015.' Available from <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015> (accessed 7 Jan 2019).
- Moj (2011) 'Family justice review: interim report.' Available from <https://www.gov.uk/government/publications/family-justice-review-interim-report> (accessed 7 Jan 2019).
- Moj (2018) 'Family Court Statistics Quarterly.' Available from <https://www.gov.uk/government/statistics/family-court-statistics-quarterly-july-to-september-2018> (accessed 4 Jan 2019).

Appendix I – data cleaning

Cafcass E (cases)

- Cafcass E dataset loaded. N = 4,621,825
- Retain only applications in CYs 2011 – 2016 inclusive (maximum FamilyMan application date is in June 2017, Cafcass in September 2017). New N = 2,175,929.
- Drop missing application type. New N = 2,175,820
- Application type rationalisation:

Original	New	n	Major new
Adoption agency (CFR), Adoption agency (CFR) (R73), Adoption agency (Guardian), Adoption agency (Guardian) (R59), Adoption Agency (RO), Adoption agency (RO) (R69), Adoption intercountry (out) (s84), Adoption Intercountry (s50/s51), Adoption non-agency (CFR), Adoption non-agency (CFR) (R73), Adoption non-agency (Guardian), Adoption non-agency (Guardian) (R59), Adoption non-agency (RO), Adoption non-agency (RO) (R69), Adoption Part 10 (FPR 2005), Guardian adoption, Guardian freeing, Placement (guardian), Placement (guardian) (R59), Placement (RO), Placement (RO) (R69), R.O. Adoption, R.O. Freeing	Adoption/Placement	77,787	Adoption/Placement
Care (s31)	CO	311,667	Public
Supervision (s31)	SO	20,913	Public
EPO (s44), Child Assessment (s43)	Emergency	25,147	Public
Secure Accommodation (s25)	SecAccom	4,885	Public
CAO – Spend Time With, CAO – Live With, Contact (s8), Prohibited Steps (s8), Specific Issue (s8), Residence (s8)	S8	1,504,493	Private
Special Guardianship (s14A)	SGO	21,748	Private
Appeal	Appeal	1,539	Dropped
Application (CA (s91(14))), Application for a FGM Protection Order (FGMA), Archived Case, Article 11(7) of Brussels 2R, Cafcass Plus, Change of Surname (s13), Change of Surname (s33(7)(a)), Child Abduction & Custody Act, Committal Application, Contact (s26), Contact (s34), Contact Warning Notice (C78 form), Contact with Adoption Order (s51), Declaration of Legitimacy (s56 FLA), Declaration of Parentage (s55(a)), Deprivation of Liberty, Direction Assessment (s38(6)), Discharge of Care Order (CA(s39(1))), Discharge of Special G'ship Order (CA(s14(D))), Discharge of Supervision Order (CA (s39(2))), Discharge or Vary of CO or SO (s39), Discharge Order, Education SO (s36), Enforcement - C79, Enforcement Order, Enquiry from court under s41 MCA 1973, Enquiry from Foreign Court, EPO Extension (s45), FAO (s16), Forced Marriage	Other	207,641	Dropped

<p>Protection Order (s63A FLA), FPR - Rule 9.5, Independent Reviewing Officer (IRO), Inherent Jurisdiction, Inherent Jurisdiction/Medical, Inherent Jurisdiction/Reporting Restriction Order, Inherent Jurisdiction/Stranded Spouse, Inherent Jurisdiction/Wardship, International Adoptions (In), Leave to Apply Section 8 Order/Form, Leave to Disclose, Leave to Withdraw, Leave/Permission to Apply-Other, Leave/Permission to Oppose Adoption Order, Leave/Permission to Revoke Placement Order, Non-Molestation Order (FLA s42), Non-Subject Child Protection Referral Made, Occupation Order (FLA s33), Other Agency Information Requests, Parental Order (s30 HF&E), Parental Order (s54 HF&E), Parental Responsibility (s4), Parental Responsibility (step-parent) (s4A), Parenting Plan Meeting, Pre-Court Care Application, Pre-Court Consent Adoption (s20), Pre-Court Consent Placement (s19), Pre-Court Consent Placement(s19)/Adoption(s20), Pre-Court Work, Pre-Proceedings SP Pilot, Recovery Order (FLA 1986 s34), Recovery Order (s50), Removal from Jurisdiction (s13), Removal from Jurisdiction (s33(7)(b)), Remove from Current Placement (ACA s38(5)), Remove the Child from the United Kingdom (s28 ACA), Revocation of Freeing Order, Revocation of Placement (s24), Revoke Contact (s26), Secure Accommodation (s25), Terminate Contact (s34(4)), Variation of Special G'ship Order (CA(s14(D))), Variation of Supervision Order (CA(s39(2))), Wardship Proceedings</p>			
--	--	--	--

- Rule 16.4 and s 37 identified from final order.
- S 7 identified from s7Ordered variable. This identified 650,604 rows as s7. Adding the “s7 ordered” in final legal order raises this number to 711,312.
- Drop non-subjects. New N = 1,138,048.
- Drop appeals and other applications. New N = 1,032,365.
- Redundant variables dropped. Deduplicated on all remaining variables. New N = 1,026,495.

Cafcass A (demographic spine)

- Cafcass A dataset loaded. N = 978,353.
- Deduplicate on Cafcass_ID, Gender, Ethnicity. New N = 880,599.
- Create major ethnicity groups and deduplicate again. New N = 876,574.
- Some duplicates remain due to inconsistent ethnicities and genders. These children could not be found in FamilyMan. Therefore an ethnicity and gender for each child was taken at random using an appropriate bimodal or multinomial distribution. Then deduplicated. New N = 856,098.

FamilyMan

- Load FM_Wide. N = 767,554.
- Load APP_DIPS_JOIN. N = 15,337,492.
- Load FM_Long. N = 2,100,334.
- Appear to be no duplicates (across all variables)
- Drop rows with missing app dates
 - FM_Wide new N = 767,551.
 - APP_DISP_JOIN new N = 15,337,375.
 - FM_Long new N = 2,100,326.
- Retain only CY 2011 to 2016 in FM_Long. New N = 1,067,454.
- Retain only rows in FM_WIDE where cyear 2011-2016, case number and mojid in FM_LONG. New N = 388,143.
- Retain only rows in APP_DIPS_JOIN where CASENUMBER in FM_LONG. New N = 4,534,077.
- Deduplicate on all variables
 - FM_LONG New N = 767,863.
 - APP_DISP_JOIN New N = 4,532,505
 - FM_WIDE – no duplicates
- Work out min and max application dates in APP_DISP_JOIN and drop <2011. New N = 4,519,012.

Census

- Load. N = 708,591.
- Deduplicate on c("Pup", "Act", "Aca", "URN", "Gen", "Eth", "NCY", "NCy", "Pri"), new n = 705,257.
- reshape to long, new n = 9,873,598.
- Only goes up to 2014/15

SEN data linked to Cafcass in the following years

Index year	Census years
2011	2010/11 to 2011/12
2012	2011/12 to 2012/13
2013	2012/13 to 2013/14
2014	2013/14 to 2014/15
2015	2014/15
2016	Nil

Note however that no SEN data linked to cases in index years 2014 to 2016. There is a considerable amount of missing data in the censuses.

CLA cleaning

- Load. N = 380,195.
- Deduplicate on all vars, new n = 380,190.
- Subset to 1 Jan 2011 onwards only, new n = 266,807.

Appendix II – ethnicity data

Table AII.1. Ethnicity of children with index cases in each calendar year per Cafcass with proportions given assuming that data are missing completely at random

	2011	2012	2013	2014	2015	2016
N children (total)	59299	63191	65017	48171	48132	52493
N (complete ethnicity data)	11352	10466	10089	9630	22888	37005
Ethnicity						
White	9287 (81.8%)	8351 (79.8%)	7731 (76.6%)	7135 (74.1%)	17238 (75.3%)	28355 (76.6%)
Black	316 (2.8%)	287 (2.7%)	377 (3.7%)	432 (4.5%)	890 (3.9%)	1425 (3.9%)
Asian	645 (5.7%)	694 (6.6%)	714 (7.1%)	790 (8.2%)	1876 (8.2%)	2748 (7.4%)
Mixed	952 (8.4%)	989 (9.4%)	1118 (11.1%)	1101 (11.4%)	2515 (11.0%)	3891 (10.5%)
Other	152 (1.3%)	145 (1.4%)	149 (1.5%)	172 (1.8%)	369 (1.6%)	586 (1.6%)

UCL Legal Epidemiology Group

UCL Great Ormond Street Institute of Child Health
Population, Policy & Practice Programme
30 Guilford Street
London WC1N 1EH

<https://www.ucl.ac.uk/child-health/research/population-policy-and-practice/child-health-informatics-group/legal-epidemiology-group>