What Works *for* Children's Social Care

KINSHIP AND RESIDENTIAL CARE

Technical Appendix

What Works for Children's Social Care

Technical Appendix

Data cleaning

Data cleaning, manipulation and analysis was conducted using R version 4.0.2 in R Studio version 1.4.1717.

Data linkage

Children who had episodes of care registered in the Children Looked After Dataset (SSDA903 collection) formed the primary population which information in other datasets was linked to. The primary identifier variable used was the Pupil Matching Reference (PMR), a nationally unique anonymised identifier based off of a child's Unique Pupil Number (UPN) assigned to children upon first entry to a state funded school (or the creation of an Education, Health and Care Plan if this is sooner) (Jay, McGrath-Lone & Gilbert, 2018). This enabled us to link children's care records to extracts of the National Pupil Database (NPD).

Where a PMR was not available, a unique child identifier was assigned by concatenating the Department for Education's unique child identifier with the Local Authority (LA) 3-digit code who the child is in the care of. While the concatenated ID is recommended by the Department for Education (DfE) to improve the matching results, this means that if a child moves to a different LA we are unable to link their records if they do not have a PMR. The tables below show the proportion of children for whom a PMR could be matched for within our cohort samples (children who turned 18 in 2019/20 and experienced types of care), and thus the maximum number of children educational data can be linked for.

Table 1: PMR matching for cohort samples used in the residential care report

Typology	Percentage of children who had an identifier linkable to the NPD		Resulting sample size	
	Children who turned 18 in 2019/20 and ever lived in residential care	Children who turned 18 in 2019/20 and were ever in care	Children who turned 18 in 2019/20 and ever lived in residential care	Children who turned 18 in 2019/20 and were ever in care
UASC	47%	21%	50	610
Children who entered care due to disability/illness	98%	93%	360	630
Early entrants to residential care (10 and under)	88%	88%	170	170
Adolescent entrants to care (11-15)	98%	93%	1,450	5,040
Late entrants to care (16+)	85%	69%	120	2,210
Early entrants to non-residential care	98%	58%	600	5,840
Overall	**95%**	**35%**	**2,770 (95% of full sample)**	**14,510 (65% of full sample)**

Table 2: PMR matching for cohort samples used in the kinship care report

	Percentage of children who had an identifier linkable to the NPD	Resulting sample size
Children who turned 18 in 2019/20 and ever lived in kinship foster care	82%	**3,020**
Children who turned 18 in 2019/20 and ever left care to a kinship special guardian	96%	**690**
Children who turned 18 in 2019/20 and were ever in care	65%	**14,510**

School census data

School Data Limitations

We use extracts from the NPD linked to children who have recorded episodes of care in the Children Looked After Dataset. Records are linked from the CLA Dataset to the NPD using the PMR, therefore where a child does not have a PMR school data is unavailable. This includes children who did not have an assigned UPN whilst they were ever in the care of a LA: this could include children in care who have not yet started school up to the most recent data (e.g. if a child was age three in 2019/20), children who left care for the final time before they started school, or children who – regardless of age – never entered state funded education. Interpretation of school data is therefore restricted: tables in the report appendices show the data availability for samples used throughout the analysis.

Furthermore, we expect missing information to arise from the particular NPD extracts requested. The schools' census was requested for child characteristics such as special educational needs, free school meal eligibility and IDACI, however our data request only included the mainstream schools' Spring Census. This excludes Pupil Referral Units (PRUs) for years 2009/10–2012/13 (Jay, McGrath-Lone & Gilbert, 2018) and Alternative Provision (AP) institutions for years 2007/08–2019/20, which have separate censuses. Children in care are over-represented in AP by a factor of 29 relative to their population size in the wider school population with 4,422 children in care, on average, attending PRUs or AP in the 2011/12–2013/14 academic years (Malcolm, 2018). As a result, we expect to undercount and have missing characteristic information within our samples for a significant number of children who are particularly vulnerable to poor outcomes (DfE, 2016).

SEN

We used the primary SEN type, reporting a child's most significant need at the point of assessment, as recorded in the schools' Spring Census: this variable is measured for children who are recorded as having a SEN statement (up to 2014/15) or Educational, Health and Care (EHC) plan (since 2014/15), or a school/early years' action/action plus (up to 2014/15) or SEN support (since 2014/15) (DfE, 2015a), therefore is only filled for children who are formally assessed as having special educational needs. The primary SEN type is a child's most "significant" need: children may have an additional secondary need recorded within the schools' census (DfE, 2015b) which is not considered in the current study; this could affect 21.9% of children who are recorded as having a secondary SEN type in addition to a primary SEN type (Lindorff & Strand, 2018). It should also be noted that children may have different primary SEN types recorded over different years of the schools' census.

Absences

Absences data obtained through the NPD records a child's attendance for each academic year. Children have an attendance record for each school they are enrolled in throughout the year. Therefore, to determine the rate of unauthorised absences per academic year, records had to be consolidated where a child was enrolled in multiple schools. This was obstructed by potentially erroneous multiple records which could not be definitively distinguished from genuine instances where a child 'should' have had multiple records, e.g. where a child had multiple records with the same number of total sessions possible but different numbers of recorded absences over different rows, or where the total number of sessions per year for a

child summed to an infeasible number. To avoid losing potentially valuable information, no records were changed/corrected or deleted. We assumed that errors were random, therefore the variables ultimately derived would not be biased because they were proportions.

Based on yearly attendance data, we calculated attendance throughout each school Key Stage. Where a child did not have an annual record, their key stage attendance was estimated from the years in which they *did* have a record. For example, Key Stage 3 (KS3) encompasses Year 7, 8 and 9: if a child had no attendance record for Year 7, their KS3attendance would be calculated from their Year 8 and Year 9 attendance. This approach could systematically overestimate or underestimate unauthorised absences depending on the reasons for attendance records being missing.

Exclusions

Exclusions data obtained through the NPD recorded every instance of a fixed term or permanent exclusion a child had recorded within an academic year; this was aggregated to an annual and key stage level to determine the total number of sessions a child was excluded from, over all fixed term exclusions and the number of permanent exclusions a child was subject to over each time period. Importantly, an exclusions record is only available per academic year if any exclusions have been recorded for a child in that particular year: we have taken the assumption that a child has never been excluded (in a particular year or every year) if they have a record in the schools' Spring Census but not in the exclusions data.

Regression analysis

Prior to regression analysis, the following strategies were used to correct missing data for both research questions described below:

- For categorical variables recorded by the Schools' census, missing values were re-coded to 'Inapplicable' if a child was below school age when they entered care for the first time, and 'Missing' otherwise.
- Multiple imputation was performed on missing continuous covariates (IDACI, sessions missed due to fixed exclusions at KS2, unauthorised absences at KS2) using predictive mean matching within the 'mice' package in R.[1]

Research Question 1b: Predicting entry to residential care in a child's first period of care

We perform binary response logistic regressions to identify factors associated with entry to residential care in a child's first period of care, using information available up to the point of entering care for the first time. Using a given set of predictors, we can predict the probability of each outcome variable. Analysis was performed on every combination of sample and covariate specifications which are detailed below. The outcome variable is a binary variable indicating whether or not a child had a care placement in a children's home subject to Children's Home Regulations during their first period of care.

The main model was produced by performing regression analysis on the sample of all children who turned 18 in 2019/20 and who were ever in care (n=22,431). Additional samples restricting the age at which children entered care for the first time were used for robustness given some of the covariates used were only available for a subset of children: the sample of children who turned 18 in 2019/20 and entered care for the first time when they were at least in school year Reception (n=16,487), and the sample of children who turned 18 in 2019/20 and entered care for the first time when they were at least in school Year 7 (n=11,574).

Regression analysis was performed on various covariate specifications which used different sets of predictor variables. Table 1 below shows which predictors were included in each specification.

Table 3: Different specifications used as the set of predictor variables for RQ1b

	Specification 1	Specification 2	Specification 3 (main specification)
Gender	✔	✔	✔
Ethnicity	✔	✔	✔
Age at first entry to care	✔	✔	✔
Primary need for services	✔	✔	✔

[1] More detail can be found here: https://cran.r-project.org/web/packages/mice/mice.pdf

UASC status	✔	✔	✔
IDACI in most recent school census before entering care		✔	✔
Primary SEN type in most recent school census before entering care		✔	✔
Ever eligible for FSM up to the point of entering care		✔	✔
Unauthorised absences in KS2			✔
Ever permanently excluded in KS2			✔
Ever received fixed term exclusion in KS2			✔

Please see Understanding formal kinship care arrangements in England - Technical Report for regression tables.

Research Question 1c: Predicting entry to residential care after the first placement in a child's first period of care using information from the first placement

In a second stage, we perform binary logistic regressions to identify factors associated with entering residential care in a child's first period of care if their first episode within the period of care was *not* residential care. This used information available up to the point of entry to care, as well as information from the child's first placement within the period of care, to predict whether the second or subsequent placement would be in a residential home subject to Children's Homes Regulations.

The outcome variable is a binary variable indicating whether or not a child had a care placement in a children's home subject to Children's Home Regulations in a second or subsequent placement in their first period of care. The model was produced by performing regression analysis on the sample of all children who turned 18 in 2019/20, were ever in care and had at least two care placements in their first period of care where their first placement was *not* in a children's home subject to Children's Home Regulations (n=8,343).

Prior to regression analysis, an additional step was taken to prepare the multiply imputed datasets: coarsened exact matching using 'cem' in R[2] was applied to create statistically equivalent treatment and control groups, using variables which were statistically in RQ1b. This included the following variables:

- Primary SEN type
- Major ethnic group
- Category of need upon care entry
- Age at entry to care
- UASC status
- IDACI at most recent school census
- Ever eligible for free school meals up to the most recent census prior to entering care
- Fixed exclusions at KS2

Observations which could not be matched were dropped from analysis, resulting in a final sample size of n=8,343.

Regression analysis was performed on the matched sample using the following covariates:

- Length of first placement episode[3] (in days)
- Type of first placement (e.g. in a residential care home, foster placement)
- Whether the first placement was in or outside of the LA (0=inside LA boundary, 1=outside LA boundary)
- Reason for placement change

[2] More detail can be found here: https://cran.r-project.org/web/packages/cem/cem.pdf

[3] An episode of care forms part of a period of care that consists of one or more episodes of care. A new episode of care is started when a child becomes looked-after, when there is a change of legal status, when there is a change of placement, or when the placement provider changes, although we will not count changes solely due to a change in legal status as a change of placement. More information can be found under: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974676/Children_looked-after_by_local_authorities_in_England_2020_to_2021.pdf

Regression tables can be found in:

- Understanding formal kinship care arrangements in England: Technical report
- Understanding residential care for children in care in England - analyses of administrative data: Technical Report

Kinship care codes

Kinship foster care placements are identified in the Children Looked After Dataset by specific placement codes. This analysis encompassed 20 years' of administrative data and consequently multiple iterations of codes indicating kinship foster care. The codes used to identify kinship foster care were: F1, F4, Q1, Q2, U1, U2, U3 and F9.

Within the Children Looked After Dataset, codes which indicate where a child left care to a special guardian have changed over time. For the 2005/06 and 2006/07 reporting years, no information was provided about the guardian (reason episode ceased code E42). All SGOs granted in this time period have been included in analysis: given the proportion of SGOs granted to non-kin carers in 2019/20, we estimate to wrongly include c.11% or approximately 100 children leaving care to a non-kin special guardian in this time period (Wade et al., 2014).

For 2007/08–2017/18, codes indicate whether or not an SGO was granted to a former foster carer. Where an SGO was granted to a former foster carer (code E43), a child is included in analysis if the foster care placement they left care from was a kinship foster care placement. All special guardianships granted to someone other than a former foster carer in 2007/08–2017/18 (code E44) were included in analysis: in 2019/20 kin made up 93% of all SGOs made by individuals who were not previously foster carers, therefore we estimate to wrongly include c.7% of SGOs granted (approximately 880) in 2007/08–2017/18 to individuals who were not previously foster carers.

For 2018/19 and 2019/20, codes available inform 1) whether an SGO was granted to a former foster carer, and 2) whether an SGO was granted to kin. We include special guardianships granted to kin, regardless of whether they were or were not a former foster carer (codes E45 and E47).

R packages used

Backports:
Lang M, R Core Team (2020). _backports: Reimplementations of Functions Introduced Since R-3.0.0_. R package version 1.2.1, <URL:https://CRAN.R-project.org/package=backports>.

Car:
Fox J, Weisberg S (2019). _An R Companion to Applied Regression_, Third edition. Sage, Thousand Oaks CA. <URL:https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.

CarData:
Fox J, Weisberg S, Price B (2020). _carData: Companion to Applied Regression Data Sets_. R package version 3.0-4, <URL:https://CRAN.R-project.org/package=carData>.

Cem:
Iacus S, King G, Porro G, Nielsen R (2021). _cem: Coarsened Exact Matching_. R package version 1.1.29, <URL:https://CRAN.R-project.org/package=cem>.

Crayon:
Csárdi G (2021). _crayon: Colored Terminal Output_. R package version 1.4.1, <URL: https://CRAN.R-project.org/package=crayon>.

Data table:
Dowle M, Srinivasan A (2021). _data.table: Extension of `data.frame`_. R package version 1.14.0, URL:https://CRAN.R-project.org/package=data.table>.

Digest:
Lucas DEwcbA, Tuszynski J, Bengtsson H, Urbanek S, Frasca M, Lewis B, Stokely M, Muehleisen H, Murdoch D, Hester J, Wu W, Kou Q, Onkelinx T, Lang M, Simko V, Hornik K, Neal R, Bell K, de Queljoe M, Suruceanu I, Denney B, Schumacher D, Chang. aW (2020). _digest: Create Compact Hash Digests of R Objects_. R package version 0.6.27, <URL:https://CRAN.R-project.org/package=digest>.

Dplyr:
Wickham H, François R, Henry L, Müller K (2021). _dplyr: A Grammar of Data Manipulation_. R package version 1.0.6, <URL:https://CRAN.R-project.org/package=dplyr>.

Farver:
Pedersen T, Nicolae B, François R (2021). _farver: High Performance Colour Space Manipulation_. R package version 2.1.0, <URL:https://CRAN.R-project.org/package=farver>.

Forcats:
Wickham H (2021). _forcats: Tools for Working with Categorical Variables (Factors)_. R package version 0.5.1, <URL:https://CRAN.R-project.org/package=forcats>.

Ggplot2:

Wickham H (2016). _ggplot2: Elegant Graphics for Data Analysis_. Springer-Verlag New York. ISBN 978-3-319-24277-4, <URL:https://ggplot2.tidyverse.org>.

Haven:
Wickham H, Miller E (2021). _haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files_. R package version 2.4.3, <URL:https://CRAN.R-project.org/package=haven>.

Hms:
Müller K (2021). _hms: Pretty Time of Day_. R package version 1.1.1, <URL: https://CRAN.R-project.org/package=hms>.

Labeling:
Talbot, J (2020). _labeling: Axis Labeling_. R package version 0.4.2, <URL: https://CRAN.R-project.org/package=labeling>.

Lmtest:
Zeileis A, Hothorn T (2002). "Diagnostic Checking in Regression Relationships." _R News_, *2*(3), 7-10. <URL:https://CRAN.R-project.org/doc/Rnews/>.

Mice:
van Buuren S, Groothuis-Oudshoorn K (2011). "mice: Multivariate Imputation by Chained Equations in R." _Journal of Statistical Software_, *45*(3), 1-67. <URL: https://www.jstatsoft.org/v45/i03/>.

Miceadds:
Robitzsch A, Grund S (2021). _miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'_. R package version 3.11-6, <URL: https://CRAN.R-project.org/package=miceadds>.

Openxlsx:
Schauberger P, Walker A (2021). _openxlsx: Read, Write and Edit xlsx Files_. R package version 4.2.4, <URL:https://CRAN.R-project.org/package=openxlsx>.

Plyr:
Wickham H (2011). "The Split-Apply-Combine Strategy for Data Analysis." _Journal of Statistical Software_, *40*(1), 1-29. <URL:http://www.jstatsoft.org/v40/i01/>.

Rcpp:
Eddelbuettel D, François R (2011). "Rcpp: Seamless R and C++ Integration." _Journal of Statistical Software_, *40*(8), 1-18. doi:10.18637/jss.v040.i08 (URL: https://doi.org/10.18637/jss.v040.i08).

Eddelbuettel D (2013). _Seamless R and C++ Integration with Rcpp_. Springer, New York. doi: 10.1007/978-1-4614-6868-4 (URL:https://doi.org/10.1007/978-1-4614-6868-4) , ISBN 978-1-4614-6867-7.

Eddelbuettel D, Balamuta JJ (2018). “Extending extitR with extitC++: A Brief Introduction to extitRcpp.” _The American Statistician_,*72*(1), 28-36. doi: 10.1080/00031305.2017.1375990 (URL:https://doi.org/10.1080/00031305.2017.1375990).

Readxl:
Wickham H, Bryan J (2019). _readxl: Read Excel Files_. R package version 1.3.1, <URL: https://CRAN.R-project.org/package=readxl>.

Reshape2:
Wickham H (2007). “Reshaping Data with the reshape Package.” _Journal of Statistical Software_, *21*(12), 1-20. <URL:http://www.jstatsoft.org/v21/i12/>.

Rio:
Chan C, Chan GC, Leeper TJ, Becker J (2021). _rio: A Swiss-army knife for data file I/O_. R package version 0.5.27.

Sandwich:
Zeileis A, Köll S, Graham N (2020). “Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R.”_Journal of Statistical Software_, *95*(1), 1-36. doi:10.18637/jss.v095.i01 (URL: https://doi.org/10.18637/jss.v095.i01).

Zeileis A (2004). “Econometric Computing with HC and HAC Covariance Matrix Estimators.” _Journal of Statistical Software_, *11*(10), 1-17. doi: 10.18637/jss.v011.i10 (URL:https://doi.org/10.18637/jss.v011.i10).

Zeileis A (2006). “Object-Oriented Computation of Sandwich Estimators.” _Journal of Statistical Software_, *16*(9), 1-16. doi:10.18637/jss.v016.i09 (URL: https://doi.org/10.18637/jss.v016.i09).

Stargazer:
Hlavac M (2018). _stargazer: Well-Formatted Regression and Summary Statistics Tables_. Central European Labour Studies Institute (CELSI), Bratislava, Slovakia. R package version 5.2.2, <URL:https://CRAN.R-project.org/package=stargazer>.

Withr:
Hester J, Müller K, Ushey K, Wickham H, Chang W (2021). _withr: Run Code 'With' Temporarily Modified Global State_. R package version 2.4.2, <URL: https://CRAN.R-project.org/package=withr>.

Zip:
Csárdi G, Podgórski K, Geldreich R (2021). _zip: Cross-Platform 'zip' Compression_. R package version 2.2.0, <URL:https://CRAN.R-project.org/package=zip>.

Zoo:
Zeileis A, Grothendieck G (2005). “zoo: S3 Infrastructure for Regular and Irregular Time Series.” _Journal of Statistical Software_,*14*(6), 1-27. doi: 10.18637/jss.v014.i06 (URL:https://doi.org/10.18637/jss.v014.i06).

What Works for Children's Social Care

References

Department for Education (2016) *Educational Excellence Everywhere: March 2016*, Department for Education: London.

Jay, M. A., McGrath-Lone, L. & Gilbert, R. (2018) 'Data Resource: the National Pupil Database (NPD)', *International Journal of Population Data Science*, 4 (1): 1--8.

What Works *for* Children's Social Care

CONTACT

info@whatworks-csc.org.uk
@whatworksCSC
whatworks-csc-org.uk