

Can administrative data be used to research health visiting in England? A completeness assessment of the Community Services Dataset

Amanda Clery^{1,*}, Catherine Bunting¹, Mengyun Liu¹, Katie Harron¹, Jenny Woodman², and Louise Mc Grath-Lone²

Submission History

Submitted:	19/01/2024
Accepted:	28/05/2024
Published:	19/06/2024

¹UCL Great Ormond Street Institute of Child Health, University College London, London, WC1N 1EH

²Thomas Coram Research Unit, Social Research Institute, University College London, London, WC1H 0AA

Abstract

Introduction

Health visiting is a community service provided to families with children under five in England and is a key focus of early years policy. Individual-level data on health visiting is captured in the Community Services Data Set (CSDS), an administrative dataset of publicly funded community services across England. Analyses of CSDS are considered experimental as the dataset matures.

Objectives

In this study, we aimed to identify health visiting contacts in the CSDS and assess the completeness of these data from 2016/17 to 2019/20 compared to external reference data.

Methods

We identified the number of the four mandated postnatal health visiting contacts delivered, excluding those scheduled but not attended, between April 2016 and March 2020. We compared counts by local authority (LA) and financial quarter against the Office for Health Improvement and Disparities' Health Visitor Service Delivery Metrics (HVSDM) to identify a subnational subset of complete CSDS data. We explored the representativeness of this subset.

Results

During the study period, 10.2 million health visiting contacts were delivered to 2.4 million children in England. Of these, we identified 3.9 million mandated contacts based on CSDS codes and age at time of contact, which represented 44.7% of all mandated contacts reported in the HVSDM for the same period. There were 63 LAs with complete CSDS data in at least one quarter, which were broadly representative of English LAs overall. Variables related to staff characteristics were highly missing and only 13 LAs had four or more successive quarters of complete data needed for longitudinal, child-level analyses.

Conclusions

We identified a subnational subset of complete CSDS data, compared to external reference data, which can be used for health visiting research. Until improvements are made to its completeness, analyses (particularly those requiring longitudinal data) may not be generalisable to the whole child population.

Keywords

administrative data; health visiting; data completeness; data representativeness; Community Services Data Set

*Corresponding Author:

Email Address: amanda.clery.14@ucl.ac.uk (Amanda Clery)

Introduction

Health visiting is a community public health service offered to all families with babies and children under 5 years aimed at improving child health and reducing inequalities [1]. The service provides universal support, through mandated contacts, as well as targeted support based on family needs. There are four mandated contacts after birth that all families with children under 5 years in England should be offered at specific ages: the new birth visit (8–14 days), 6–8 week review (42–56 days), 1-year review (9–15 months), and 2-2½-year review (23–30 months) [2]. Health visitors may provide additional contacts to support families with more complex needs or to help with specific challenges such as infant feeding or sleeping.

Describing and evaluating health visiting is currently a key interest for early years policy with a focus on health and development outcomes from conception to age 2 [3, 4]. Several research projects are underway that aim to understand the processes, variation and impact of health visiting in England [5, 6]. To do so requires complete and robust data.

The primary source of information about health visiting services provided to families is CSDS, a national, individual-level administrative dataset of all publicly funded community services delivered to adults and children in England, which has been collected since 2015 [7]. As part of their commissioning contracts from local authorities (LAs), all providers of community services (including NHS Trusts, LAs, and independent providers) are responsible for submitting data to CSDS monthly via NHS England. For various reasons, including incompatibility of computer systems, not all providers of community services submit their data in full to CSDS. For some LAs, neither coverage of the population nor individual data fields are complete, and within an LA, data completeness may vary over time [8]. As such, statistics published using these data have been classified as ‘experimental’, defined by the UK Statistics Authority as ‘a subset of newly developed or innovative official statistics undergoing evaluation’ [9].

These administrative data have the potential to provide public health researchers with a wealth of information about service use and delivery, outcomes and risk factors. However, researchers need to take care in understanding the completeness and accuracy of these data. Previous research assessed the completeness of the number of 2-2½-year reviews recorded in CSDS in 2018/19 compared to the number reported in external reference data, namely the aggregate health visiting metrics published by (ex) Public Health England, now Office for Health Improvement and Disparities (OHID), which found that 33 out of 150 LAs had highly completed 2-2½-year review data [10, 11].

In this study, we aimed to (1) identify all four postnatal mandated health visiting contacts in the CSDS and (2) expand on the previous methods to assess the completeness of these data from 2016/17 to 2019/20 compared to the external aggregate health visiting metrics, to create a study sample for health visiting research.

Methods

Data overview

We received anonymised CSDS data from 2015/16 to 2020/21 via the NHS Data Access Request Service. We excluded 2015/16 as there was a high level of incomplete data in the first year of collection and 2020/21, as patterns of health visiting and data submission were affected by the COVID-19 pandemic.

The CSDS contains information on service user demographics, care activities, referrals, diagnoses, and assessment scores, all of which are recorded in separate linkable tables, consolidated by financial year (April to March). Our study used data from four demographics tables, one per financial year from 2016/17 to 2019/20, containing details of all individuals known to community services within each LA in that year, with a pseudonymised unique person identifier (*Token Person ID*) for each record. We used four care activities tables containing details of all scheduled care activities, including health visiting activities, for each financial year from 2016/17 to 2019/20. Care activities may be face-to-face, over the telephone or via another medium such as email. An activity may have taken place as planned, in which case information may be available on the activity location and duration, or it may be recorded as cancelled or not attended. Care activities scheduled for individual children can be extracted by using their person ID to link the demographics and care activities tables.

Study period and cohort

Using the four demographics tables, we extracted all children aged 5 years and under living in England during the study period (April 2016 to March 2020). Exact date of birth of individuals was not provided, so age eligibility was determined from month and year of birth. Residence in England was defined by the first recorded LA of residence. At the time of analysis, there were 149 upper tier LAs included in CSDS (with Isles of Scilly combined with Cornwall, and City of London combined with Hackney). Children with missing data for LA of residence or recorded as living outside of England were excluded from the cohort. We removed duplicate child records using the unique person ID.

Identifying health visiting contacts

The CSDS care activities table contains records of activities for all publicly funded community services, not just health visiting services. We identified health visiting activities based on the recorded *team type* variable (code 16: health visiting service) and/or *activity type* variable (codes 8-12: health visitor health reviews). Activities with missing data for both team and activity type were excluded. We identified attended activities using the *attend or not* variable (codes 5: attended on time and 6: arrived late but was seen). Activities with missing attendance data were categorised as attended if the duration was recorded as being 5 minutes or longer.

A single care 'contact' between a health visiting team and a child may include multiple care activities. For example, if a health visitor carries out a child assessment and provides advice to parents during the same appointment, these can be included in CSDS as separate activity records, linked by the same care contact ID. We reduced all care activities that were provided to a child within a contact to a single contact record. If there were multiple contacts provided to the same child on the same day, we consolidated these to a single contact record, so that each child had no more than one recorded contact with the health visiting service per day.

Distinguishing mandated and additional contacts

We identified the four mandated contacts (new birth visit, 6-8-week review, 1-year review and 2-2½-year review) using activity codes 8-11 respectively. If a contact record included a care activity of type 8, 9, 10 or 11 we coded the overall contact with the corresponding mandated contact type. As activity codes are subject to human error during data entry, we sense-checked the age at which the mandated contacts occurred to ensure they were plausible, using broader age ranges to allow for mandated contacts that were provided earlier or later than expected (Table 1). Any mandated contacts that fell outside these broad plausible time windows were re-categorised as additional contacts (e.g., a new birth visit recorded for a 14-month-old child was considered implausible and re-categorised as an additional contact). If a child had multiple mandated contacts of the same type recorded within the plausible age range, the earliest mandated contact was retained and others were re-categorised as additional contacts. This ensured that each child was recorded as receiving only one of each mandated contact type.

Due to a low proportion of mandated contacts captured by these activity codes, we also developed a data supplementation approach to identify probable mandated contacts based on the age of the child on the date when the contact took place.

CSDS includes 'date flag' variables that indicate whether a contact took place within a particular age range. For children who did not have a mandated contact recorded, we used date flags to determine whether a contact occurred during a time window in which we would expect a mandated contact to occur (Table 2). For example, if a child did not have a new birth visit recorded in CSDS but had an additional health visiting contact between 8 and 30 days after birth, we re-categorised this contact as the child's new birth visit. If a child had multiple contacts flagged as taking place between 8 and 30 days, the earliest of those contacts was identified as the new birth visit.

Completeness assessment

We assessed whether the health visiting data for a given LA and financial quarter were complete by assessing agreement with reference data, namely the Health Visitor Service Delivery Metrics (HVSDM) published by OHID [12]. These statistics include indicators related to the four postnatal mandated contacts (e.g., the percentage of births that receive a new birth visit within 14 days) and are based on aggregate counts submitted by LAs to OHID on a quarterly basis. In May 2023, we downloaded HVSDM data for Q1 2016/17 to Q4 2019/20 and extracted aggregate counts of the four postnatal mandated contacts [13]. This meant that for each of the four mandated contacts, we had reference data for 2,384 LA-quarters (16 quarters × 149 LAs).

HVSDM are experimental data and the aggregate counts of mandated contacts are not subject to any validation checks. Therefore, before using HVSDM as reference data, we carried out additional validation and cleaning. For example, for the 1-year review, we checked that the number reported by the LA as being completed by age 15 months was larger or equal to the number completed by age 12 months. If the number completed by 12 months was larger, we assumed that the figures had been transposed and corrected the error.

We counted the number of each type of mandated contact in CSDS by quarter for each LA. For each mandated contact

Table 1: Age ranges for determining plausibility of mandated health visiting contacts coded by practitioners in Community Services Dataset, 2016/17 to 2019/20

Mandated contact type	Age range when mandated contact should be offered	Plausible age range specified in our analysis
New birth visit	8–14 days	0–6 months
6–8-week review	42–56 days	0–6 months
1-year review	9–15 months	6–23 months
2–2½-year review	23–30 months	15–40 months

Table 2: Date flag variables in the Community Services Dataset that were used to identify potentially mis-recorded mandated contacts based on child's age at the time of contact

Mandated contact type	Age range included	Date flag variable used for identification
New birth visit	8-14 days	ContactBetween8_14Days_Flag
	2-4 weeks	ContactBetween15_30Days_Flag
6-8-week review	6-9 weeks	ContactBetween42_63Days_Flag
1-year review	9-15 months	ContactBetween270_457Days_Flag
2-2½-year review	23-30 months	ContactBetween691_914Days_Flag

type, we compared the count in CSDS with the count in HVSDM for the same LA in the same quarter and calculated the percentage agreement. If the agreement between CSDS and HVSDM counts was within $\pm 15\%$ for all four mandated contacts, we defined the LA as having 'complete' data for that quarter. The 15% margin has been used in previous health visiting research using CSDS data and is based on the observed difference between counts of mandated contacts recorded in HVSDM and local health visiting activity data held by three LAs [10].

We assessed the representativeness of the LAs we defined as having 'complete' data for at least one quarter by comparing to publicly available national statistics for region, ethnicity, deprivation, and indicators of childhood adversity. We examined levels of missing data for key variables relevant to health visiting activity, including duration, medium and location of contacts, as well as variables related to staff mix.

Results

All counts of children and contacts from CSDS have been rounded to the nearest 5 to comply with NHS statistical disclosure rules for subnational data.

Study cohort

For the four financial years from 2016/17 to 2019/20, the CSDS demographics table contained more than 72 million records, including 17 million records for children under 5 years. Of these, there were 425,340 records (2.5%) with missing LA data, and 20,180 (0.1%) with LAs recorded in other UK nations, which were excluded. Most children were duplicated across multiple demographics tables as they were aged under 5 for multiple years between 2016/17 and 2019/20. We removed 11.8 million duplicate child records, leaving 4.7 million unique children under 5 years living in England for inclusion in the cohort (Supplementary Appendix 1).

Identifying health visiting contacts

There were 59.1 million care activities recorded for our cohort of 4.7 million children. In total, we identified 31.7 million health visiting activities: 29.5 million activities were recorded

as being provided by health visiting service teams and a further 2.2 million had no team recorded but were recorded as one of the mandated health visiting activities. Overall, 24.7 million of these activities were attended. After consolidating activities that occurred on the same day for each child, we identified 10.2 million health visiting contacts attended by 2.4 million children (Supplementary Appendix 1). This represented about half of the eligible cohort of children under 5 years, which was expected given that there are no mandated contacts after age 2.5 years and so many children in the cohort would have had no contact with health visiting services during the study period.

Of the 10.2 million health visiting contacts attended by our study cohort, there were 2.2 million contacts coded as one of the four postnatal mandated contacts (based on *activity type* 8-11). Of these, 25,570 (1.2%) were deemed implausible based on the age at contact and 74,735 (3.5%) were identified as repeated mandated contacts (Table 3). Following re-categorisation of these contacts as additional, we identified 2.1 million mandated contacts attended by our study cohort between 2016/17 and 2019/20. Over the same 4-year period, there were 8.6 million mandated contacts reported in HVSDM indicating that using the *activity type* field in CSDS identified only 24.0% of the mandated contacts that LAs reported delivering between 2016/17 and 2019/20.

After using our data supplementation approach to additionally identify age-derived mandated contacts, we identified a total of 3.9 million mandated contacts attended by our study cohort between 2016/17 and 2019/20 (Table 3). This was 44.7% of the 8.6 million mandated contacts reported in HVSDM for the same period.

Variation in CSDS completeness at local authority level

When comparing the pooled counts of CSDS-coded and age-derived mandated contacts to the aggregate mandated contacts reported in HVSDM, 9.9% of LA-quarters (237/2,384) had high agreement ($\pm 15\%$) to be categorised as 'complete' data. The proportion of LA-quarters that were complete increased over time from 4.0% in 2016/17 to 14.9% in 2019/20 (Supplementary Appendix 2). Overall, we identified a subset of 63 LAs (out of 149; 42%) that had at least one

Table 3: Number of mandated health visiting contacts identified for children aged <5 years in England in the Community Services Dataset from 2016/17 to 2019/20, by method of identification and coverage in relation to reference data

Mandated contact type	HVSDM reference data	Community services data set (CSDS)				
		Raw: based on <i>activity type</i> field	Cleaned: based on <i>activity type</i> field		Cleaned: including age-derived method	
			N	% of HVSDM	N	% of HVSDM
New birth visit	2,203,681	523,065	504,210	22.9	1,027,960	46.6
6-8-week review	2,173,045	467,930	448,530	20.6	885,590	40.8
1-year review	2,170,797	595,195	563,890	26.0	989,415	45.6
2-2½-year review	2,075,237	581,350	550,605	26.5	947,215	45.6
All	8,622,760	2,167,540	2,067,235	24.0	3,850,180	44.7

HVSDM = Health Visiting Service Delivery Metrics.

Table 4: Number and percentage of local authorities by total number of quarters of complete health visiting data recorded in the Community Services Dataset, 2016/17 to 2019/20

Number of quarters of data assessed as complete	Number of LAs	% of LAs with at least quarter of complete data (N = 63)	% of all LAs (N = 149)
1	13	20.6%	8.8%
2	16	25.4%	10.7%
3	10	15.9%	6.7%
4+	24	38.1%	16.1%

LA = local authority.

quarter of complete data (at least 85% of their delivered mandated contacts were recorded in the CSDS) between April 2016 and March 2020. The proportion of LAs that had complete data also increased over time from 7.9% in 2016/17 to 28.9% in 2019/20 (Supplementary Appendix 2).

The total number of quarters of complete data per LA for the 4-year study period ranged from 1 to 16 with a median of 3. Overall, 24 LAs had four or more quarters of data assessed as complete (Table 4; Figure 1). Of these, 13 had four or more continuous quarters of complete data, enabling longitudinal analysis of the health visiting contacts of a cohort of children from birth to 12 months. Two LAs had 10 or more continuous quarters, enabling tracking of health visiting experiences from birth to age 2.5 years.

The 63 LAs that had complete data for at least one quarter were broadly representative of England in terms of region, deprivation and other indicators of childhood adversity (Table 5). However, the population of LAs with complete data was less ethnically diverse than the English population overall (16.5% ethnic minority groups vs 20.9%; z-test, $p = 0.03$).

Among the subset of 63 LAs, the level of missing data for key variables related to health visiting activity ranged from 0% for provider organisation code to 97.5% for occupation type of the person providing the activity (Supplementary Appendix 3). The levels of missing data varied between LAs, but for most LAs variables related to staff characteristics had high levels of missingness; for example, just five LAs had <20% missing data for the 'staff type' variable.

Figure 1: Total number of quarters of complete health visiting data recorded in Community Services Dataset between 2016/17 and 2019/20, by local authority

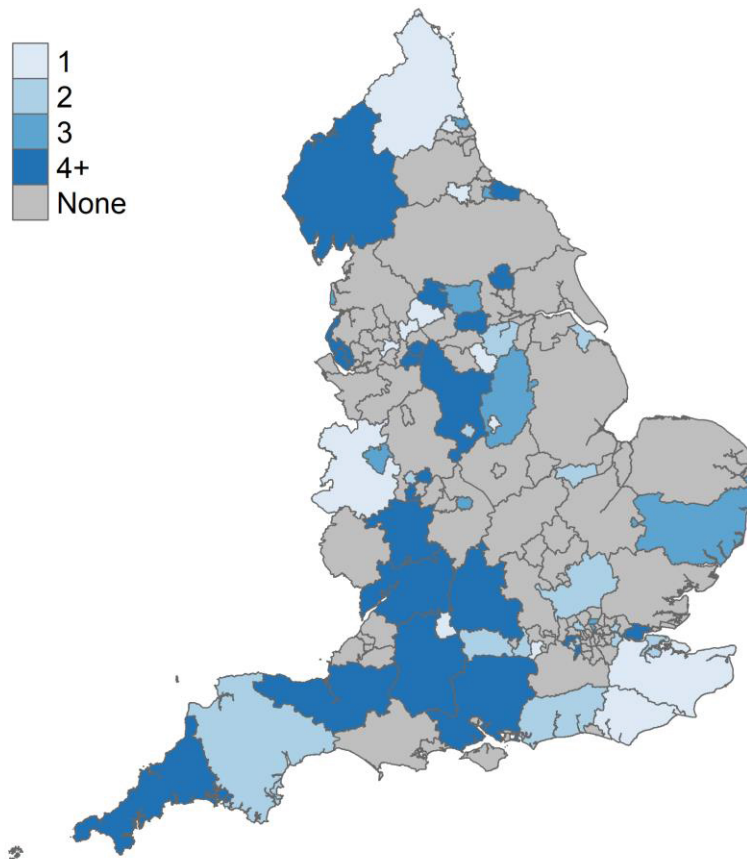


Table 5: Comparison of characteristics of local authorities with complete health visiting data in Community Services Dataset between 2016/17 and 2019/20 and all local authorities in England

	All local authorities in England (N = 149)		Local authorities with complete data (N = 63)		p-value
	n	%	n	%	
Region					
East Midlands	9	6.0%	4	6.3%	
East of England	12	8.1%	4	6.3%	
London	32	21.5%	8	12.7%	
North East	12	8.1%	6	9.5%	
North West	23	15.4%	9	14.3%	
South East	18	12.1%	9	14.3%	
South West	14	9.4%	8	12.7%	
West Midlands	14	9.4%	7	11.1%	
Yorkshire and The Humber	15	10.1%	8	12.7%	0.48
Geographical area classification^a					
Predominantly Rural	20	13.4%	9	14.3%	
Predominantly Urban	108	72.5%	45	71.4%	
Urban with significant rural	21	14.1%	9	14.3%	0.96
Index of multiple deprivation (IMD)^b					
Lowest quintile	30	20.1%	12	19.0%	
2 nd quintile	30	20.1%	13	20.6%	
3 rd quintile	30	20.1%	10	15.9%	
4 th quintile	30	20.1%	13	20.6%	
Highest quintile	29	19.5%	15	23.8%	0.69
Income deprivation affecting children index (IDACI)^b					
Lowest quintile	30	20.1%	12	19.0%	
2 nd quintile	30	20.1%	11	17.5%	
3 rd quintile	30	20.1%	12	19.0%	
4 th quintile	30	20.1%	14	22.2%	
Highest quintile	29	19.5%	14	22.2%	0.88
Prevalence of childhood adversities^c					
Lowest quintile	30	20.1%	13	20.6%	
2 nd quintile	30	20.1%	15	23.8%	
3 rd quintile	29	19.5%	13	20.6%	
4 th quintile	30	20.1%	9	14.3%	
Highest quintile	29	19.5%	13	20.6%	0.60
Ethnic diversity (mean %)					
White		79.1%		83.5%	
Ethnic minority groups		20.9%		16.5%	
Asian		10.6%		8.3%	
Black		4.8%		3.6%	
Mixed		3.1%		2.7%	
Other		2.5%		2.0%	0.03

Bold highlighting indicates p-value <0.05 for z-test for difference in proportions.

^aGeographical area type was categorised using the Office for National Statistics (ONS) Rural Urban Classification lookup table for local authority areas.

^bIMD and IDACI quintiles were based on national ranks from 2015, as published by the Ministry of Housing, Communities and Local Government.

^cQuintiles of the prevalence of childhood adversities was based on the estimated prevalence of living in a household where parent is suffering domestic abuse, alcohol/drug dependence or suffering a severe mental health problem per 1,000 children aged <18 years as published in the Children's Commissioner for England local authority vulnerability profiles 2019. Ethnic diversity was based on pooled ONS Census 2021 ethnic group classifications.

Discussion

In this study, we found that mandated contact codes in CSDS only identified 24.0% of all mandated contacts reported in OHID metrics from 2016 to 2020. Through a novel data supplementation approach, we were able to identify 44.7% of mandated contacts reported by OHID. This means that the 10.2 million health visiting contacts that we found in CSDS for all children in England during a 4-year period is likely to be a significant underestimate of the true level of activity provided by health visiting services.

We also created a study sample, through comparison with the national, aggregate reference data, which included a subset of LAs and time periods with complete health visiting data in the CSDS. The subset of data we derived for analysis includes only LAs and quarters for which the mandated contacts recorded in the CSDS were within $\pm 15\%$ of those reported by OHID. This enables use of a sample of data for research from which findings can be considered reliable. Between November 2023 and March 2024, we presented findings from nine LAs in our data to clinical service leads and/or commissioners in these same LAs. All these stakeholders agreed that the findings we had generated for their local area looked plausible and as they would broadly expect. We also spoke to stakeholders in three further LAs who did not have complete data in CSDS during our time period, presenting the overall results. These were also considered plausible. We caution that, if the full data source is used for analyses, findings will differ [14]. This will continue to be a limitation until completeness of CSDS improves.

Our subnational subset of data was broadly representative of all English LAs in terms of region, deprivation and other indicators of childhood adversity and can be reliably used for research related to health visiting practice and policy. We also show that it is possible to use CSDS data for longitudinal analyses of health visiting contacts in England with 13 LAs having continuously complete data for a year or longer. However, findings from longitudinal analyses on this small subset may be limited in their generalisability to the whole of England. There may also be challenges with using this subset for subgroup analyses or to explore rare outcomes given the reduced sample size.

Our analyses focused on the four postnatal mandated health visiting contacts delivered after birth. We did not include the antenatal mandated contact as this is recorded as a service provided to the mother and is not available in children's CSDS records. Future work could link CSDS to other datasets that connect children to mothers such as hospital admissions for births. This would allow examination of the full range of ante- and postnatal contacts and exploration of variation in patterns of health visiting according to maternal characteristics and background.

Limitations

During data cleaning and management, we encountered conflicting information between variables (e.g., mandated contacts at implausible ages). In this study, we aimed to provide transparency in our decision-making. However more information from data providers about how data are collected and submitted would help to guide this process.

Submission of incomplete data is a key issue for CSDS and restricts its utility as a research resource. Until overall data quality and completeness improves, we recommend that analyses using CSDS should be conducted only on a subset of complete data. This helps to ensure that inferences are reliable and form a robust evidence base for policymaking. The HVSDM reference data are essential in order to identify complete data, and should not be replaced by CSDS estimates until data are more reliably complete [15].

Among the subset of LAs with complete data are high levels of missing data for variables related to staff characteristics. This limits investigation of issues relating to the health visiting workforce and models of staff mix. Research with local data managers is ongoing to explore the barriers to submitting complete health visiting information (and Ages and Stages Questionnaire data collected at the 2-2 $\frac{1}{2}$ -year review) to CSDS and to identify best practice by providers and how to drive improvements [16]. Previous research has identified time restraints, data entry systems, and lack of workforce engagement as challenges in collection of data [17–19]. Addressing these barriers, from data entry through to submission and access for research, should be a key priority to enable the best use of the CSDS for health visiting research and community services research more broadly.

Conclusions

The CSDS is a powerful administrative data source that can be used for research when managed carefully. This management requires pre-analysis quality assessment, including careful identification of mandated contacts. While restricting to a subnational sample reduces the overall sample size and may limit the generalisability of findings, it also minimises the risk of underestimating activity and drawing spurious conclusions about health visiting and other community health services.

Acknowledgements

We would like to thank our team members, Eirini-Christina Saloniki and Jayu Jung, for helping with decision making throughout data preparation.

Ethics statement

The studies underpinning this work have been approved by University College London Research Ethics Committee (20561/002), University College London Institute of Education Research Ethics Committee (1531), and the National Research Ethics Service (21/SW/0159).

Conflicts of interest

None

Data availability

Access to the CSDS was approved and provided by NHS England (NIC-393510 and NIC-381972). The data do not

belong to the authors and may not be shared by the authors, except in aggregate form for publication. Data can be obtained by submitting a data request through the NHS England Data Access Request Service.

Funding statement

This work is supported by the National Institute for Health and Care Research (NIHR) Public Health Research Programme (NIHR129901) and Policy Research Programme (NIHR203450). This research benefits from and contributes to the NIHR Children and Families Policy Research Unit, but was not commissioned through this Policy Research Unit (PR-PRU-1217-21301). This research was supported in part by the NIHR Great Ormond Street Hospital Biomedical Research Centre.

The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

References

- Public Health England. Health visiting and school nursing service delivery model. 2021. Available from: <https://www.gov.uk/government/publications/commissioning-of-public-health-services-for-children/health-visiting-and-school-nursing-service-delivery-model>.
- Office for Health Improvement and Disparities. Healthy child programme. 2023. Available from: <https://www.gov.uk/government/collections/healthy-child-programme>.
- Department of Health and Social Care. The best start for life: a vision for the 1,001 critical days. 2021. Available from: <https://www.gov.uk/government/publications/the-best-start-for-life-a-vision-for-the-1001-critical-days>.
- Harron K, Kendall S, Bunting C, Cassidy R, Atkins J, Clery A, et al. How should policymakers, funders and research teams mobilise to build the evidence base on universal early years services? Primary Health Care Research & Development. In press.
- Woodman J, Mc Grath-Lone L, Clery A, Weatherly H, Jankovic D, Appleton JV, et al. Study protocol: a mixed-methods study to evaluate which health visiting models in England are most promising for mitigating the harms of adverse childhood experiences. *BMJ Open*. 2022;12(9):e066880. <https://doi.org/10.1136/bmjopen-2022-066880>.
- Harron K, Cavallaro FL, Bunting C, Clery A, Kendall S, Cassidy R, et al. Study protocol: evaluation of the 0-5 public health investment in England - a mixed-methods study integrating analyses of national linked administrative data with in-depth case studies. *BMJ Open*. 2023;13(4):e073313. <https://doi.org/10.1136/bmjopen-2023-073313>
- NHS Digital. Community Services Data Set. 2023. Available from: <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/community-services-data-set>.
- NHS Digital. Community Services Data Set Submission Tracker. 2023. Available from: <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/community-services-data-set/submission-update>.
- UK Statistics Authority. Code of Practice for Statistics. 2022. Available from: <https://code.statisticsauthority.gov.uk/>.
- Fraser C, Harron K, Barlow J, Bennett S, Woods G, Shand J, Woodman J. How can we use the community services dataset (CSDS) for research into health visiting? London: NIHR Policy Research Unit Children and Families; 2020.
- Fraser C, Harron K, Barlow J, Bennett S, Woods G, Shand J, et al. Variation in health visiting contacts for children in England: cross-sectional analysis of the 2-2(1/2) year review using administrative data (Community Services Dataset, CSDS). *BMJ Open*. 2022;12(2):e053884. <https://doi.org/10.1136/bmjopen-2021-053884>
- Office for Health Improvement and Disparities. Health visitor service delivery metrics experimental statistics: quarterly data for 2022 to 2023. 2023. Available from: <https://www.gov.uk/government/statistics/health-visitor-service-delivery-metrics-experimental-statistics-quarterly-data-for-2022-to-2023>.
- Office for Health Improvement and Disparities. Child and maternal health statistics: health visitor service delivery. 2017. Available from: <https://www.gov.uk/government/collections/child-and-maternal-health-statistics#health-visitor-service-delivery>.
- Public Health England. Contacts other than those mandated and referrals by health visitors. 2021. Available from: <https://www.gov.uk/government/publications/contacts-other-than-those-mandated-and-referrals-by-health-visitors>.
- Office for Health Improvement and Disparities. Interim data submission for the universal health visiting service: 2023 to 2024. 2023. Available from: <https://www.gov.uk/government/publications/childrens-public-health-0-to-5-years-national-reporting/interim-data-submission-for-the-universal-health-visiting-service-2023-to-2024>.
- UCL Children and Families Policy Research Unit. At a glance - Improving data completeness in the Community Services Dataset (CSDS). 2024. Available from: <https://www.ucl.ac.uk/children-policy-research/projects/start-life-tackling-inequalities-child-development-age-2>.

17. Lezard R, Deave T. The factors influencing community nurses' usage of electronic health records: findings from focus groups. *Br J Community Nurs.* 2021;26(12):604–10. <https://doi.org/10.12968/bjcn.2021.26.12.604>.
18. Keith J, Grimm F, Steventon A. How better use of data can help address key challenges facing the NHS2022. Available from: <https://www.health.org.uk/publications/long-reads/how-better-use-of-data-can-help-address-key-challenges-facing-the-nhs>.
19. Hargreaves S. "Moving Beyond Bean Counting": Improving Analytical Capability in Health Visiting. Institute of Health Visiting; 2021. Available

from: <https://ihv.org.uk/our-work/ihv-projects-and-evaluation/completed-projects/2021-2/analytics-in-health-visiting-moving-beyond-bean-counting-project/>.

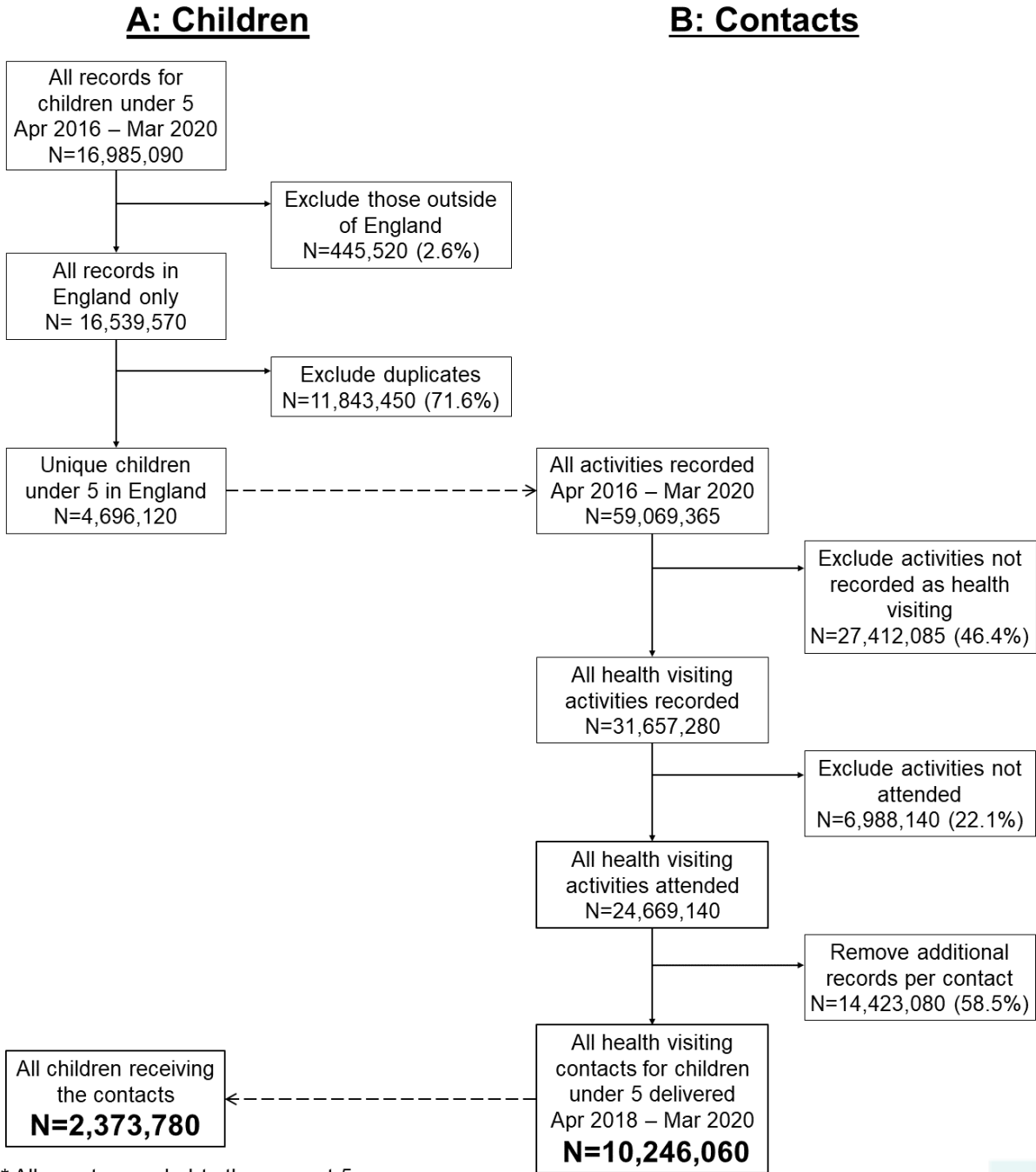
Abbreviations

COVID-19:	Coronavirus disease
CSDS:	Community Services Data Set
HVSDM:	Office for Health Improvement and Disparities Health Visitor Service Delivery Metrics
LA:	Local authority
NHS:	National Health Service
OHID:	Office for Health Improvement and Disparities
UK:	United Kingdom



Supplementary Appendices

Supplementary Appendix 1: Identifying the eligible cohort of children (A) and their health visiting contacts (B) in Community Services Dataset from April 2016 to March 2020



* All counts rounded to the nearest 5



Supplementary Appendix 2: Proportion of LA-quarters and local authorities that were assessed as having complete health visiting data in Community Services Dataset between 2016/17 and 2019/20

Year	LA-quarters (N = 596 per year)		Local authorities (N = 149)	
	n	%	n	%
2016/17	24	4.0%	11	7.4%
2017/18	49	8.2%	23	15.4%
2018/19	75	12.6%	39	26.2%
2019/20	89	14.9%	43	28.9%
Overall	237	9.9%	63	42.3%

LA = local authority. Complete data was defined as being within +/-15% of the reference Health Visitor Service Delivery Metrics for the specified local authority and quarter.

Supplementary Appendix 3: Missingness for key variables related to health visiting service delivery in Community Service Dataset for local authorities with complete data, 2016/17 to 2019/20

	Contacts with missing data		Range of missing data by LA		LAs with <20% missing data	
	n	%	Min	Max	n	%
Contact characteristics						
Duration (minutes)	247,110	9.3%	1.3%	100.0%	56	88.9%
Medium (e.g. face-to-face, phone)	177,420	6.7%	2.5%	99.9%	50	79.4%
Group therapy contact (yes/no)	416,420	15.6%	0.7%	99.9%	49	77.8%
Person who was the subject of the contact (i.e. patient or proxy)	86,905	3.3%	1.0%	100.0%	55	87.3%
Location of face-to-face contacts	341,340	16.3%	0.5%	99.9%	51	81.0%
Distance from home to contact location	2,286,620	85.7%	0.0%	99.1%	8	12.7%
Staff and service characteristics						
Staff type	2,155,715	80.8%	0.0%	97.8%	5	7.9%
Registration body	2,151,355	80.7%	0.0%	90.4%	4	6.3%
Occupation type	2,599,445	97.5%	0.0%	89.2%	1	1.6%
Job role	2,115,880	79.3%	0.1%	89.4%	5	7.9%
Provider organisation code	0	0.0%	n/a	n/a	63	100.0%
Commissioner organisation code	25,700	1.0%	6.0%	100.0%	61	96.8%

LA = local authority. N = 2,666,845 for contacts; 2,096,650 for face-to-face-contacts; 63 for local authorities.

