



Data Explained

Who are the children and how are their needs recorded? Methodological lessons from using the Mental Health Services Dataset in ECHILD

Author: Dr Simona Skripkauskaite

Date: August 2026

DOI: <https://doi.org/10.5281/zenodo.23535891>

Summary

This Data Explained summarises experiences and learning from working with the Education and Child Health Insights from Linked Data (ECHILD) linked dataset while researching pathways through education, health and social care services for neurodivergent children with mental health needs. It focuses particularly on challenges encountered when using the Mental Health Services Dataset (MHSDS), including inconsistencies in demographic information across MHSDS tables and limitations in diagnostic recording. This publication is intended to help guide future researchers using these data and to provide feedback to inform future dataset development and documentation.

Project details

This research used linked administrative data from the ECHILD project, which brings together health, education and social care records for children and young people in England. ECHILD has near-population coverage, containing linked data for an estimated 21 million individuals overall, with very high linkage rates (approximately 99% of pupil records are linked to a health record). National Data Opt-Outs are applied by NHS England to health data before linkage within ECHILD, meaning that health records are not available for linkage for individuals who have opted out of their confidential patient information being used for research and planning.

This ADR UK fellowship project aimed to build a clearer picture of children's pathways through educational support, social care, and health services, and to examine when different points of contact occur across childhood (Skripkauskaitė, 2025). The goal was not only to describe service use, but to identify where there may be opportunities for earlier recognition and support before children's needs escalate or reach crisis. Specifically, it examines timing and likelihood of first contact with services including early years provision, special educational needs support, Education, Health and Care Plans, school exclusion, social care, emergency care for self-harm, outpatient, and unplanned admitted patient care (Skripkauskaitė & Ali, under review). The project uses a quasi-experimental approach with three groups:

- neurodivergent children with recorded mental health needs (ND+MH);
- neurodivergent children without recorded mental health needs (ND-only);
- non-neurodivergent children with recorded mental health needs (MH-only).

The original plan

The fellowship was originally designed around clinical diagnoses recorded in MHSDS. The intention was to use demographic information to determine which records represented eligible children. I then planned to use clinical diagnoses recorded using International Classification of Diseases (ICD) or SNOMED CT coding systems to classify those children according to neurodivergence- and mental health-related phenotype codelists and to compare how different groups moved through support systems over time. Specifically, the original plan was to define the cohort as all children born in or after 2010 who were recorded in MHSDS between 2016 and 2022.

This approach offered several advantages.

First, defining the cohort of children directly from MHSDS would have allowed the study to include all children who met the inclusion criteria, including those who may not be captured in education records due to homeschooling, private education, or otherwise being absent from the National Pupil Database (NPD). Second, diagnoses aligned with existing clinical frameworks and reflected established categories and phenotypes used in research and practice. Finally, diagnoses appeared to offer a relatively objective way of distinguishing neurodivergence and mental health presentations.

The initial plan, therefore, assumed that MHSDS could answer two fundamental questions:

1. Which records represented children meeting the study eligibility criteria?
2. How could these children be classified according to neurodivergence- and mental health-related needs?

Early exploration of the data revealed that answering both questions was more complicated than anticipated.

Although these questions are considered separately below for clarity, the problems emerged together. Exploring referral and diagnostic records also raised concerns about whether records of eligible children could be reliably distinguished from MHSDS demographic information, while the eventual decision to establish the cohort using census records determined the population in which referral and diagnostic information could be evaluated.

Challenge one: determining which records represent children

The first major challenge concerned demographic information.

Information contained within the Master Patient Index (MPI) did not always align with demographic information recorded elsewhere in MHSDS, including referral records. Initially, using demographic information from the MPI, I selected a sample of around 600,000 who I thought represented children under 12 years old. Concerns about these inconsistencies arose partly through exploration of the referral data. However, when I started exploring the referral tables, some of these same records appeared to relate to adults, including people in their 30s. In other cases, the age recorded in the referral data produced combinations of children's demographic information and reasons for referral, including toddlers being referred for 'Drug and alcohol issues' or 'Preconception perinatal mental health concern'. These records were not necessarily incorrect. For example, some referral reasons could potentially reflect concerns relating to a parent rather than the child themselves, errors in the demographic information itself, or in how records were attributed to the same individual within MHSDS. However, these discrepancies across MHSDS tables raised uncertainty about whether demographic information within MHSDS could be used reliably to determine whether records represented an eligible child, and, if so, which source should be prioritised.

The compromise: constructing the cohort through linkage

Rather than distinguishing records relating to children directly from MHSDS, the cohort was constructed through linkage with census records in the NPD. The cohort spine had to be defined by first considering anyone who had been recorded in the MHSDS between 2016 and 2022, which resulted in over 3 million individual cases. Then, I had to assess whether any of these cases could be linked with pupils recorded as born after 2010 in Spring, Summer, Autumn Census between 2012-2022 or Alternative Provision Census, which narrowed it down to 601,361 children born after 2010 who had been recorded in MHSDS. These were not the same 600,000 'children' as selected earlier directly in MHSDS MPI, as only about a quarter (23.4%) of the MPI sample had been recorded as pupils in the NPD spring census for 2022. It was beyond the scope of the fellowship to systematically assess which source of demographic information within MHSDS most closely aligned with demographic information recorded in NPD.

This decision improved confidence in cohort construction and ensured that analyses were based on a consistent demographic framework. However, it also introduced important trade-offs:

- **Computational costs** - Constructing the cohort spine through such linkage required extensive processing and quality checks across multiple datasets with server capacity being constantly compromised (*my apologies to any other users who were trying to complete their data analysis at the same time*). This was substantially more computationally intensive than the original approach and required considerable server time with scripts often falling apart without any data saved unless split into smaller and more manageable chunks.
- **Coverage costs** - More importantly, this approach restricted the study population to children who could be linked to education records. Children who were not captured in the NPD could no longer be included.

As a result, some groups of children in contact with mental health services may be underrepresented in the final study population.

Challenge two: missing diagnostic information

The second major challenge concerned diagnostic recording.

Initial analyses demonstrated that diagnostic information was available for only a minority of children.

Among the final census-linked cohort of 601,361 children:

- only 60,047 (10.0%) had any recorded primary, secondary, provisional, or previous diagnoses;
- only very few (~1-2%) of children then had a recorded diagnoses that aligned with conditions included in 'neurodevelopmental disorders' section of ICD-11.

These findings raised an important methodological question: if diagnoses were available for fewer than one in ten children, could they realistically be used to classify children into the study groups?

Upon consulting colleagues and patient and public involvement (PPI) advisors, I decided the answer was ultimately no.

Restricting analyses to children with recorded diagnoses would have excluded the vast majority of children in contact with services and would likely have produced a highly selected sample. Recent research confirms that only a small proportion of children and young people referred to CAMHS have their referrals accepted and receive a diagnosis within 18 months and that whether they do varies based on their age, previous referrals, and neighbourhood deprivation (Sayal et al., 2026).

The compromise: moving from diagnoses to referrals

To address this limitation, the study shifted from diagnoses to referral information.

Within this census-linked cohort, referral information was available for substantially more children and offered broader coverage across the dataset. After data cleaning (removing duplicated, unknown, undisclosed, or missing referral records), a primary or other reason for referral was available for 407,043 (67.7%) children. For the purposes of my fellowship study, the reason for referrals were then reviewed in consultation with clinicians and PPI contributors. Those considered unlikely, irrelevant, or too ambiguous to classify were excluded (see Skripkauskaitė & Ali, under review). This resulted in 396,762 (66.0%) children with recorded neurodivergence- or mental health-related referrals.

Among these children, 41,368 had both referral and diagnostic data. Of these, 67.9% were classified into the same group using both data sources, indicating moderate concordance (see the table below).

Referrals	Diagnoses					
	MH-only		ND-only		ND+MH	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
MH-only	12065	29.2%	3064	7.4%	1068	2.6%
ND-only	455	1.1%	10824	26.2%	284	0.7%
ND+MH	1708	4.1%	6710	16.2%	5190	12.5%

However, this change represented more than a technical adjustment. It also changed what the resulting groups represented. Diagnoses and referral reasons are not competing measures of the same underlying construct: they capture different stages and aspects of children's interactions with services. Diagnoses reflect clinical classification following assessment. Their recording may depend on factors including access to diagnostic pathways, diagnostic thresholds, clinician judgement and local recording practices. In contrast, reasons for referral describe the needs or concerns through which a child comes to the attention of mental health services. They may reflect concerns identified by children and families, schools or other professionals, as well as decisions about referral and service access.

This required a subtle but important shift in how the research question was operationalised and interpreted. Rather than asking: *“How do children with particular diagnoses move through services?”* the project became more about *“How do children whose needs are recognised in different ways move through services?”*.

Summary of comments on specific variables

Variable name(s)	Comments
MHSDS Primary and secondary diagnoses	Very limited coverage. Using these alone would have excluded the large majority of children in contact with MHSDS.
MHSDS Previous and provisional diagnoses	Improved diagnostic coverage, but even after including these fields only 60,047 of 601,361 children (10.0%) had any recorded diagnosis.
MHSDS Primary and other reason for referral	Much better coverage than diagnoses and thus ultimately used to define the comparison groups. However, referral reasons represent the reason for service referral rather than a confirmed clinical diagnosis and required substantial PPI work for classification.
MHSDS Age at reporting period or referral	Demographic information was inconsistent across MHSDS files, particularly between the MPI and referral tables. This prevented reliably determining the cohort of eligible children using MHSDS alone.
NPD Census-derived demographic information	Used to determine which records represented children and select those born from 2010 onwards. This improved confidence in cohort definition but restricted the analysis to children recorded in NPD census data.

Suggested improvements recommended to data owners

Two improvements arise directly from these learnings that would substantially increase the value of MHSDS for research using linked administrative data. As these recommendations concern the recording and consistency of information within MHSDS, they are primarily directed to NHS England as the data owner, rather than to the ECHILD team responsible for linking MHSDS with other datasets.

Establish the reliability of demographic information across MHSDS tables

Further work by NHS England is needed to assess the consistency and reliability of core demographic information across MHSDS tables, including the MPI and referral tables, and establish the reasons for discrepancies between them. This could include examining repeated demographic information over time to determine whether some sources or combinations of records provide more reliable information than others and whether discrepancies arise through recording practices or the way records are attributed or linked within MHSDS. Clear guidance on how the researchers should interpret and resolve conflicting demographic information would substantially improve the usability of MHSDS for secondary research.

Improve the recording of diagnostic information

More complete and consistent recording of diagnostic information would substantially increase the research value of MHSDS. In this study, fewer than 10% of children in the final census-linked cohort had any diagnosis recorded, even after primary, secondary, provisional, and previous diagnosis fields were considered. This made it impossible to implement the original diagnosis-based study design and required referral reasons to be used instead.

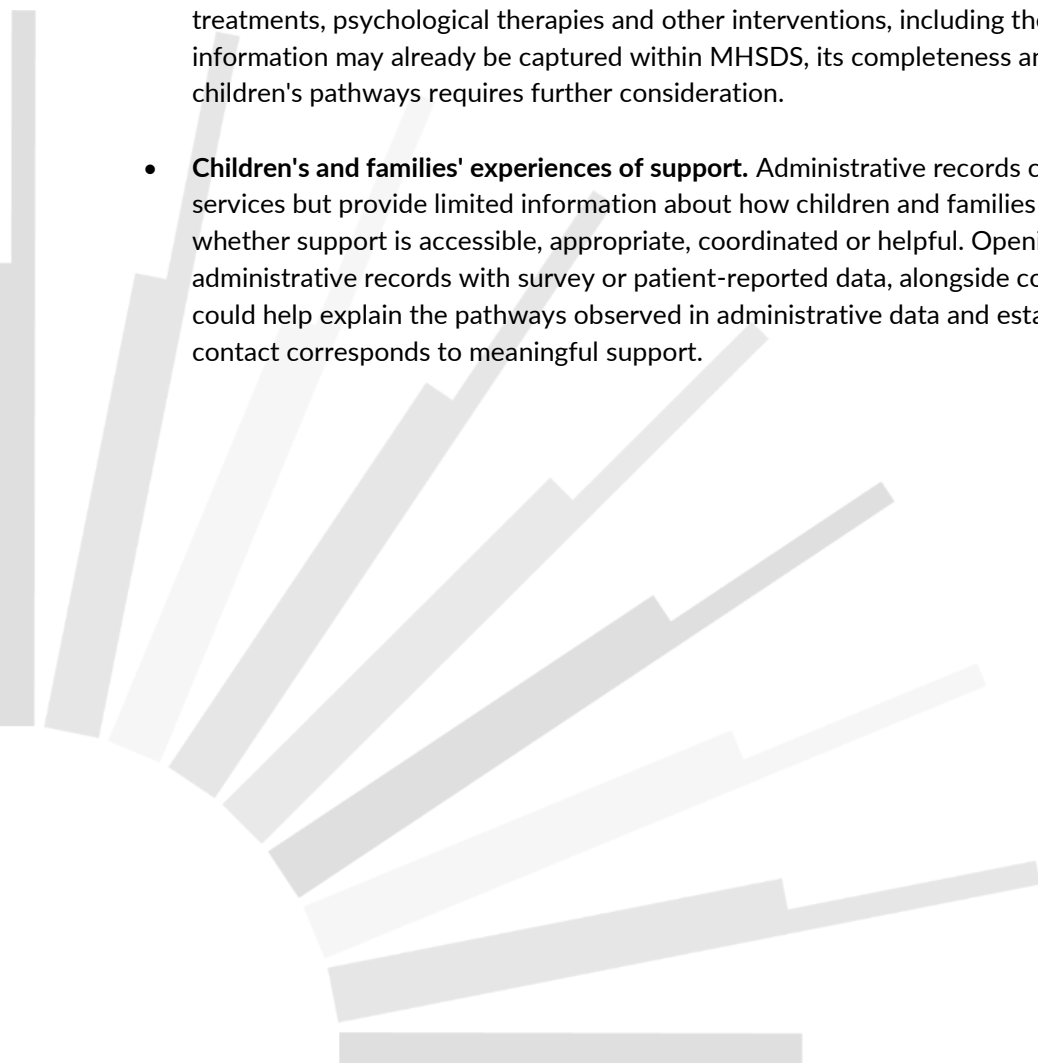
This does not imply that every child accessing mental health services should necessarily receive a formal diagnosis. Diagnostic practices in CAMHS vary, and some services may appropriately work within needs-based approaches without a formal diagnosis. However, the limited availability of recorded diagnostic information creates an important gap in understanding who is accessing services and what support they receive. Improving diagnostic data capture in MHSDS, including clarity about whether a diagnosis has been assessed, established, considered but not assigned, or simply not recorded, would help determine more accurate differences in access, treatment and outcomes without assuming that absence of a recorded diagnosis means absence of a diagnosable condition.

There is also a wider debate about the relative roles of diagnosis and needs- or functioning-based approaches to understanding and supporting neurodivergent children. Diagnostic information provides important information about clinically recognised conditions, but does not in itself capture functional impact or day-to-day support needs. The recent independent review into mental health conditions, ADHD, and autism highlights the lack of reliable information on severity and functional impact within routine data and argues for support to be better aligned with need, functional impact and severity, while retaining access to specialist diagnostic assessment and treatment where required (Department of Health and Social Care, 2026). This distinction was also reflected in discussions with lived-experience contributors and other stakeholders during this project. While earlier access to support based on need was generally supported, contributors also emphasised that diagnosis can have important functions for children and families, including providing understanding and validation, supporting identity development, and facilitating access to services.

Additional data which would help to further develop the research

Discussions with neurodivergent young people, their parents and carers, practitioners, commissioners and researchers during the project and roundtable event highlighted several areas where more complete information, either within routinely collected service data or through linkage to additional data sources, could extend the research.

- **Recognition, referral and diagnosis occurring outside the services captured by MHSDS.** Neurodivergence and mental health needs may be identified, referred, or diagnosed through primary care, community services, or private providers. These pathways are not necessarily captured within MHSDS, particularly where care is not NHS-funded. Linkage to data from these settings could provide a more complete picture of when children's needs are first recognised and assessed, when diagnoses are made, and the different routes through which children access support.
- **Support and interventions received.** More complete and usable information within MHSDS on what happens following referral and the interventions children receive would also strengthen research such as this. This includes whether referrals lead to assessment or support and more detailed information on pharmacological treatments, psychological therapies and other interventions, including their duration and intensity. While such information may already be captured within MHSDS, its completeness and usability for reconstructing children's pathways requires further consideration.
- **Children's and families' experiences of support.** Administrative records can help describe contacts with services but provide limited information about how children and families experience those contacts, including whether support is accessible, appropriate, coordinated or helpful. Opening the possibility of linking administrative records with survey or patient-reported data, alongside complementary qualitative research, could help explain the pathways observed in administrative data and establish whether recorded service contact corresponds to meaningful support.



Works cited

- Department of Health and Social Care. (2026, June 10). *Independent review into mental health conditions, ADHD and autism: Interim report*. GOV.UK. <https://www.gov.uk/government/publications/independent-review-into-mental-health-conditions-adhd-and-autism-interim-report/independent-review-into-mental-health-conditions-adhd-and-autism-interim-report>
- Sayal, K., Holt, G., Partlett, C., Wyatt, L., Bhardwaj, A., & Gledhill, J. (2026). Investigating inequalities in children and young people's mental healthcare and outcomes: prospective longitudinal analysis from the STADIA trial. *The British Journal of Psychiatry*, 1–10. <https://doi.org/10.1192/BJP.2026.10617>
- Skripkauskaite, S. (2025, March 27). *A roadmap to better support for neurodivergent children* [Blog]. ADR UK. <https://www.adruk.org/news-publications/news-blogs/a-roadmap-to-better-support-for-neurodivergent-children/>
- Skripkauskaite, S., & Ali, Z. (under review). Becoming visible to services: First contact with education, health, and social care among children with histories of neurodivergence- and mental health-related referrals in England.

Cite this publication: Skripkauskaite, S., & Ali, Z. et al. (2026). Data Explained: Who are the children and how are their needs recorded? Methodological lessons from using the Mental Health Services Dataset in ECHILD. Zenodo. <https://doi.org/10.5281/zenodo.23535891>

Acknowledgements

This work is funded by ADR UK (Administrative Data Research UK). ADR UK is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives. ADR UK is an Economic and Social Research Council (ESRC) investment (part of UK Research and Innovation). [Grant number: ES/Z503198/1].

We would like to acknowledge the contribution of the wider ECHILD Database support and programme management. This work uses data provided by children and young people collected by the NHS or the Department for Education as part of their services. We gratefully acknowledge all children and families whose de-identified data are used in this analysis. We would also like to thank all the advisors, including the neurodivergent young people, parents and carers, practitioners, voluntary sector representatives, and service commissioners, who have supported this project throughout with their wisdom and expertise.

About ADR England and ADR UK

ADR England is a portfolio of data linking and research projects, delivered by a range of academic and government partners to provide policy-relevant insights using data held by UK Government departments and public bodies. The geographic scope of ADR England projects varies depending on the data holder, but is typically England-only, England and Wales, Great Britain or UK-wide. ADR England is part of ADR UK, an ESRC investment and part of UKRI.

ADR UK (Administrative Data Research UK), ESRC's flagship data-sharing partnership, enables secure access to linked UK public-sector data for researchers. From crime and justice to education, the economy, and health, this UK-wide partnership is transforming public sector data into research insights and policy evidence to improve lives.

Disclaimer

The ECHILD Database uses data from the Department for Education (DfE). The DfE does not accept responsibility for any inferences or conclusions derived by the authors. This work uses data provided by patients and collected by the National Health Service as part of their care and support. Source data can also be accessed by researchers by applying to NHS England. We are grateful to the Office for National Statistics (ONS) for providing the trusted research environment for the ECHILD Database. ONS agrees that the figures and descriptions of results in the attached document may be published. This does not imply ONS' acceptance of the validity of the methods used to obtain these figures, or of any analysis of the results. The views in this publication do not necessarily reflect the views of UCL.

This work uses research datasets which may not exactly reproduce National Statistics aggregates. National Statistics follow consistent statistical conventions over time.

Publication contact

Name: Dr Simona Skripkauskaite
Email: simona.skripkauskaite@psych.ox.ac.uk or
s.skripkauskaite@surrey.ac.uk

