Martin-Denham, Sarah, Scott, Nathan, Pinney, Anne and Horridge, Karen (2026) Have children with learning disabilities, as identified in school census Children in Need administrative data, had appropriate paediatric clinical assessments completed and reported across agencies? Project Report. University of Sunderland.

Version: UNSPECIFIED

Available under license: UNSPECIFIED

Downloaded from: https://sure.sunderland.ac.uk/id/eprint/20496/

Which version am I reading (articles, book chapters and books)

Pre-print (or submitted version): the version of an article or chapter before any peer-review. Can be submitted to a publisher or added to a pre-print server.

AAM: Author Accepted Manuscript - The version of the article formally accepted for publication after the peer-review process. It includes the changes made by the author(s) in line with peer-reviewers' comments. It has not undergone the publisher's formatting, type-setting and copyediting.

VOR: Version of Record - The version published on the publisher's website which includes formatting, typesetting and copyediting.

Usage guidelines

Please refer to the usage guidelines at http://sure.sunderland.ac.uk/policies.html or alternatively

contact sure@sunderland.ac.uk.

University of Sunderland

Have children with learning disabilities, as identified in school census and Children in Need administrative data, had appropriate paediatric clinical assessments completed and reported across agencies?

Dr. Sarah Martin-Denham , Nathan Scott,
Anne Pinney, Dr. Karen Horridge

Funded by the True Colours Trust and the National Institute of Health Research, School of Public Health and the University of Sunderland

November 2025

fuse

The True Colours Trust

NIHR | School for Public Health Research

Table of Contents

Acronyms 3
Glossary 4
Executive Summary 5
1.0 Background 6
1.1 What is data linkage? 6
1.2 Why link local administrative data? 6
1.3 Policy context 8
1.4 Prevalence, terminology, and data gaps in learning disability documentation 8
2.0 Aims and objectives 12
3.0 Methodology and methods 13
4.0 Datasets and triangulating learning disability data 14
4.1 Intersection of cohorts throughout the data 16
4.2 Demographics for the whole linked dataset 17
4.3 Identifying learning disability terminology 18
5.0 Results 20
5.1 Matching terminology across datasets 20
5.2 Learning difficulty and disability by demographic characteristics 23
5.3 Distribution of learning disabilities throughout key datasets 26
5.4 Objective 1 findings 28
5.5 Objective 2 findings 29
5.6 Objective 3 findings 30
5.7 Objective 4 findings 30
5.8 Objective 5 findings 36
6.0 Concluding remarks 37
7.0 Limitations 39
8.0 Future research 40
Recommendations 41
References 42
Appendices 47

Acronyms

AP - Alternative Provision

CAMHS - Child and Adolescent Mental Health Services

CiN - Children in Need

CSDS - Community Services Data Set

DfE - Department for Education

EHCP - Education, Health and Care Plan

FSM - Free School Meals

ICB - Integrated Care Board

IQ - Intelligence Quotient

LA - Local Authority

LD - Learning Disability

MLD - Moderate Learning Difficulty

NHS - National Health Service

NECS - North of England Commissioning Support Unit

PMLD - Profound and Multiple Learning Difficulty

SEN - Special Educational Needs

SEND - Special Educational Needs and Disabilities

SEMH - Social, Emotional and Mental Health

SLCN - Speech, Language and Communication Needs

SLD - Severe Learning Difficulty

SpLD - Specific Learning Difficulty

STSFT - Sunderland and South Tyneside Foundation Trust

TfC - Together for Children

UoS - University of Sunderland

Glossary

Learning difficulty/disability terminology	Definition
Learning Difficulty	Term used in education datasets to describe challenges in learning; includes categories such as Moderate Learning Difficulty (MLD), Severe Learning Difficulty (SLD), and Profound & Multiple Learning Difficulty (PMLD).
Learning Disability (LD)	Term used in health and social care to describe significant intellectual impairment, often confirmed through clinical assessment.
Specific Learning Difficulty (SpLD)	Includes conditions such as dyslexia, dyscalculia, and dysgraphia; does not imply general intellectual disability.
EHCP (Education, Health and Care Plan)	A statutory plan for children with complex special educational needs (SEN) requiring coordinated support.
SEN (Special Educational Needs)	Additional support needs in education at 2 levels: SEN Support and EHCP.
Data and data linkage terminology	
AP (Alternative Provision)	Education outside mainstream schools, often for children excluded or unable to attend regular school settings.
Batch Tracing	Method of matching records across datasets using demographic details to retrieve identifiers (e.g., NHS numbers).
CiN (Children in Need)	Legal definition under the Children Act 1989; includes disabled children entitled to assessment for support.
CSDS (Community Services Data Set)	National dataset capturing information on individuals in contact with community health services.
Data Linkage	Process of combining data from multiple sources for the same individual to create a richer dataset.
Pseudonymisation	Technique that replaces identifiable information with irreversible pseudonyms to protect privacy.
SEN2	A statutory data collection conducted by the Department for Education (DfE) that gathers detailed information on children and young people with Education, Health and Care Plans (EHCPs).
Statistical terminology	
AIC (Akaike Information Criterion)	A statistical measure used to compare the goodness of fit of different models while penalizing for model complexity. Lower AIC values indicate a better balance between model fit and simplicity.
Confidence Interval (CI)	Range within which the true value of a parameter is expected to fall with a given level of confidence (usually 95%).
ICC (Intra-Class Correlation)	Proportion of variance explained by group-level differences (e.g., schools) in multi-level models.
Marginal R^2 / Conditional R^2	Measures of variance explained by fixed effects alone (marginal) and by both fixed and random effects (conditional).
Mixed-Effects Model	Statistical model that accounts for both fixed effects (predictors) and random effects (e.g., school-level variation).
Odds Ratio (OR)	A measure of association between a predictor and an outcome in logistic regression; OR > 1 indicates that a predictor is associated with increased odds of the outcome
p-value	Statistical measure indicating the probability that an observed effect occurred by chance; $p < 0.05$ typically considered significant.

Executive Summary

This study explored whether children identified in education and social care datasets as having learning difficulties or disabilities had corresponding paediatric clinical assessments documented in health records. Using a linked dataset combining school census, Children in Need (CiN) census, and paediatric health data for Sunderland, the research aimed to evaluate consistency in identification and reporting of learning disabilities across agencies. The data linkage was facilitated by a batch-tracing exercise which attached NHS numbers to local authority datasets, allowing these records to be linked to paediatric appointment records.

The study revealed significant discrepancies in how learning needs are described in education and health records. Education datasets use loosely defined terms such as Moderate Learning Difficulty (MLD), Severe Learning Difficulty (SLD), and Profound and Multiple Learning Difficulty (PMLD), while health records apply ICD-11 classifications such as Mild, Moderate, Severe, and Profound Intellectual Developmental Disability, based on clinical assessments. These differences mean that severity levels do not align between sectors, creating confusion and limiting the ability to build a coherent, cross-cutting picture of children's needs to inform service planning and commissioning or to understand trends in prevalence, at any level.

Furthermore, a substantial proportion of children with a learning disability identified in the paediatric data had no corresponding learning difficulty recorded in the school census. In 2023/24, only 36% of children with a paediatric record of a learning disability had any learning difficulty record in their education data, and often only autism or other SEN needs were reported. This suggests that learning disabilities may be obscured by other SEN categories in education records, potentially affecting access to tailored support and reasonable adjustments. While CiN records of learning disability were more likely to align with education and health data, overall consistency across datasets was low. Only 4–5% of those identified in paediatric data appeared in CiN records. These findings highlight systemic gaps in cross-agency visibility of childrens' learning disabilities. At a local area level, this also means that learning disabilities may be less visible in crucial local authority data used for planning and commissioning specialist provision.

Inferential statistics also confirmed well-documented patterns in learning difficulties in the school census, with boys and those eligible for free school meals significantly more likely to have a learning difficulty recorded in their school census data. These effects persisted across all three years of data, as was the finding that around 15% of variability in learning difficulties in the education data can be attributed to school-level differences.

These findings have implications for children and young people, their families, and services' planning and commissioning. Families may face inconsistent support because agencies operate with different definitions and thresholds, and fragmented recording and documentation limits the effectiveness of joint planning and commissioning. There is a need for clearer, cross-sector terminology and improved data-sharing practices to ensure children's needs are accurately identified and supported, and that local areas have robust prevalence data on learning disabilities among children to inform planning and commissioning decisions.

1.0 Background

In 2021, Dr Sarah Martin-Denham, together with local area partners, sought and secured funding to begin the process of obtaining the necessary information governance approvals to link individual-level health, education, and social care data for children in Sunderland, North East England.

Funding to date has been provided by the University of Sunderland, the Department for Levelling Up, Housing and Communities (through the Local Data Accelerator Fund) (department is now the Ministry of Housing, Communities and Local Government), True Colours Trust and the National Institute of Health and Care Research (NIHR) School of Public Health.

Health data were received in 2024, followed by Local Authority (LA) data in 2025. The pseudonymised data were transferred to the University of Sunderland (UoS) and linked by the research team. The processes involved in this data linkage, including information governance barriers and solutions, are documented in a briefing paper to support local areas and inform policy development (Martin-Denham et al., 2025). The information governance arrangements and processes required were long and complex to an extent that could not be predicted at the start of the project.

1.1 What is data linkage?

Data linkage is the process of combining data from various sources relating to the same individual to create a new, enhanced dataset. This linked data enables agencies to identify trends in service access and evaluate the effectiveness of these services (Downs et al., 2017). Linked data can be used to answer questions that require large sample sizes and is particularly effective for analysis about populations that are hard to reach (Chiu et al., 2016; Aldridge et al., 2019). Through linked data, local areas could explore relationships between education, social care, and healthcare outcomes for children. However, routinely collected data by education, health, and social care sectors are rarely linked by Local Authorities (LAs) (Martin-Denham et al., 2023) or Integrated Care Boards (ICBs) (Scott et al., 2024).

In 2023, Freedom of Information (FOI) requests revealed a considerable disparity in data linkage practices between LAs and ICBs (Martin-Denham et al., 2023). The findings highlighted that 73% of LAs (n=91) were linking their internal administrative datasets, but only 24% also linked their data with external healthcare data. Similarly, 58% of ICBs (n=31) were linking their internal datasets, but only 16% were linking their data with LA data. LAs and ICBs primarily used data linkage for strategic planning and the creation of data dashboards, with some utilising it for joint commissioning and service reviews (Martin-Denham et al., 2023; Scott et al., 2024).

1.2 Why link local administrative data?

The breadth and depth of local data sets provide opportunities to understand the impact of interventions on improving outcomes for children (ADR UK, 2023). Linking these datasets can support the identification of those at risk of adverse outcomes and support early intervention (Atherton et al., 2015). Local authorities and partner organisations can use linked data to improve children's quality of life in multiple ways, informing strategic planning, targeted interventions, and joint commissioning (Downs et al., 2017; Sohal et al., 2022; Martin-Denham et al., 2023; Scott et al., 2024; Pinney et al, 2024). By combining information from multiple sources, data linkage strengthens the evidence base of decision making. For example, during the COVID-19 pandemic, the lack of ethnicity information on death registrations was identified and rectified by linking this data with the 2011 census (Government Actuary's Department (GAD), 2021).

Research consistently shows that children with special educational needs and disabilities (SEND) are exposed to increased risk of poverty, unemployment, poor housing, and discrimination (Parsons and Platt, 2022; Krasniki et al., 2023). Yet, their multi-faceted needs are often poorly captured and described in administrative data (Horridge, 2018) and may therefore not be visible in key administrative data used for planning and commissioning services, which are often collected in silos (Wilcock, Elliott and Symons, 2022).

An important indirect benefit of data linkage is the improvement of data quality across services. Linkage projects rely on accurate, rich, and high-quality data. The process and analysis lead to the identification and resolution of inconsistencies, gaps, and data blockages (Holman et al., 2008). For example, this project identified missing diagnostic data that should have been flowing into the Community Services Data Set (CSDS), a blockage that has now been resolved.

National-level initiatives have also sought to link children's data. The most noteworthy is University College London's Education and Child Health Insights from Linked Data (ECHILD, 2025), in partnership with the Department for Education (DfE), the National Health Service (NHS) England and the Office for National Statistics (ONS). ECHILD was the first study to combine health, education, and social care data for all children in England. If ECHILD datasets become sufficiently comprehensive, timely, and accessible, local areas may not need to undertake data linkage themselves, as they could extract their own local area's statistics from the ECHILD datasets. However, ECHILD does not include the data required for this study – lacking the diagnostic breadth and depth of data about individual, multi-faceted needs. Effective reporting of these data relies on systems being available at all points of clinical care across all providers and settings to capture and record them easily. While these data should be collected and processed through the CSDS, it was missing from the dataset, instead flowing into the Acute Outpatient Data. This further highlights the barriers posed by inconsistencies in administrative data collection.

Another promising initiative is the Pan London Data Sharing Agreements Project, which was delivered by the London Office of Technology & Innovation in collaboration with the London Safeguarding Children Partnership (LOTI, 2025). All 32 London boroughs signed a 'London Multi-Agency Safeguarding Data Sharing Agreement for Safeguarding and Promoting the Welfare of Children'. Although the full benefits of this agreement are yet to be fully understood, feedback from the London borough responses indicate positive impacts, including:

- a consistent approach that reduces the need for multiple formal agreements
- stronger working relationships that build confidence
- a reliable framework for data sharing
- increased efficiency in processes

Case studies of LAs highlight the tangible benefits of linking data across services (Pinney et al., 2023):

Bedford Borough: By linking NHS data with Education, Health and Care Plans (EHCPs), Bedford identified delays in access to speech and language therapy, long wait times between appointments, and instances of some children being incorrectly discharged from services. The linked data enabled the identification of children with SEND within health data, revealing that children were more likely to access Child and Adolescent Mental Health Services (CAMHS), rather than the neurodevelopmental team within CAMHS. This led to changes within CAMHS, including more SEND staff training, and a reduction in paediatrician caseloads.

Hartlepool Borough Council: Hartlepool operates an integrated 0-19 service covering health visiting, school nursing, and Early Help. The service uses NHS numbers as consistent identifiers across its two information systems employed by the teams. A data sharing agreement enables the sharing of information about babies with disabilities, either antenatally or after birth. This pro-active approach ensures that specialist support can be arranged in advance, ready for when the child starts nursery or primary school.

Bradford Council: Bradford has developed a live data dashboard that refreshes daily and brings together multiple LA data sets. This dashboard facilitates the identification of vulnerable children based on their SEN/social care status, and their presence in youth offending data. This data dashboard has enabled the council to monitor changes within vulnerable cohorts and proved especially valuable during the COVID-19 pandemic for tracking and supporting at-risk children.

1.3 Policy context

The SEND review (Her Majesty's (HM) Government, 2022) proposed to make better use of data in the SEND system as current data capture is 'inconsistent' and does not 'enable local systems and leadership to respond to local needs before it is too late' (p. 69). The Independent Review of Social Care (MacAlister, 2022) recommended that the DfE develop a proactive strategy to better utilise data in children's social care, including a strategy for data linking with other data sources such as education, hospital, and justice data. Then the SEND and Alternative Provision Improvement Plan (HM Government, 2023) set out proposed national and local inclusion dashboards to present performance data across education, health, and care.

Parliamentary debate on the Health and Care Act (2022) prompted the Government to develop proposals to improve data sharing between health and social care to safeguard children. Their report to Parliament (DfE, 2023) cautiously supported using the NHS number as a consistent child identifier, proposing regional pilots and further work to improve the 'interoperability' of data systems and to build practitioners' confidence in sharing information.

The Labour manifesto (2024) set out to improve data sharing across services, with a single unique identifier (SUI), to better support children and families:

> *'Labour will improve coordination between education, social care and the wider services that support families by piloting the expansion of a children's number like the NHS number that stays with children not just for their school career but for their whole childhoods, ensuring that their needs are better met, and any issues are addressed early' (Labour Party, July 2024).*

A DfE policy paper, Keeping Children Safe, Helping Families Thrive: Breaking down barriers to opportunity (DfE, 2024), commits to developing a broader programme to improve data-sharing and linking, including piloting the implementation of the SUI, identifying and addressing 'system and process challenges', and a new duty 'that provides absolute clarity on the legal basis to share information for the purposes of safeguarding children' (p. 9-10).

Building on these commitments, the Children's Wellbeing and Schools Act has entered the statute book as of April 2026. This bill includes provisions to enable the introduction of a Consistent Identifier for children and to require its inclusion on certain types of data; however this is restricted to children's data to be used for the direct purpose of safeguarding or promoting the welfare of children. The Act also introduces a clearer duty to share information for these purposes, and provides for Information Standards on information-sharing and a Code of Practice on Consistent Identifiers.

1.4 Prevalence, terminology, and data gaps in learning disability documentation

This study utilises linked data to investigate the identification and documentation of learning disabilities across agencies in Sunderland. In the UK, around 2.5% of all children (around 349,000) are believed to have a learning disability (Mencap, 2024). The percentage of pupils in England with an EHCP has risen from 2.8% to 5.3% in the last 10 years, and the percentage of those on SEN support has risen in a similar fashion from 11.6% to 14.2%. A child's primary and secondary SEN needs, if they have any, are recorded in the school census data, and the rates of some learning difficulties have fluctuated over time as well. The percentage of children with a 'Moderate Learning Difficulty' has decreased over the last 10 years from 3.2% to 2.3%, while the percentage of those with a 'Specific Learning Difficulty' has risen from 17.7% to 20.0%. Percentages of those with 'Severe Learning Difficulty' and 'Profound and Multiple Learning Difficulties' have not considerably changed over the last 10 years (DfE, 2025).

These prevalence figures predominantly use education data, which refer to learning "difficulties" rather than "disabilities". While NHS GP registers primarily report prevalence of learning disabilities for all ages of patients (around 0.6% of registered patients), national child-specific averages are derived from population-based estimates rather than routine health datasets (Public Health England, 2024). The lack of child-specific health data for learning disabilities has been raised in other publications, such as a recent data mapping report published by King et al. (2022) through the NIHR's policy research unit. This report emphasises that, while there are a variety of sources of health data on learning disabilities (administrative datasets, longitudinal cohorts, surveys, registers, etc.), there are many gaps in learning disability documentation. The report also compares national prevalence figures for learning disability to the General Practice Learning Disability (GP LD) registers, and estimates that only 23% of those with a learning disability have a corresponding record in the GP LD register (King et al., 2022). This demonstrates the inconsistencies in identifying and reporting learning disabilities between datasets. This can have implications for children with learning disabilities, as they may be eligible for extra support, health checks, and reasonable adjustments (under the Equality Act, 2010) if their needs are recorded by relevant registers and agencies (NHS England, 2025).

While there are challenges involved in identifying children with learning disabilities, terminology inconsistencies compound these difficulties. Multiple terms are used to describe and document learning disabilities, and these vary across services and datasets. While it is understandable that each sector tailors definitions to its own priorities, fragmented terminology makes it difficult to build a coherent picture, supported by local and national data, of children with learning disabilities. This ambiguity surrounding learning disability terms, and the criteria or thresholds for their classification, is not unique to England; research indicates that these challenges persist internationally (Gabriel & Börnert-Ringleb, 2023; Archibald, 2024).

Problems with terms associated with learning difficulties and disabilities are particularly evident in defining different levels of learning disability severity. There are known differences in terms used to describe levels of severity used in health data, which takes its definitions from the World Health Organisation's International Classification of Diseases (ICD-11) (NICE, 2025), compared to the education data which uses its own set of severity band definitions (DfE, 2015). National guidance in education does not map IQ ranges to learning difficulty bands, however, there are some local areas who have produced documentation which does this. (Bradford Metropolitan District Council, 2020). The IQ ranges reported in these documents are offset from the definitions used in the ICD-11, as shown in Table 1. However, it should be clarified that no national guidance associates IQ ranges with education definitions of learning difficulties, and no formal IQ testing is used in the assessments of these needs:

Table 1. Cross-sector definitions of learning disability and difficulty severity levels

Term	Sector	Typical IQ Range	Description
Mild intellectual disability	Health (ICD-11 based)	50–69	Individuals typically experience difficulties in academic learning and abstract reasoning but can achieve independence in daily living with appropriate support.
Moderate Learning Difficulties (MLD)	Education (UK SEN)	50–70	Pupils who perform significantly below age-related expectations in most areas of the curriculum. Associated with general cognitive delay rather than specific deficits.
Moderate Intellectual Disability	Health (ICD-11 based)	35–49	A more severe form of intellectual impairment than mild ID. Individuals show marked developmental delays and limited academic attainment.
Severe Learning Difficulties (SLD)	Education (UK SEN)	35–50	Pupils with profound cognitive impairment affecting most aspects of learning and development. Major difficulties with communication, mobility, and self-care.
Severe Intellectual Disability	Health (ICD-11 based)	20–34	Indicates very limited intellectual functioning. Individuals have minimal academic skills and require pervasive support for all aspects of daily life.
Profound & Multiple Learning Difficulties (PMLD)	Education (UK SEN)	<35	Pupils with extremely complex needs across cognitive, physical, sensory, and communication domains. Require intensive, multi-disciplinary support and highly specialised educational environments.
Profound Intellectual Disability	Health (ICD-11 based)	<20	Represents the most severe level of intellectual impairment. Individuals have extremely limited functioning and require constant care.

*Source for health descriptions and IQ ranges associated with learning disabilities: Learning Disabilities Clinical Knowledge Summaries (NICE, 2025);
Source for education descriptions of learning difficulties: SEND Code of Practice (DfE, 2015);
Source for education IQ ranges associated with learning difficulties (**not** nationally defined): Working Definitions of Learning Difficulties (Bradford Metropolitan District Council, 2020)

In education, there is also confusion around some of these terms, particularly around the term Moderate Learning Difficulty (MLD). Studies and practitioner guidance note that MLD is often used as a residual or "catch-all" category for pupils who perform significantly below age-related expectations but do not fit neatly into other diagnostic groups (UCL Institute of Education, 2011). One study investigated the link between the statutory definition for MLD and the etymology of pupils who have been assigned this classification, and found that "the label of MLD is often used in an over-generalised way in schools" and that, with regards to the reliability and utility of the MLD level, there is "no evidence base for using the MLD category in some secondary schools" (Norwich, Ylonen and Gwernan-Jones, 2014).

Similar inconsistencies are evident in social care datasets, where disability status is often under-recorded. The CiN census contains data on disabilities and can specify a learning disability. Disabled children are entitled to have their needs assessed under s.17 of the Children Act (1989); however, it has been reported that this is often not the case. CIN Census guidance for 2022-23 (p.13) states that 'Only disabled children who have been assessed as requiring children's social care services should be included in the collection', so this dataset omits many disabled children. However, children with other primary needs (such as abuse or neglect) may also have a disability recorded. This is supported by a cohort study on the CiN census data subjects, showing that only around 5% of children in the CiN census are referred for a disability Emmott et al., 2022).

Existing literature has highlighted the challenges involved in identifying and describing children with learning difficulties or disabilities, which leads to gaps and shortcomings in the services provided for these children. Many of these challenges relate to inconsistencies between datasets in how children with learning disabilities are identified and documented, with further difficulties arising in the different terms used to describe these children and their specific needs. Data linkage can help evaluate the adequacy of support for children with learning disabilities, but this is only meaningful if their needs are documented consistently across all agencies, an issue which the present study aims to clarify.

2.0 Aims and objectives

This study has one aim and five objectives, as follows:

Aim 1:

To develop a linked cross-agency dataset for children to determine if those identified as having learning difficulties/learning disabilities have had a clinical paediatric assessment and investigations.

Objective 1:

Determine the rate of children recorded as having a learning difficulty in school census and/or learning disability in the children in need (CiN) census that have undergone a paediatric disability assessment and investigations as evidenced by representation in health data

Objective 2:

Determine the rate of children identified as having a learning disability in paediatric disability data who appear in school census and/or CiN census data

Objective 3:

Determine the rate of children with a learning disability in the CiN census who appear in school census data as having a learning difficulty

Objective 4:

Determine the characteristics of children with learning disabilities in the paediatric disability data compared to their learning difficulties recorded in the school census data

Objective 5:

Assess whether any clinical features (e.g. severity of learning disability) or demographic characteristics (e.g. school type, deprivation status) are associated with receiving a paediatric medical assessment

3.0 Methodology and methods

The datasets used in this study were provided by the North of England Commissioning Support Unit (NECS) and Together for Children (TfC). NECS received paediatric data from Sunderland and South Tyneside Foundation Trust (STSFT), including children's NHS numbers, which were processed using a pseudonymisation tool. This tool applies a one-way encryption algorithm to generate irreversible pseudonyms, ensuring that researchers cannot re-identify individuals without access to a separately held encryption key.

The same pseudonymisation process was applied to LA administrative data; however, these records typically do not contain children's NHS numbers. To enable linkage, NECS conducted a batch-tracing exercise, using children's demographic details (e.g. name, date of birth, postcode) to retrieve corresponding NHS numbers from health records. While not all matches were successful due to the limitations of batch-tracing (outdated address information, etc.), batch-tracing rates were still relatively high (see Table 2). Once batch-traced, TfC transferred the LA data to NECS for pseudonymisation. The LA data and the paediatric data were pseudonymised with the same pseudo key, meaning that an NHS number in the paediatric data that was also in the LA data would be encrypted into the same irreversible pseudonym. These pseudonymised LA datasets were returned to TfC and then shared with the research team.

To complete the data linkage process, the researchers used the pseudonymised NHS numbers as a unique identifier to link each individual's data across the various datasets. Most of the datasets did not originally use a structure of one row per individual; therefore, various transformations were carried out using packages in R Studio (R Core Team, 2025; Wickham et al. 2025). During these transformations, the priority was to retain as much data as possible and only perform aggregations when necessary, due to excessive numbers of rows per child. For example, the school exclusions data had one row per permanent exclusion/suspension. If a child experienced 35 school suspensions, aggregated columns were used (such as average length of exclusion, longest exclusion, total sessions missed) rather than including 35 separate columns in the dataset to exclusively capture all of this individual child's permanent exclusion/suspension details.

Most of the datasets were separated into three academic years, 2021/22, 2022/23 and 2023/24. SEN2 data was only processed at an individual-level from 2022/23 onwards[1] , and the health data had all acute outpatient appointments from 2019 onwards. Three linked datasets were created, with the 2021/22 dataset having one fewer dataset (SEN2). Health records were linked using the appointment dates. The decision was made to link all health records (if a respective pseudo-NHS number was matched) for the 2023/24 dataset to maximise the amount of health data available for records in that dataset. In the 2022/23 dataset, only health records from before the final day of the 2022/23 academic year were included to prevent conclusions being drawn from an individual's education/social care records from 2022/23 using health records from their 'future'. The same logic was applied to the 2021/22 dataset.

In NHS electronic medical records, there is a facility to add an electronic flag if a person of any age is identified as having a learning disability. These data were not included in the linked dataset, which only included those children with learning disabilities whose needs were reported when attending paediatric clinics during the time window of data capture: 2019-2024.

[1] The requirement for all local authorities to report pupil-level data in the SEN2 survey was only introduced in early 2023. Prior to this, it was a cohort-level survey.

4.0 Datasets and triangulating learning disability data

Datasets were linked using pseudonymised NHS numbers that were batch-traced from demographic information held by NECS. The LA supported this process through the application of a pseudonymisation tool before sharing datasets with NECS. Table 2 details the datasets received by the University of Sunderland (UoS), the years, total rows of data, the number of pseudonymised NHS numbers, and the overall match rate. The research team at the UoS undertook a data linkage exercise to produce one linked data set for the years (2022, 2023, and 2024) with a single row of data representing each child. This data allows for individual child-level and population-level analysis.

Table 2. The datasets provided to the project team by TfC and STSFT

Dataset	Years	Total rows	NHS numbers	NHS number match rate
Children in Need (CiN) Census	2021/22	5,046	4,402	87.2%
	2022/23	4,706	3,835	81.5%
	2023/24	4,752	4,034	84.9%
Attendance	2021/22	35,392	34,099	96.3%
	2022/23	36,768	35,271	96.0%
	2023/24	36,498	34,772	95.3%
Exclusions	2021/22	3,308	3,169	95.8%
	2022/23	6,125	5,926	96.8%
	2023/24	9,340	8,984	96.2%
SEN2	2022/23	3,003	2,835	94.4%
	2023/24	4,680	4,478	95.7%
School census	2021/22	48,812	42,622	87.3%
	2022/23	42,318	39,483	93.3%
	2023/24	42,675	39,816	93.3%
Alternative Provision Census	2021/22	142	131	92.3%
	2022/23	147	137	94.0%
	2023/24	191	180	94.2%
CSDS/ Acute outpatient data	Appointments from 2019-2024	20,887	All	N/A

The CiN census is a social care dataset derived from records held by English local authorities (only Sunderland in this study) on any child or young person referred to social care for a 'needs assessment', even if they are not ultimately assessed as needing social care support. A child may be referred for a needs assessment if they are identified as being exposed to (or are at risk of exposition to) adverse childhood experiences which require social care intervention (Emmott et al., 2019). Under the Children Act 1989 (s.17), disabled children are considered children in need and are eligible for an assessment for support.

Both the attendance and exclusion datasets are collected by Sunderland City Council and the data is held along with the school census by Together for Children (TfC). Attendance data are primarily structured on a one row per enrolment basis (a child will have two rows if they have attended two different establishments during the time period), while the exclusions data contains one row per exclusion.

The school census data is this study's primary source of demographic information, such as gender, ethnicity, free school meal eligibility, etc., and includes the most unique individuals of any dataset by a significant number. The dataset also includes SEN provision and SEN need information.

The SEN2 dataset includes any children who have an EHCP, and has only been available as a child-level dataset (rather than an aggregated dataset) since 2022/23. It includes detailed information about EHCP referrals, SEN needs, and any associated placements.

The alternative provision (AP) dataset records any children who are enrolled at a small number of AP establishments in Sunderland. It includes data on reasons for placement and the type of AP setting attended.

The CSDS is a patient-level, output-based, secondary uses data set which aspires to deliver 'robust, comprehensive, nationally consistent and comparable person-centred information' (NHS, 2025) for people who are in contact with publicly funded Community Health Services. The Acute Outpatient Data set (diagnostic data) was also provided.

4.1 Intersection of cohorts throughout the data

In the 2022 to 2024 school census datasets, there was variance in how many of these children appeared in other datasets, as follows:

There are 35,095 unique children in the 2022 school census dataset. The number of these children found in other datasets was:

- AP census (118 in dataset)
  - 4 children in both school census and AP census datasets
- CiN census (4,402 in dataset)
  - 2,710 children in both school census and CiN census datasets
- SEN2
  - No child-level SEN2 data for 2022
- CSDS (5,682 in dataset)

2,912 children in both school census and CSDS datasets

There are 39,136 unique children in the 2023 school census dataset. The number of these children found in other datasets was:

- AP census (121 in dataset)
  - 2 children in both school census and AP census datasets
- CiN census (3,834 in dataset)
  - 2,397 children in both school census and CiN census datasets
- SEN2 (2,707 in dataset)
  - 1,755 children in both school census and SEN2 datasets
- CSDS (7,745 in dataset)
  - 3,891 children in both school census and CSDS datasets

There are 39,128 unique children in the 2024 school census dataset. The number of these children found in other datasets was:

- AP census (132 in dataset)
  - 10 children in both school census and AP census datasets
- CiN census (4,034 in dataset)
  - 2,415 children in both school census and CiN census datasets
- SEN2 (3,024 in dataset)
  - 1,906 children in both school census and SEN2 datasets
- CSDS (8,874 in dataset)
  - 4,426 children in both school census and CSDS datasets

4.2 Demographics for the whole linked dataset

Table 3. Key demographic information from the linked data

	2022	2023	2024
Number of Records	39,059	39,136	39,128
Age mean (SD) range	9.86 (4.00) 2-19	9.88 (3.98) 2-19	9.86 (4.00), 2-20
Male - N (%)	20,048 (51.3%)	20,086 (51.3%)	20,117 (51.4%)
Female - N (%)	19,011 (48.7%)	19,049 (48.7%)	19,011 (48.6%)
English 1st lang – N (%)	36,921 (94.5%)	36,726 (93.8%)	36,492 (93.3%)
Ethnicity WB – N (%)	34,486 (88.6%)	33,975 (86.8%)	33,188 (84.8%)
Eligible FSM - N (%)	11,235 (28.8%)	11,538 (29.5%)	11,734 (30.0%)
Attending AP – N (%)	1,982 (5.1%)	1,716 (4.4%)	1,933 (4.9%)
EHCP – N (%)	1,498 (3.8%)	1,646 (4.2%)	1,789 (4.6%)
SEN Support – N (%)	5.673 (14.5%)	5,880 (15.0%)	6,186 (15.8%)
No SEN – N (%)	31,887 (81.6%)	31,610 (80.8%)	31,153 (79.6%)

In Table 3, there are some observable year-on-year trends throughout the linked dataset that contextualise later analyses of the data. As expected, the proportion of males and females in the data does not change over the course of the three academic years. The same can be said for the proportion of children whose first language is English. The proportion of children who were white British (WB) slightly decreased by around 2% each year over the three years. The number of children eligible for free school meals (FSM) saw a slight increase over the three years, although the increase is marginal. The proportion of children attending AP dropped from 2021/22 to 2022/23, but rose again in 2023/24. It is unclear why this would be the case and may have to do with data collection variance. The AP data has extremely low overlap with the school census dataset, which is used as the primary dataset for this study; therefore, the fluctuations in AP proportions will not be analysed or discussed further in this report. The number and proportion of children with an EHCP have seen a considerable relative increase year-on-year, while there has been a similar rise in children on SEN support. This corresponds with a fall in proportion of children with no type of SEN support, and reflects national trends (DfE, 2025).

4.3 Identifying learning disability terminology

Both school census and SEN2 census have a data field named 'SEN Type' that refers to learning 'difficulties' rather than 'disabilities'.

Learning difficulties values in this field are:

- Specific Learning Difficulty (SpLD)
- Moderate Learning Difficulty (MLD)
- Severe learning difficulty (SLD)
- Profound & Multiple Learning Difficulty (PMLD)

CiN census has a data field called 'Disability':

- The only value this field can record that is relevant to learning disabilities is "LD"
- It should be noted that, while 'Children in Need' does refer to a legal definition in the Children Act (1989, s 17(10)), which includes 'disabled children', a child will only be recorded as having a disability in the CiN census if they are receiving social care support.[2]
- As explained in section 1.4, children in the CiN census who have a disability are likely to only have this reported if this is their primary need.

The CSDS data has the data field 'Diagnosis'. There are numerous diagnoses associated with the broad term 'learning disabilities':

- Learning difficulties
- Developmental academic disorder
- Dyscalculia
- Dyslexia
- Dysgraphia
- Early Developmental Impairment/Unspecified intellectual developmental disability
- Borderline intellectual ability (IQ 70-85)
- Mild Intellectual Developmental Disability (IQ 50-70)
- Moderate Intellectual Developmental Disability (IQ 35-49)
- Severe Intellectual Developmental Disability (IQ 20-34)
- Profound Intellectual Developmental Disability (IQ <20)
- Specific Learning Disability
- Significant intellectual disability

While the variability in terminology referring to learning disabilities is vast, the researchers elected to include as many of these terms as possible when identifying children as 'having a learning disability' or 'not having a learning disability', which is the primary outcome measure used in the results section.

[2] It is unclear what proportion of all children with disabilities also receive support from social care. Between 2008 and 2016, only 4.8% of all CiN referrals were primarily due to disabilities (Emmott et al., 2019). There are also different local thresholds for access to social care, who may only get involved if children have the most severe or complex needs (CDC, 2024)

Furthermore, the DfE specifically includes 'Learning disability' as a factor that can be identified and recorded at the end of a CiN assessment:

"The learning disability factor should be recorded where a learning disability impacts on a child. This could include, but is not limited to, a reduced ability to understand new or complex information or to learn new skills, a reduced ability to cope independently and a lasting effect on development. Code (5A) should be used for concerns about the child's learning disability, (5B) for concerns about the parent's or carer's learning disability and (5C) for concerns about the learning disability of another person in the family/household." (DfE, 2021)

NB: Initially, the researchers considered filtering out 'Specific Learning Difficulties' and related subgroups such as dyslexia, dyscalculia, and dysgraphia, as well as related broad terms such as 'Developmental Academic Disorder'. This is because previous descriptive analyses on children with learning disabilities have excluded specific learning difficulties on the basis that this term does not imply an intellectual disability (Emerson et al., 2012). However, for reasons explained in section 5.1, the terms were retained when defining learning difficulty and disability outcomes in the data.

The following passage from the British Institute of Learning Disabilities' fact sheet explains intricacies of the terms 'Learning difficulty', 'Specific learning difficulty', and 'Learning disability':

> *"Many people with learning disabilities prefer to use the term 'learning difficulty'. The two terms are interchangeable when used in the context of health and social care for adults. However, in UK education services, the term 'learning difficulty' also includes people who have 'specific learning difficulties' (e.g., dyslexia), but who do not have a significant general impairment in intelligence. However, the SEN codes of 'moderate learning difficulty', 'severe learning difficulty' and 'profound multiple learning difficulty' all refer to generalised learning difficulty of varying severity. Taken together they can be considered to be interchangeable with the adult health and social care term 'learning disability'. However, people with specific learning difficulties such as dyslexia do not have 'learning disabilities'." (Holland, 2011)*

5.0 Results

5.1 Matching terminology across datasets

Prior to undertaking any analysis, the researchers investigated whether there was consistency in the level or severity of learning difficulty and disability terminology across the school census classifications and the paediatric diagnoses. Tables 4 and 5 demonstrate the extent to which the key severity terms (Moderate, Severe, Profound) are used consistently throughout a child's data.

Table 4. Breakdown of how all individuals (2023/24who have some type of "Intellectual Developmental Disability" documented in health records map to school census learning difficulty classifications (Primary need)

		Paediatric data – Intellectual Developmental Disabilities			
		Health classification			
		Mild (N = 7)	Moderate (N = 122)	Severe (N = 198)	Profound (N = 13)
School census – Primary need	Moderate LD	2	22	3	0
	Severe LD	1	18	70	2
	Profound & multiple LD	0	1	20	9
	Specific LD	1	4	0	0
	Autism	0	39	91	1
	SLCN	1	20	3	0
	SEMH	1	5	1	0
	Other	1	13	10	1
	None	0	0	0	0
Terminology match %		N/A	18.0%	35.4%	69.2%

Table 5. Breakdown of how all individuals (2023/24) who have some type of "Intellectual Developmental Disability" identified in health records map to school census learning difficulty classifications (Secondary need)

	Paediatric data – Intellectual Developmental Disabilities				
	Health classification				
		Mild (N = 7)	Moderate (N = 122)	Severe (N = 198)	Profound (N = 13)
School census – Secondary need	Moderate LD	0	13	3	0
	Severe LD	0	7	20	0
	Profound & multiple LD	0	0	1	0
	Specific LD	0	4	0	0
	Autism	1	8	14	0
	SLCN	3	23	55	4
	SEMH	0	7	2	0
	Other	0	9	31	5
	None	3	51	72	4

As shown in Table 4, having a 'Profound Intellectual Developmental Disability' in the healthcare data does fairly consistently map (69.2% of the time) to a primary need of 'Profound and multiple learning difficulties' in the school census (albeit from a small sample size). The same cannot be said for the larger groups of children with 'Severe Intellectual Developmental Disability' (35.4% match to 'SLD' as recorded primary SEN) or 'Moderate Intellectual Developmental Disability' (18.0% match to 'MLD' in SEN data). The consistency appears to increase as the level of severity increases.

These inconsistencies reflect known differences between learning difficulty terminology in education and learning disability terminology in health data (Section 1.4, Table 1). A core reason for the inconsistencies between these terms is that SEN reported in the school census does not need to be based on a formal assessment of cognitive abilities, but is rather used to help with planning and provision for these children (DfE, 2015).

While there is some evidence of shared terminology between these two datasets (and overlap of records when severity is higher), it is clear that they are not applied consistently if a child has a learning disability in both datasets (see Appendix 1 for a more detailed breakdown). Furthermore, these four health diagnoses only account for a portion of total learning disabilities recorded in the paediatric data. Consequently, the researchers elected to use a simple indicator for whether a child had any learning disability in the school census, CiN census, and/or the paediatric data.

At this stage of the study, the researchers decided to investigate the relationship between autism and learning difficulties, as there appeared to be numerous children who had a learning disability identified in the paediatric data but only had autism recorded in the school census, with no record of learning difficulty. As the school census allows for the recording of primary and secondary needs, it was necessary to check how frequently children with a paediatric record of a learning disability had no learning difficulties recorded as a primary or secondary need, and whether an autism record was present for either of these instances. Table 6 demonstrates the various permutations between having autism and/or having a learning difficulty in the school census primary or secondary SEN need records, and whether specific paediatric records of a learning disability were present.

Table 6 Number of children with autism and/or any type of learning difficulty (LD) in their school census data by intellectual disability identification in the paediatric data (2023/24).

	School census classification			
Paediatric learning disability classification	**Autism & LD**	**Autism and no LD**	**No Autism and LD**	**No Autism and no LD**
Mild intellectual disability	1	0	3	3
Moderate intellectual disability	19	28	50	25
Severe intellectual disability	32	73	85	8
Profound intellectual disability	0	1	11	1
Total – N (%)	52 (15.3%)	102 (30.0%)	149 (43.8%)	37 (10.9%)

NB: Table 6 includes only the 340 children (2023/24) with one of the four ICD-11 "Intellectual Developmental Disability" categories recorded in the paediatric data. This is a subset of the wider cohort of 1,038 children with any paediatric record of a learning disability that year (see Section 5.5)

The figures in Table 6 show that in 2023/24, 15.3% of all children who had some level of intellectual disability documented in the paediatric data had a learning difficulty and autism co-morbidity in their school census records. Crucially, the figures demonstrate that there is a large cohort of children (30%) who have some level of intellectual disability documented in the paediatric data (which falls under the learning disability umbrella), but only have an autism SEN need, without any mention of a learning difficulty in their school census records. With the addition of the children with no learning difficulty recorded and no autism recorded in their school census, the total percentage of children in 2023/24 who had a paediatric record of an intellectual developmental disability, but no record of any learning difficulty in the school census, rises to 40.9%. Further interrogation of the data showed that in 2023/24, 28.3% of children with a paediatric record of autism had no record of autism in their school census records, and this figure has decreased considerably year-on-year, indicating an improvement in the identification of autism in school census records.

Finally, while there is precedent to exclude children with a 'Specific learning difficulty' from learning disability analysis (see section 4.3), there is an observable lack of consistency in learning disability terminology between datasets, which would render this exclusion problematic. For example, there are instances in which a child has a dyslexia diagnosis in the paediatric data but has a 'Moderate learning difficulty' classification in the school census, rather than the expected 'Specific learning difficulty' classification. Excluding specific learning difficulties from the outcome measures would result in these children being identified as 'has a learning disability in the school census' and 'does not have a learning disability in the paediatric data', even though there is some paediatric data available about their learning needs. There were also numerous instances of children with a learning disability recorded in the health data who had a specific learning difficulty in the education data, and vice versa. This led the researchers to lean towards a wider approach to identifying children with learning disabilities, in order to capture as many children with learning needs as possible, regardless of their specific type of need. In this report, **'learning difficulty' refers to education data records, while 'learning disability' refers to health data records.**

5.2 Learning difficulty and disability by demographic characteristics

Table 7. Demographic breakdown of children with a learning difficulty recorded in the school census

		2021/22		2022/23		2023/24	
	Learning difficulty	**Has LD**	**No LD**	**Has LD**	**No LD**	**Has LD**	**No LD**
Gender – N (%)	Male	1397 (6.98%)	18629 (93.02%)	1392 (6.94%)	18675 (93.06%)	1478 (7.35%)	18620 (92.65%)
	Female	849 (4.47%)	18154 (95.53%)	868 (4.56%)	18173 (95.44%)	900 (4.74%)	18090 (95.26%)
SEN provision – N (%)	EHCP	451 (30.58%)	1024 (69.42%)	477 (29.23%)	1155 (70.77%)	533 (30.18%)	1233 (69.82%)
	SEN Support	1747 (30.80%)	3925 (69.20)	1743 (29.65%)	4135 (70.35%)	1817 (29.38%)	4368 (70.62%)
	No SEN	48 (0.15%)	31834 (99.85%)	40 (0.13%)	31559 (99.87%)	28 (0.09%)	31101 (99.91%)
FSM eligibility – N (%)	FSM Eligible	1180 (10.52%)	10039 (89.48%)	1184 (10.27%)	10346 (89.73%)	1287 (10.98%)	10433 (89.02%)
	Not FSM Eligible	1066 (3.83%)	26744 (96.17%)	1076 (3.90%)	26503 (96.10%)	1091 (3.99%)	26277 (96.01%)
Ethnicity – N (%)	White	2106 (6.00%)	32969 (94.00%)	2119 (6.14%)	32385 (93.86%)	2209 (6.55%)	31540 (93.45%)
	Asian	71 (4.28%)	1588 (95.72%)	69 (4.08%)	1621 (95.92%)	82 (4.67%)	1673 (95.33%)
	Black	10 (1.85%)	531 (98.15%)	13 (1.31%)	979 (98.69%)	30 (2.01%)	1460 (97.99%)
	Mixed	33 (4.25%)	743 (95.75%)	35 (4.26%)	787 (95.74%)	33 (3.84%)	827 (96.16%)
	Other/ unknown	26 (2.66%)	952 (97.34%)	24 (2.18%)	1077 (97.82%)	24 (1.94%)	1210 (98.06%)
School type attended – N (%)	Academy	1645 (5.69%)	27243 (94.31%)	1628 (5.62%)	27318 (94.38%)	1667 (5.79%)	27112 (94.21%)
	LA maintained	465 (4.98%)	8881 (91.05%)	481 (5.19%)	8783 (94.81%)	514 (5.50%)	8828 (94.50%)
	Free School	62 (4.98%)	631 (95.02%)	72 (9.27%)	705 (90.73%)	75 (8.90%)	768 (91.10%)
	Special school	74 (72.55%)	28 (27.45%)	79 (64.75%)	43 (35.25%)	74 (72.55%)	28 (27.45%)

Table 8. Demographic breakdown of children with a learning disability identified and documented in the paediatric data

		2021/22		2022/23		2023/24	
	Learning difficulty	**Has LD**	**No LD**	**Has LD**	**No LD**	**Has LD**	**No LD**
Gender – N (%)	**Male**	551 (2.75%)	19475 (97.25%)	673 (3.35%)	19394 (96.65%)	702 (3.49%)	19396 (96.51%)
	Female	269 (1.42%)	18734 (98.58%)	329 (1.73%)	18712 (98.27%)	336 (1.77%)	18654 (98.23%)
SEN provision – N (%)	**EHCP**	445 (30.17%)	1030 (69.83%)	551 (33.76%)	1081 (66.24%)	596 (33.75%)	1170 (66.25%)
	SEN Support	320 (5.64%)	5352 (94.36%)	392 (6.67%)	5486 (93.33%)	380 (6.14%)	5805 (93.86%)
	No SEN	55 (0.17%)	31827 (99.83%)	59 (0.19%)	31540 (99.81%)	62 (0.20%)	31075 (99.80%)
FSM eligibility – N (%)	**FSM Eligible**	391 (3.49%)	10828 (96.51%)	482 (4.18%)	11048 (95.82%)	509 (4.34%)	11211 (95.66%)
	Not FSM Eligible	429 (1.54%)	27381 (98.46%)	520 (1.89%)	27059 (98.11%)	529 (1.93%)	26839 (98.07%)
Ethnicity – N (%)	**White**	741 (2.11%)	34334 (97.89%)	910 (2.64%)	33594 (97.36%)	928 (2.75%)	32821 (97.25%)
	Asian	36 (2.17%)	1623 (97.83%)	46 (2.72%)	1644 (97.28%)	55 (3.13%)	1700 (96.87%)
	Black	6 (1.11%)	535 (98.89%)	13 (1.31%)	979 (98.69%)	16 (1.07%)	1474 (98.93%)
	Mixed	15 (1.93%)	761 (98.07%)	15 (1.82%)	807 (98.18%)	16 (1.86%)	844 (98.14%)
	Other/ unknown	22 (2.25%)	956 (97.75%)	18 (1.63%)	1083 (98.37%)	23 (1.86%)	1211 (98.14%)
School type attended – N (%)	**Academy**	507 (1.76%)	28381 (98.24%)	628 (2.17%)	28318 (97.83%)	653 (2.27%)	28126 (97.73%)
	LA maintained	197 (2.11%)	9149 (97.89%)	228 (2.46%)	9036 (97.54%)	223 (2.39%)	9119 (97.61%)
	Free School	32 (4.62%)	661 (95.38%)	54 (6.95%)	723 (93.05%)	71 (8.42%)	772 (91.58%)
	Special school	84 (82.35%)	18 (17.65%)	92 (75.41%)	30 (75.41%)	91 (73.39%)	33 (26.61%)

Tables 7 and 8 show relatively similar demographic patterns throughout the learning disability data, considering that far fewer children have a learning disability identified in the paediatric data than have a school census learning difficulty classification. More boys have a learning disability than girls throughout both datasets, and the same goes for children who are eligible for FSM (although the difference is greater in the school census data).

Differences arise when looking at the relationship between SEN provision and learning disabilities, as well as ethnicity and learning disabilities. The school census data shows that around the same proportion (29-31%) of those on SEN support and those with an EHCP have a learning difficulty recorded in the education data, but in the paediatric data, there is a considerable difference between the two groups. The proportion of those with an EHCP who had an identified learning disability was still around 30-34%, but only around 5-7% of children on SEN support had a learning disability identified in the paediatric data. This is perhaps as one would expect, as EHCPs are intended for children requiring a higher level of specialist support, based on a statutory needs assessment. Additionally, if more years of paediatric data were available it is likely that more children with SEN support would be identified as having a learning disability in the paediatric data. (DfE, 2015).

Furthermore, the proportion of white children who have a learning difficulty (6-7%) was higher than other ethnicities (1-5%) in the school census data, but these differences are smaller in the paediatric data (proportion of around 1-3% for all ethnicities). It should be noted, however, that ethnic subgroups are exceedingly small in the paediatric learning disability data, and this could contribute to the difference between the trends across the two datasets.

Finally, there are some small differences between the datasets in how those with learning disabilities are distributed throughout different types of school. In both datasets, children in special schools are more likely to have a learning disability (around 65-75% across all years and both datasets). The proportion of those with learning disabilities across other types of school fluctuates year-by-year and there are some differences between the datasets, such as the sudden drop in proportion of those with a learning difficulty classification in LA maintained schools after 2022/23, a temporal trend which is not present in the paediatric data.

Other temporal trends exist within the data on school types; these figures fluctuate considerably between years in a fashion that cannot be identified in any other demographic or school-related characteristics. Firstly, across both datasets, the proportion of children in special schools who have a learning disability dropped from 2021/22 to 2022/23, however, in the paediatric data, this proportion does not change between 2022/23, while the proportion drastically rose in the school census data to nearly 100%. Secondly, the proportion of those with learning disabilities in free schools rises considerably after 2021/22 in both datasets from around 5% to 9%, although this rise happens over one year in the school census data, and happens over two years in the paediatric data.

For all of these temporal trends, it is not possible with descriptive statistics to infer whether changes are attributable to subgroup-specific trends in learning disability prevalence, or due to population-level fluctuations in learning disability prevalence (of which there is some documentation in section 1.4). The effect of demographic characteristics on likelihood of having an identified learning difficulty and/or disability is explored in section 5.7 through multi-level statistical modelling.

5.3 Distribution of learning disabilities throughout key datasets

Figure 1. Venn diagram showing the distribution of children with learning disabilities throughout the three key datasets in 2021/22

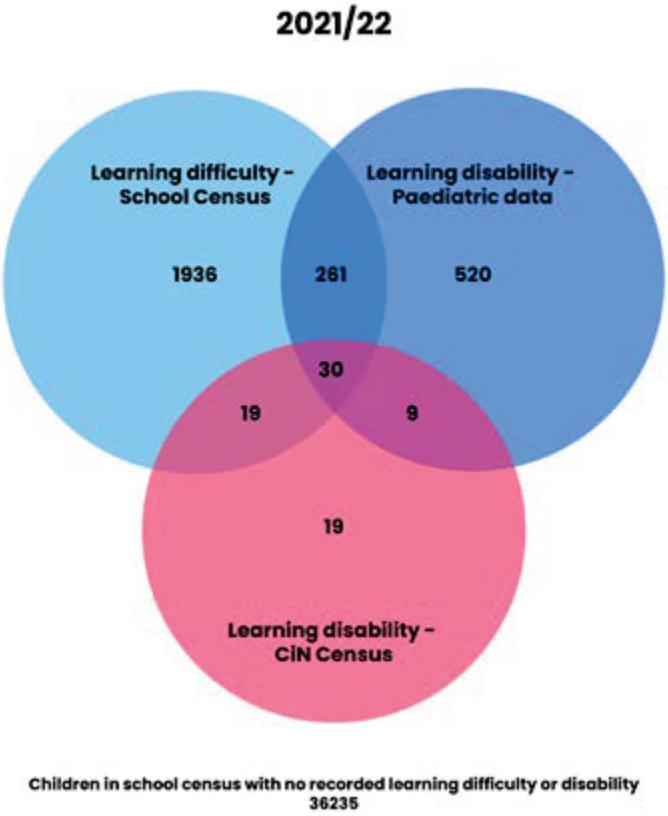


Figure 2. Venn diagram showing the distribution of children with learning disabilities throughout the three key datasets in 2022/23

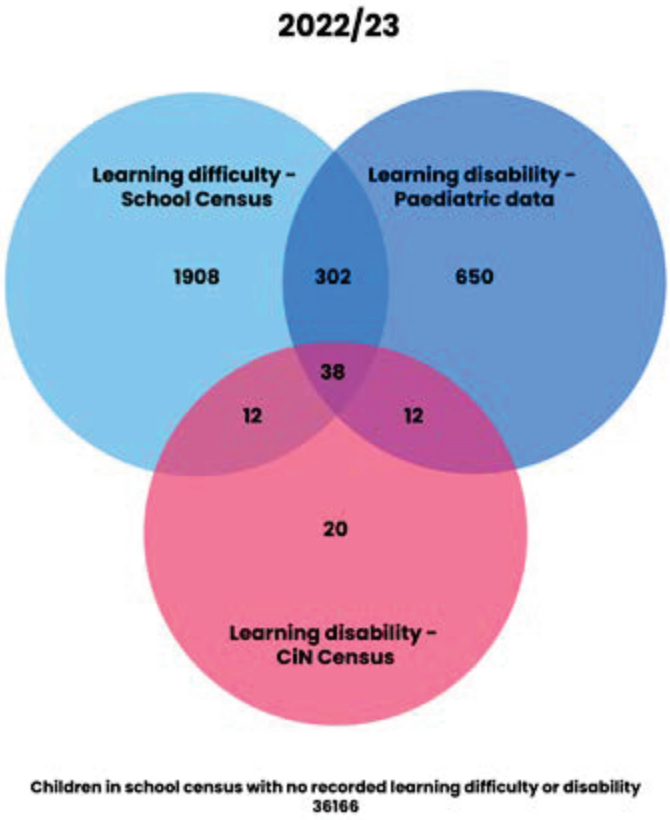
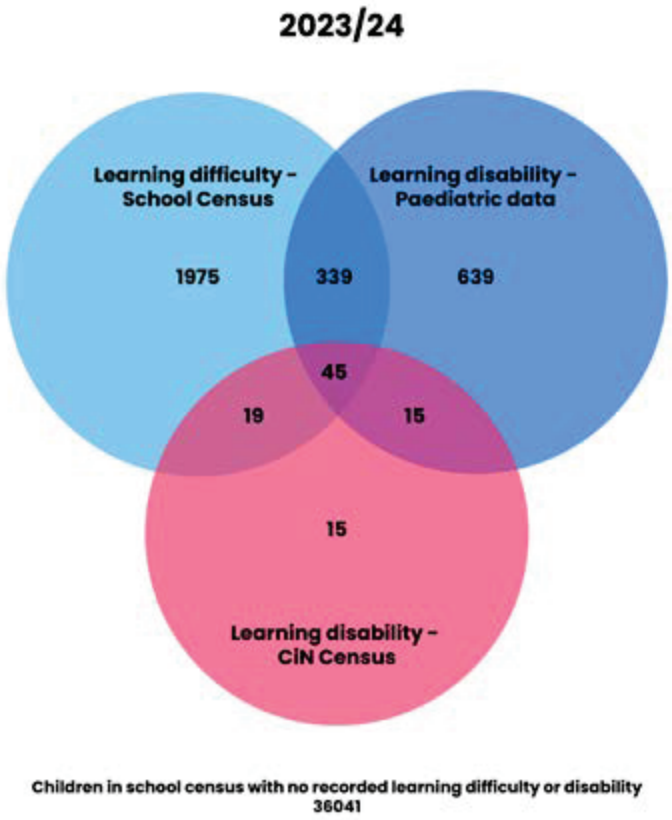


Figure 3. Venn diagram showing the distribution of children with learning disabilities throughout the three key datasets in 2023/24



Figures 1, 2, and 3 illustrate the extent to which children's learning disabilities are consistently recorded across datasets. There is little that separates the three years of academic data in this regard, other than an increase in numbers within all of the 'Learning disability – CSDS' segments, which is almost certainly due to the fact that there is more paediatric data associated with the 2023/24 academic year linked dataset.

For both the school census and the paediatric data, there are high numbers of children whose learning disabilities are only found in one of those datasets, and the level of overlap in each of the three years is relatively low. It is notable that there are considerably more children with learning disabilities recorded across multiple datasets when their disability is known to social care: this could reflect high eligibility thresholds for support from children's social care (see section 1.4), meaning that children known to social care for a disability tend to have more clearly documented and severe needs that are also visible to other agencies. This pattern is further supported by findings in section 5.6. The Venn diagram demonstrates that there is some level of overlap between learning disability identification in the school census and in the paediatric data, but there are a large number of those in the school census whose disabilities are not recorded in the paediatric data. Again, this could be explained by the limited time window of data capture for paediatric data (2019-2024).

5.4 Objective 1 findings

Objective 1: To determine the rate of children recorded as having a learning difficulty in school census and/or learning disability in the CiN census that have undergone a paediatric disability assessment and investigations, as evidenced by representation in health data between 2019 and 2024.

The rate of children with a "learning difficulty" in the school census who also had a learning disability in the paediatric data are:

0.129 in 2021/22

- 2246 children had a learning difficulty according to their school census records. 291 of these children also had a learning disability documented in health records (12.9%).

0.150 in 2022/23

- 2260 children had a learning difficulty according to their school census records. 340 of these children also had a learning disability documented in health records (15.0%).

0.161 in 2023/24

- 2378 children had a learning difficulty according to their school census records. 384 of these children also had a learning disability documented in health records (16.1%).

The rate of children with a "learning disability" in the CiN census who also had a learning disability in the paediatric data are:

0.506 in 2021/22

- 77 children had a learning disability according to their CiN census records. 39 of these children also had a learning disability documented in health records (50.6%).

0.610 in 2022/23

- 82 children had a learning disability according to their CiN census records. 50 of these children also had a learning disability documented in health records (61.0%).

0.638 in 2023/24

- 94 children had a learning disability according to their CiN census records. 60 of these children also had a learning disability documented in health records (63.8%).

While the linked dataset can provide an answer as to the representation of these two groups of children in the paediatric data, there is insufficient data on the specific assessments and investigations they have undergone, and only the presence of an identified learning disability can be used to address this objective. Furthermore, it is crucial to keep in mind that any figures regarding paediatric learning disability identification only refer to children who had a paediatric appointment between 2019-2024.

A relatively low rate of children with a school census learning difficulty classification are known to have a learning disability identified in the paediatric data. Conversely, by 2023/24, the majority of children with a known learning disability according to the CiN census have a learning disability identified in the paediatric data. This likely reflects a number of factors: the different size of these cohorts (with education being a universal service while access to support from children's social care is restricted to those who meet eligibility criteria) and the looser approach to recording type of SEN in the school census (DfE, 2015), as compared to clinical assessment.

5.5 Objective 2 findings

Objective 2: To determine the rate of children identified as having a learning disability in paediatric disability data who also appear in school census and/or CiN census data.

School Census

The rate of children who had any learning disability reported in the paediatric disability data who are also recorded as having learning difficulties in the school census data was:

0.355 in 2021/22

- 820 children had an identified learning disability in the paediatric data. 291 of these children also had some type of learning difficulty in their school census records (35.5%).

0.339 in 2022/23

- 1002 children had an identified learning disability in the paediatric disability data. 340 of these children also had some type of learning difficulty in their school census records (33.9%).

0.370 in 2023/24

- 1038 children had an identified learning disability in the paediatric disability data. 384 of these children also had some type of learning difficulty in their school census records (37.0%).

CiN Census

The rate of children who had any learning disability reported in the paediatric disability data who are also recorded as having a learning disability in the CiN census data was:

0.047 in 2021/22

- 820 children had an identified learning disability in the paediatric disability data. 39 of these children also had a learning disability in their CiN census records (4.7%).

0.050 in 2022/23

- 1002 children had an identified learning disability in the paediatric disability data. 50 of these children also had a learning disability in their CiN census records (5.0%).

0.058 in 2023/24

- 1038 children had an identified learning disability in the paediatric disability data. 60 of these children also had a learning disability in their CiN census records (5.8%).

The findings related to Objective 2 offer important insight into the consistency with which learning disabilities are documented across key administrative datasets. While it is reasonable to expect that some children recorded as having learning difficulties in the school census may not present needs severe enough to warrant paediatric intervention (as shown in the section **5**.4 findings), it is notable that only 33–37% of children with a learning disability identified and documented in the paediatric data lack any corresponding record of learning difficulties in the school census. It should also be noted that there will be children with a documented learning disability who are not present in this dataset (if they have not attended an appointment between 2019-2024), and it is not known how many of these children would have a corresponding school census learning difficulty classification.

The low rates observed in the CiN census findings (4-6% of learning disability records in the paediatric data had corresponding CiN census learning disability records) are more understandable, as although disabled children are entitled to have their needs assessed (Children Act 1989, s.17), they do not always require social care intervention, and these children would not have their learning disabilities recorded in the CiN census (see section 4.3). It could also be explained by different thresholds for access to children's social care, who are more likely to be involved with families where children have the most severe learning disabilities.

5.6 Objective 3 findings

Objective 3: To determine the rate of children with a learning disability in the CiN census who appear in school census data as having a learning difficulty.

The rate of children who had a learning disability in the CiN census and also had a learning difficulty in their school census data was:

0.636 in 2021/22

- 77 children had a learning disability according to their CiN census data. 49 of these children also had some type of learning difficulty in their school census records (63.6%).

0.610 in 2022/23

- 82 children had a learning disability according to their CiN census data. 50 of these children also had some type of learning difficulty in their school census records (61.0%).

0.681 in 2023/24

- 94 children had a learning disability according to their CiN census data. 64 of these children also had some type of learning difficulty in their school census records (68.1%).

The rates calculated to address Objective 3 demonstrate one of the more consistent links in learning disability identification across the linked dataset. 61-68% of children who had a learning disability recorded in the CiN census had a corresponding learning difficulty classification in their school census records. While this relationship is stronger than most other associations of learning disability identification across datasets, there are nevertheless a considerable number of children whose learning disability is known to social care and recorded in their CiN census, but this disability is not recorded in their school census records.

5.7 Objective 4 findings

Objective 4: To determine the characteristics of children with learning disabilities recorded in the paediatric disability data compared to their learning difficulties recorded in the school census data.

While there was confidence that the education data used in the project captured the majority of children with "learning difficulties recorded in the school census data", there were known limitations around the extent to which the healthcare data contained the majority of children with "learning disabilities recorded in the paediatric disability data", due to the timeframe limitations laid out in section 3.0. For this reason, results that used inferential statistics to investigate the characteristics of children with a paediatric record of a learning disability are omitted from section 5.7, with the caveated findings reported in Appendix 2. The focus in this section will be limited to the characteristics of children with a learning difficulty recorded in the school census data.

To address this objective, the researchers chose to fit and compare mixed-effect general linear models. Analyses were carried out using the R packages (mlmRev, lme4, sjPlot, DHARMa) (Bates et al., 2025; Lüdecke, 2025; Hartig, 2025), and the school a child attended was included as a random effect. This decision was made to allow each school to have its own baseline odds of its children having a learning disability (LD), regardless of predictors. These school-level differences are well-documented, and a recent study by the Educational Policy Institute suggests that up to two thirds of variation in SEND identification could be attributed to the school a child attended (Hutchinson, Downs and Ford, 2025).

Figure 4.

Caterpillar plot of school-level random effects from a multilevel logistic regression model. Each point represents a different school, and its position in the chart represents the estimated baseline log-odds of a child having a learning difficulty recorded for a given school, after accounting for fixed effects in the model. Horizontal lines indicate 95% confidence intervals around each estimate. Schools are anonymised and ordered by their estimated log-odds. The red-dotted vertical line at zero represents the average log-odds across all schools; schools to the right have higher-than-average baseline odds, while those to the left have lower-than-average odds.

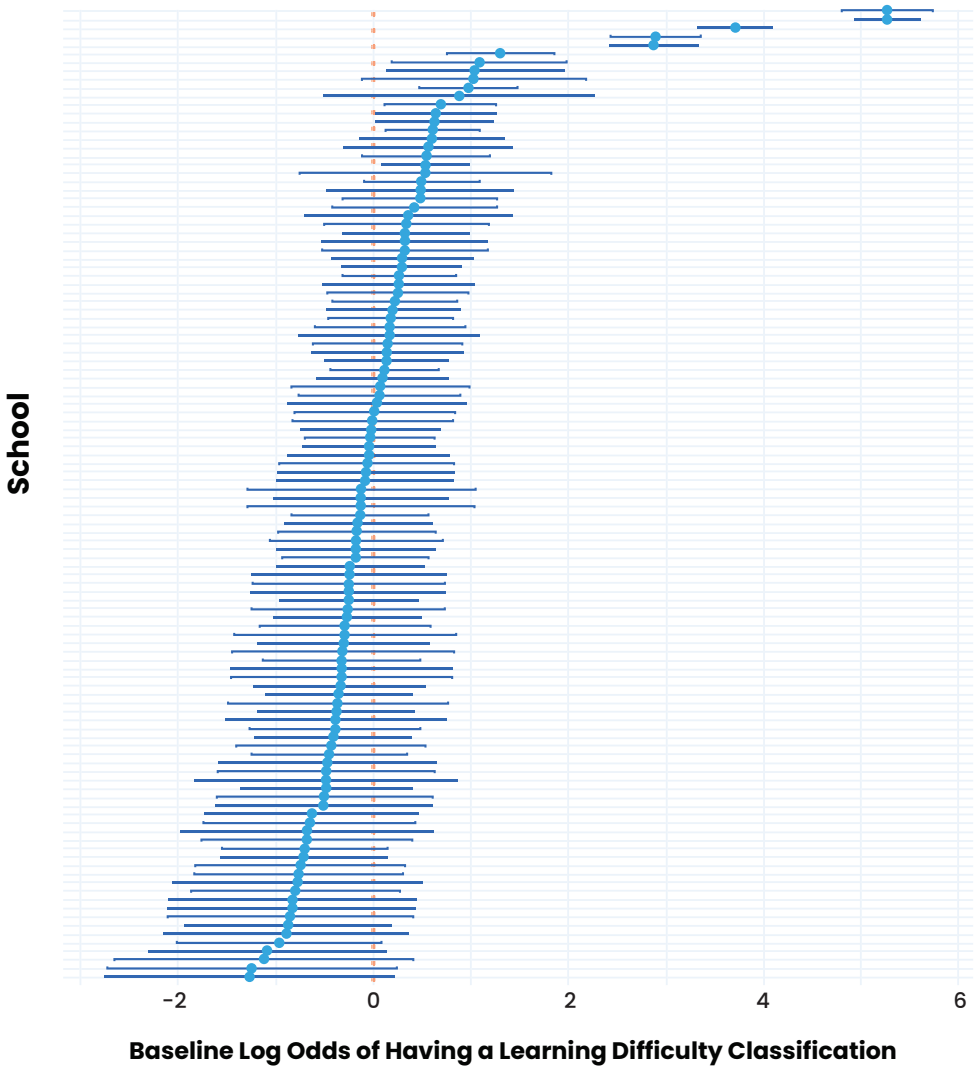


Figure 4 illustrates the extent to which the odds of a child having a learning disability classification (school census) can vary based on their school. Particularly, there are some schools (mostly special schools) that are highly skewed in the sense that children at these schools have much higher odds of having a learning difficulty compared to the average across all schools (and the smaller confidence interval ranges of these schools indicate that this is a robust association). The inclusion of this random effect was carried over into all other mixed-effect models fitted in this study. School type (e.g., Academy, special school, etc.) was initially considered as a potential random effect or predictor, however sensitivity analyses confirmed that its inclusion did not substantially alter the estimates of other predictors compared to using schools as the random effect.

The random effect of school was built into the three general linear models, one for each academic year. The models use learning difficulty recorded in the school census as the outcome variable.

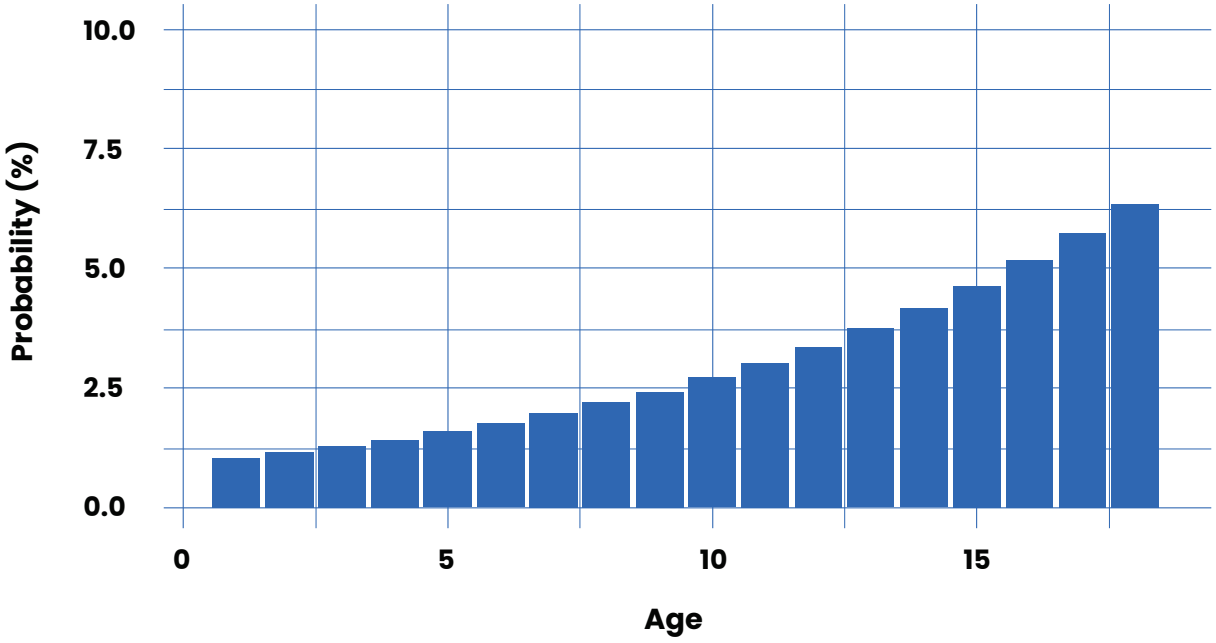
2021/22

Table 9. Logistic mixed-effects model predicting classification of a learning difficulty in the school census (2021/22) using demographic predictors and school as a random effect (N = 39,029)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.01	0.00	0.01 - 0.01	<0.001
Age	1.12	0.01	11.09 - 1.14	<0.001
Gender (Male)	1.91	0.13	1.67 - 2.18	<0.001
FSM Eligibility (Eligible)	2.82	0.21	2.44 - 3.27	<0.001
Asian (vs White)	0.67	0.09	0.51 - 0.89	0.005
Black (vs White)	0.35	0.11	0.19 - 0.66	0.001
Mixed (vs White)	0.62	0.11	0.43 - 0.89	0.009
Other or unknown (vs White)	0.54	0.11	0.36 - 0.81	0.003
Gender × FSM eligibility	0.75	0.07	0.63 - 0.90	0.02
Random Effects				
σ2	3.29			
τ00 School	0.61			
ICC	0.16			
N School	114			
Observations	39029			
Marginal R^2 / Conditional R^2	0.117 / 0.256			
AIC	15429.266			

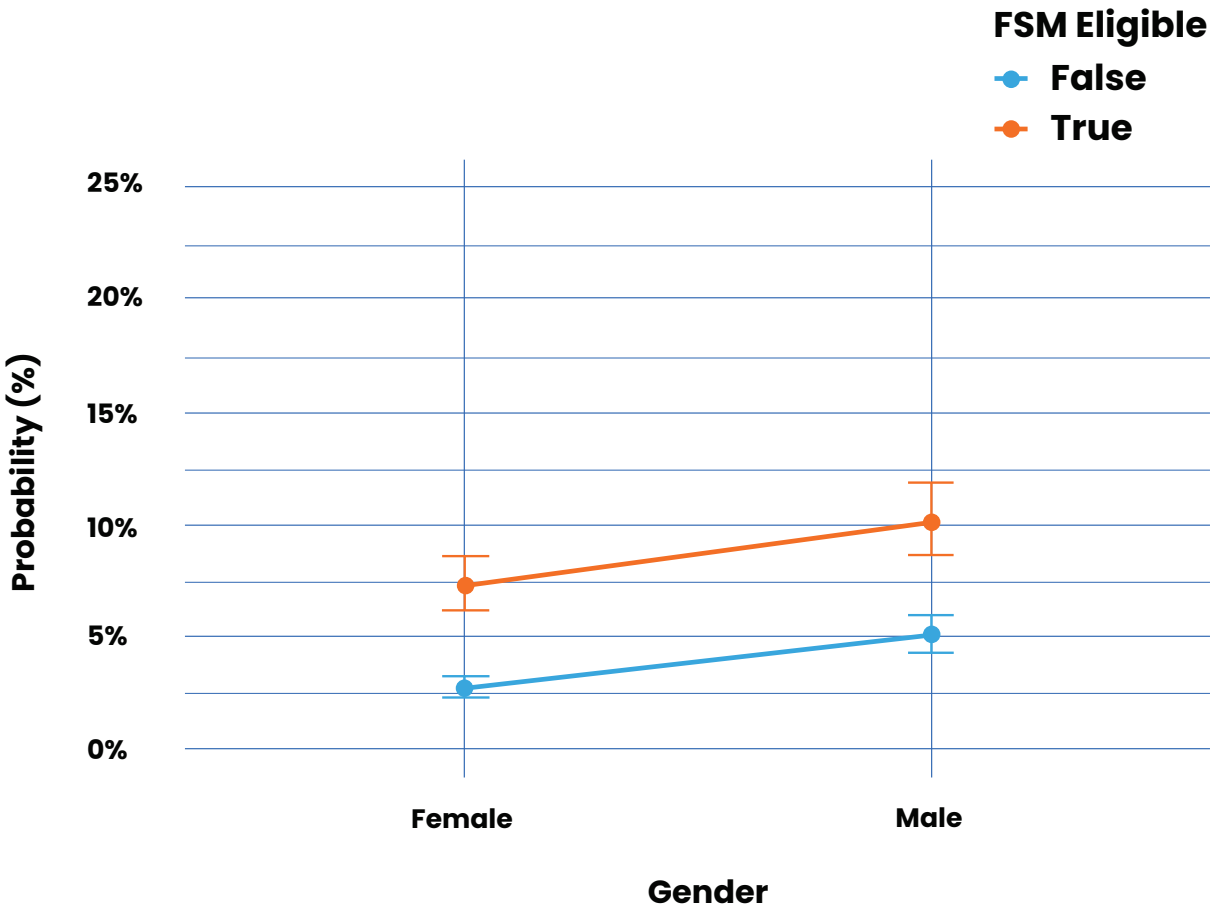
In the 2021/22 data, the odds of whether a child had a recorded learning difficulty in the school census were significantly higher for males (91% higher) and for those eligible for FSM (182% higher) as indicated by the extremely low p-values. Age was included as a covariate, as there is a relationship between age and being eligible for FSM (see Appendix 3). Note that this association was not an interaction term that affected the outcome. However, it is pertinent to note that age also had a significant association with having a recorded school census learning difficulty (Figure 5), with the odds of this outcome increasing by around 12% with each year of age. This likely reflects assessment practices and age milestones used in classifying SEND (e.g., the gradual accumulation of evidence across years) rather than a difference in prevalence of learning difficulties depending on age (DfE, 2015).

Figure 5. Predicted probabilities of learning difficulty classification in the school census by age.



The interaction between FSM eligibility and gender was included because it improved model fit (lower AIC). In the 2021/22 data, the combined effect of being male and FSM-eligible was slightly less than additive, as indicated by the interaction term. While the interaction is statistically significant, plotting the predicted probabilities of the four subgroups illustrates the small size of the interaction compared to the main effects of gender and FSM eligibility on their own (see Figure 6).

Figure 6. Plotted predicted probabilities for the FSM eligibility and gender interaction found in the mixed-effects model in Table 9.



Another characteristic that significantly predicted the odds of having a school census record of a learning difficulty was ethnicity. Identifying as any ethnicity other than white significantly reduced the odds of having a school census learning difficulty classification. However, it should be noted that the ethnicity subgroups are considerably smaller than other demographic subgroups, and caution should be taken in interpreting these ethnicity effects. More discrete ethnicity data was available; however, these subgroups would have been even smaller. Therefore, these groups were aggregated into higher-order ethnicity groups, a practice that is common but is often cautioned against. A longitudinal study by Strand and Lindorff (2021) exposes these difficulties in investigating the effects of ethnicity, but also reports significant 'disproportionality' of ethnic minority children in SEND identification (including underrepresentation of these children in learning difficulty data).

Furthermore, the intra-class correlation coefficient (ICC) in Table 9 suggests that school-level differences account for 16% of the variance in learning difficulty classification. This ICC value contextualises the final conditional R2 value of 0.256, suggesting that the demographic predictors and school-level differences account for just over 25% of the variability in whether a child has a school census record of a learning difficulty.

2022/23

Table 10. Logistic mixed-effects model predicting classification of a learning difficulty in the school census (2022/23) using demographic predictors and school as a random effect (N = 39,108)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.01	0.00	0.01 – 0.01	<0.001
Age	1.12	0.01	1.10 – 1.15	<0.001
Gender (Male)	1.76	0.12	1.54 – 2.01	<0.001
FSM Eligibility (Eligible)	2.53	0.19	2.19 – 2.92	<0.001
Asian (vs White)	0.60	0.09	0.45 – 0.79	<0.001
Black (vs White)	0.23	0.06	0.13 – 0.40	<0.001
Mixed (vs White)	0.63	0.11	0.44 – 0.90	0.011
Other or unknown (vs White)	0.40	0.08	0.26 – 0.60	<0.001
Gender × FSM eligibility	0.79	0.07	0.66 – 0.94	0.010
Random Effects				
σ2	3.29			
τ00 School	0.66			
ICC	0.17			
N School	115			
Observations	39108			
Marginal R^2 / Conditional R^2	0.127 / 0.274			
AIC	15510.782			

The fitted model for the 2022/23 data closely mirrors that of the 2021/22 data. Males and those eligible for FSM were significantly more likely to have a learning difficulty in the school census than their counterparts (females and those not eligible for FSM). The age effect also persists in the same direction in this year of data, and the same can be said for the gender x FSM eligibility interaction. Children in all ethnic groups other than white were still significantly less likely to have a learning difficulty recorded in the school census than white children. The same level of caution should be observed in interpreting these results due to small subgroup sizes. The conditional R2 for this model is slightly higher than the previous model, indicating that around 27% of the variation in learning difficulties recorded can be accounted for by the model's predictors and school-level differences. These school-level differences accounted for a marginally higher amount of variation in the 2022/23 model (ICC = 0.17).

2023/24

Table 11. Logistic mixed-effects model predicting classification of a learning difficulty in the school census (2023/24) using demographic predictors and school as a random effect (N = 39,088)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.01	0.00	0.01 – 0.02	<0.001
Age	1.09	0.01	1.07 – 1.12	<0.001
Gender (Male)	1.70	0.12	1.49 – 1.94	<0.001
FSM Eligibility (Eligible)	2.51	0.19	2.17 – 2.90	<0.001
Asian (vs White)	0.63	0.08	0.48 – 0.81	<0.001
Black (vs White)	0.32	0.06	0.22 – 0.47	<0.001
Mixed (vs White)	0.53	0.10	0.37 – 0.76	<0.001
Other or unknown (vs White)	0.33	0.07	0.21 – 0.49	<0.001
Gender × FSM eligibility	0.87	0.08	0.72 – 1.04	0.116
Random Effects				
σ2	3.29			
τ00 School	0.89			
ICC	0.21			
N School	114			
Observations	39088			
Marginal R^2 / Conditional R^2	0.110 / 0.299			
AIC	15805.025			

In the 2023/24 data, the same strong effects of gender and FSM eligibility on learning difficulties recorded in the school census are present. The differences in ethnicity effects between the school census model and the paediatric data model mirrored the differences in the 2022/23 data as well. The main difference between the 2023/24 model and the previous two models is that the initially observed interaction between gender and FSM eligibility in the school census model is no longer statistically significant in 2023/24. The overall R2 and conditional R2 figures, as well as the ICC, are also broadly similar to previous academic years.

5.8 Objective 5 findings

Objective 5: To assess whether any clinical features (e.g., severity of learning disability, other medical conditions) or demographic characteristics (e.g., school type, deprivation status) are associated with receiving a paediatric medical assessment.

Due to the limitations of the paediatric data available for this study (see section 3.0), it was not possible to investigate associations between demographic or clinical predictors and paediatric medical assessments. Multi-level models were fitted using demographic predictors, with the addition of some clinical predictors such as other school census SEN needs (other than learning difficulties). As the paediatric data available are only a snapshot of children in Sunderland who have received a paediatric medical assessment for a learning disability, the findings related to Objective 5 will not be reported. The limited inferential analysis that was undertaken using paediatric record of an identified learning disability as the outcome measure can be found in Appendix 4, and should be interpreted with caution due to the limitations of the paediatric data laid out in prior sections (Section 3.0, Section 5.7).

6.0 Concluding remarks

This study provides insight into the potential of data linkage, and how its utility can reach beyond its clear administrative benefits. While the information governance processes were arduous, the project allowed for the linking of numerous data sets across education, social care, and healthcare, and an evaluation of consistency in learning disability records across these domains.

While addressing the service evaluation objectives was the main focus of this study, one of the key findings concerns the disparities in learning disability terminology between education and paediatric care records. While there is some shared terminology regarding the severity of a learning disability, they take on different meanings for each agency (Table 1), and this is reflected in the data; the terms are not used consistently within individual children's records, which could be confusing for families whose children have different levels of learning disability 'severity' within their records. This is exacerbated by the use of some legacy diagnostic terms that appear infrequently in the paediatric data, but do indicate the presence of needs relating to a learning disability.

Evaluating the consistency in learning disability records across health and education also revealed that a large proportion of children with a learning disability identified in the paediatric data did not have a corresponding learning difficulty classification. While all of these children did have some type of recorded SEN need, they usually had a classification of autism, Speech, Language and Communication Needs (SLCN) or Social, Emotional and Mental Health needs (SEMH) instead. This highlights the limitations of SEN primary needs data. Some children could have been misidentified as having one of these alternate needs or, likely more commonly, they have co-morbidities with learning difficulties, which have been recorded as their primary need. There is concern that an autism classification could be obscuring needs related to a learning disability, and that these children may not receive the adjustments or support they need regarding their learning disability.

Objectives 1, 2, and 3 related to whether children with any recorded learning disability in one or more data sets had a corresponding learning disability recorded in another data set. While having a learning disability record in the CiN census usually meant that a child had a learning disability record in their school census data and/or paediatric data, thought to reflect higher eligibility criteria for access to social care services, this was not always the case. However, the rate at which children's learning disability records were matched between the school census data and the paediatric data was much lower (Section 5.1 and 5.2). Only around 10-16% of those who had a school census learning difficulty classification also had a corresponding learning disability identified in the paediatric data. However, some of this disparity can be attributed to the limited timeframe of data provided for the health records (Section 3.0).

A finding that is harder to attribute to the data's limitations is the fact that only around 33-37% of children who have a learning disability reported in the paediatric data also have any learning difficulty classification in their school census data (Section 5.2). This suggests that there are children with paediatrically documented learning disabilities whose schools have not accurately identified and/or recorded their associated SEN needs. This finding is supplemented by the descriptive analyses in Section 6.1 (and Appendix 1) around which alternative SEN needs may be obscuring these learning disabilities. The rate of children with a learning disability identified in the paediatric data who have a social care learning disability record is also extremely low (4-6%). However, this is more consistent with precedent and does not necessarily capture all paediatric diagnoses. The rates calculated to address Objectives 1, 2, and 3 did not considerably vary between the three academic years of data.

Objective 4 poses the question of whether there are any differences in which demographic characteristics can be used to predict either a learning disability identified in the paediatric data, or a school census learning difficulty classification. A mixed-effects general linear model was fitted to the data for each academic year, with the aim of estimating the odds of having a learning difficulty classification in the school census depending on the presence of certain demographic characteristics. The school a child attended was added as a random effect, as there is considerable variation in baseline odds of having a learning disability between schools.

Across all three years, being male or being eligible for FSM significantly increased the odds of having either of the two outcomes, while FSM eligibility was the stronger of the two effects in each model. The odds of having a learning difficulty record in the school census also increased with age; however, there are various contextual factors around when assessments of needs happen in schools that could account for this association. With regards to ethnicity effects, non-white ethnic groups were significantly less likely to have a learning difficulty record in the school census; however, the number of individuals in some ethnic groups who also had a record of a learning difficulty in the school census was small, and these particular results should be interpreted with caution. Overall, the models were able to explain around 25-30% of variation in which children had a record of a learning difficulty in the school census, with a considerable amount of this variation explained by school-level differences.

The present study was not able to confidently address elements of Objective 5 due to limitations of the paediatric data, namely the fact that the snapshot of data does not include all children with a paediatric record of a learning disability. Findings regarding this objective are omitted from the main report.

7.0 Limitations

The chief limitation of this study, and the linked dataset as a whole, is that the paediatric data is only presented for children who had a paediatric appointment between 2021 and 2024. This means there could be many more children with a learning disability identified in the paediatric data who are not present in the data simply because they did not engage with a paediatrician in the last 5 years. Furthermore, as explained in section 3.0, the linked dataset did not include permanent health records, which can include children with learning disabilities who do not necessarily have more detailed data about their needs captured in the timeframe of the paediatric data included in this study.

Furthermore, sample sizes are small in the ethnicity subgroups (and some other demographic characteristics), particularly when these groups are further divided by the presence or absence of an identified learning difficulty or disability. For example, in 2022, our data included only 6 black children who had a paediatric record of a learning disability. While these sample sizes are expected in Sunderland, which is not a particularly ethnically diverse LA, they should still be considered when interpreting statistically significant results.

It would also have been beneficial to have geographic data, such as the index of multiple deprivation (IMD), which was only included in the paediatric data and not linked to all children in the school census data. This would allow for analysis of any geographic clustering of learning difficulties throughout Sunderland, to compare potential findings with known literature on the 'postcode lottery' associated with SEND.

As always, it is impossible to know whether differences in learning records between groups is due to lower prevalence of learning disability in those groups, or whether there is a greater proportion of unidentified needs in these groups.

8.0 Future research

This study's objectives could not have been addressed without the rich linked dataset that was created by leveraging the strong local and regional partnerships in Sunderland. The significance of the study's findings provides evidence as to the far-reaching utility of linked data, and future research at a local and regional level should continue to take advantage of linked data as it provides a more holistic view of the health, wellbeing, and learning outcomes of children across domains than siloed records from individual services.

This particular study could have benefited from additional years of health data, or a more permanent record of each child's known health investigations, assessments, and diagnoses. The researchers were also made aware that certain health assessments may be carried out in Sunderland by organisations such as Children and Young People's Services (CYPS), whose records and data are not currently linked to the data used for this study.

While there are certainly opportunities to expand this linked data and evaluate other public health issues affecting the population of children in Sunderland, there is also a need for more granular research on some groups identified in this study. In particular, while Sunderland is a predominantly White-British area, there is some evidence that children from ethnic minority groups could currently be overlooked with regards to identification of learning difficulties.

Furthermore, the paediatric data contained numerous diagnoses and identified needs related to learning disabilities, and it would be worth exploring whether these terms are being accurately and consistently applied; this would also be useful in relation to recording SEN needs in the school census, particularly in relation to mild to moderate learning disabilities (and the use of the Moderate Learning Difficulties category). It would also be useful to explore, with the help of families, what difference having a paediatric record of learning disability (as opposed to a recorded learning difficulty in education) makes in terms of reasonable adjustments, transition planning, and access to services. Moreover, this work should investigate how inconsistent terminology impacts the support and choices that families are offered.

Finally, while there are many benefits to academic uses of linked data, it would be advantageous to consult with individual services to identify gaps in their understanding that could be rectified by data sharing, and this study could serve as a proof of concept of the types of questions that data linkage can help to answer.

The researchers are currently using the same linked dataset to investigate patterns in access to interventions across healthcare, social care, and education settings, which will use data that was not relevant to the objectives of the present study. It is preferable that the information governance processes involved in linking and sharing data across services, and with academics, will be facilitated and improved in the future, as these served as the most significant barriers to completing the present study.

Recommendations

1. Clarify the definitions of the different terms used across services and agencies to describe learning difficulties and disabilities, and the assessment pathways that underpin these.
2. Create an evidence-based terminology guide that aligns definitions across education, healthcare, and social care. Where it is not possible to standardise terminology, agencies should provide cross-sector terminology mapping tools to allow practitioners and families to better understand differences in terms, and to enable a robust approach to joining up data on this cohort.
3. Encourage agencies to describe specific learning needs and functional impacts rather than relying solely on categorical labels. This approach supports more tailored interventions and planning.
4. Schools should be asked to record all of a child's needs (similar to the approach used in other datasets), not just the perceived primary and secondary need. This will ensure a more complete picture of each child's profile and reduce the risk of important needs being obscured.
5. Mandate consistent data capture and documentation of all needs at all points of care and services across agencies to ensure that quality data are available to link and analyse to improve outcomes for children, including transparent consent processes for documentation, storage, and use of these data.
6. Invest in systems and governance frameworks that allow secure, efficient sharing of data across agencies. This includes adopting a unique identifier (such as NHS numbers or proposed Single Unique Identifier) to support linkage.
7. Establish a cross-sector expert working group, including representatives from education, healthcare, social care, and family-carers, to facilitate the development and implementation of consistent terminology and data-sharing practices.

References

ADR UK (2023)
Linked local data on children and young people.
Available at: https://www.adruk.org/our-work/browse-all-projects/linked-local-data-on-children-and-young-people/
(Accessed: 31 October 2025).

Aldridge, R.W. et al. (2019)
'Causes of death among homeless people: a population-based cross-sectional study of linked hospitalisation and mortality data in England.', Wellcome Open Research, 4, p. 49.
Available at: https://doi.org/10.12688/wellcomeopenres.15151.1.

Archibald, L.M.D. (2024)
'On the many terms for developmental language and learning impairments', Discover Education, 3(1), p. 33.
Available at: https://doi.org/10.1007/s44217-024-00112-y.

Atherton, I.M. et al. (2015)
'Barriers and Solutions to Linking and Using Health and Social Care Data in Scotland', British Journal of Social Work, 45(5), pp. 1614–1622.
Available at: https://doi.org/10.1093/bjsw/bcv047.

Bates, D. et al. (2015)
'Fitting Linear Mixed-Effects Models Using lme4', Journal of Statistical Software, 67(1), pp. 1–48.
Available at: https://doi.org/10.18637/jss.v067.i01.

Bates, D., Maechler, M. and Bolker, B. (2020)
mlmRev: Examples from Multilevel Modelling Software Review.
Available at: https://CRAN.R-project.org/package=mlmRev.

Bradford Metropolitan District Council (2020)
'Working Definitions of Learning Difficulties'.

Children Act (1989).

Chiu, M. et al. (2016)
'Describing the linkages of the immigration, refugees and citizenship Canada permanent resident data and vital statistics death registry to Ontario's administrative health database', BMC Medical Informatics and Decision Making, 16(1), p. 135.
Available at: https://doi.org/10.1186/s12911-016-0375-3.

Council for Disabled Children (2024)
Disability and additional learning needs. CDC.
Available at: https://councilfordisabledchildren.org.uk.

Department for Education (2015)
'Special educational needs and disability code of practice: 0 to 25 years'.
Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_ data/file/398815/SEND_Code_of_Practice_January_2015.pdf
(Accessed: 22 November 2021).

DfE (2021)

Department for Education (2021) Children in Need Census: Additional Guide on the Factors Identified at the End of Assessment. Technical Guidance. London: Department for Education.

DfE (2023)
Improving multi-agency information sharing, GOV.UK.
Available at: https://www.gov.uk/government/publications/improving-multi-agency-information-sharing (Accessed: 31 October 2025).

DfE (2025)
'Pupils in all schools, by type of SEN provision - 2016 to 2025'. GOV.UK.
Available at: https://explore-education-statistics.service.gov.uk/data-tables/special-educational-needs-in-england/2024-25?subjectId=27403e26-aca1-4ef2-3b3f-08dd97c89364.

Downs, J. et al. (2017)
'Linking health and education data to plan and evaluate services for children', Archives of Disease in Childhood, 102(7), pp. 599–602.
Available at: https://doi.org/10.1136/archdischild-2016-311656.

Downs, J.M. et al. (2019)
'An approach to linking education, social care and electronic health records for children and young people in South London: a linkage study of child and adolescent mental health service data', BMJ Open, 9(1), p. e024355.
Available at: https://doi.org/10.1136/bmjopen-2018-024355.

ECHILD (2025)
ECHILD: Education and Child Health Insights from Linked Data, ECHILD: Education and Child Health Insights from Linked Data.
Available at: https://www.echild.ac.uk
(Accessed: 31 October 2025).

Emerson, E. et al. (2012)
'Improving health and lives: The Learning Disabilities Public Health Observatory', Advances in Mental Health and Intellectual Disabilities, 6(1), pp. 26–32.
Available at: https://doi.org/10.1108/20441281211198835.

Emerson, E. et al. (2014)
'Estimating the number of children in England with learning disabilities and whose behaviours challenge', Kent Academic Repository [Preprint].

Emmott, E.H., Jay, M.A. and Woodman, J. (2019)
'Cohort profile: Children in Need Census (CIN) records of children referred for social care support in England', BMJ Open, 9(2), p. e023771.
Available at: https://doi.org/10.1136/bmjopen-2018-023771.

Fleming, M. et al. (2021)
'Educational and health outcomes of schoolchildren in local authority care in Scotland: A retrospective record linkage study', PLOS Medicine. Edited by M. Tomlinson, 18(11), p. e1003832.
Available at: https://doi.org/10.1371/journal.pmed.1003832.

Gabriel, T. and Börnert-Ringleb, M. (2023)
'The intersection of learning difficulties and behavior problems – a scoping review of intervention research', Frontiers in Education, 8.
Available at: https://doi.org/10.3389/feduc.2023.1268904.

Government Actuary's Department (GAD) (2021)
Mortality Insights from GAD - September 2021, GOV.UK.
Available at: https://www.gov.uk/government/publications/mortality-insights-from-gad-september-2021/mortality-insights-from-gad-september-2021
(Accessed: 31 October 2025).

Harron, K. (2022)
'Data linkage in medical research', BMJ Medicine, 1(1), p. e000087.
Available at: https://doi.org/10.1136/bmjmed-2021-000087.

Hartig, F. (2025)
DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models.
Available at: https://github.com/florianhartig/DHARMa.

HM Government (2022)
'SEND Review: Right support, right place, right time'. London: HM Government.

HM Government (2023)
SEND and alternative provision improvement plan, GOV.UK.
Available at: https://www.gov.uk/government/publications/send-and-alternative-provision-improvement-plan
(Accessed: 31 October 2025).

Holland, K. (2011)
Factsheet: Learning Disabilities. British Institute of Learning Disabilities.
Available at: https://studylib.net/doc/18790584/factsheet--learning-disabilities---british-institute-of-l...

Holman, C.D.J. et al. (2008)
'A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system', Australian Health Review, 32(4), pp. 766–777.
Available at: https://doi.org/10.1071/AH080766.

Horridge, K. (2018)
SNOMED CT Case Study. NHS Digital.

King, D. et al. (2022)
Review of existing sources of evidence about people with Learning Disability. NIHR Policy Research Unit.
Available at: https://ascru.nihr.ac.uk/output/review-of-existing-sources-of-evidence-about-people-with-learning-disability-summary/
(Accessed: 5 December 2025).

Krasniqi, P., Carr, M. and Stevens, M. (2023)
"People don't understand" - The impact of the cost-of-living crisis on children and young people with special educational needs and disabilities. The Childhood Trust.

London Office of Technology and Innovation (LOTI) (2025)
Pan-London Data Sharing Agreements Project.
Available at: https://loti.london/projects/pan-london-dsa-project/
(Accessed: 31 October 2025).

Long, R., Roberts, N. and Foster, D. (2025)
Children's Wellbeing and Schools Bill - Parliamentary Bills - UK Parliament. Available at: https://bills.parliament.uk/bills/3909
(Accessed: 31 October 2025).

Lüdecke, D. (2025)
sjPlot: Data Visualization for Statistics in Social Science.
Available at: https://CRAN.R-project.org/package=sjPlot.

MacAlister, J. (2022)
Independent review of children's social care: final report, GOV.UK.
Available at: https://www.gov.uk/government/publications/independent-review-of-childrens-social-care-final-report
(Accessed: 31 October 2025).

Martin-Denham, S. et al. (2023)
Putting together the data jigsaw: The extent of the linking of administrative datasets on children by local authorities in England. Responses to a Freedom of Information request. University of Sunderland.
Available at: https://sure.sunderland.ac.uk/id/eprint/16347/.

Mencap (2024)
How common is learning disability?, Mencap.
Available at: https://www.mencap.org.uk/learning-disability-explained/research-and-statistics/how-common-learning-disability
(Accessed: 4 December 2025).

NHS (2025)
Community Services Data Set, NHS Data Model and Dictionary.
Available at: https://archive.datadictionary.nhs.uk/DD%20Release%20November%202025/data_sets/clinical_data_sets/community_services_data_set.html
(Accessed: 18 December 2025).

NHS England (no date)
'NHS England: Find out about the Learning Disability Register'.
Available at: https://www.england.nhs.uk/long-read/find-out-about-the-learning-disability-register/
(Accessed: 5 December 2025).

NHS England Digital (2025)
Health and Care of People with Learning Disabilities, Experimental Statistics 2024 to 2025, Health and Care of People with Learning Disabilities.
Available at: https://digital.nhs.uk/data-and-information/publications/statistical/health-and-care-of-people-with-learning-disabilities/experimental-statistics-2024-to-2025
(Accessed: 5 December 2025).

NICE (2025)
Learning disabilities: What is it?, Clinical Knowledge Summaries.
Available at: https://cks.nice.org.uk/topics/learning-disabilities/background-information/definition/
(Accessed: 19 December 2025).

Norwich, B., Ylonen, A. and Gwernan-Jones, R. (2014)
'Moderate learning difficulties: searching for clarity and understanding', Research Papers in Education, 29(1), pp. 1–19.
Available at: https://doi.org/10.1080/02671522.2012.729153.

Parish, N., Bryant, B. and Swords, B. (no date)
Have we reached a 'tipping point'? Trends in spending for children and young people with SEND in England | Local Government Association. Local Government Association.
Available at: https://www.local.gov.uk/have-we-reached-tipping-point-trends-spending-children-and-young-people-send-england
(Accessed: 31 October 2025).

Parsons, S. and Platt, L. (2022)
Special educational needs and disability: a lifetime of disadvantage in the labour market? CLS working paper 2022/24. University College London.

Pinney, A. (2017a)
Looked after children data bulletin: What can we learn from data on children who are looked after under the Children Act 1989 due to their disability? Council for Disabiled Children.
Available at: https://councilfordisabledchildren.org.uk/resources/all-resources/filter/inclusion-send/understanding-needs-disabled-children-complex-needs.

Pinney, A. (2017b)
Understanding the needs of disabled children with complex needs or life-limiting conditions: What can we learn from national data? Council for Disabled Children.
Available at: https://councilfordisabledchildren.org.uk/resources/all-resources/filter/inclusion-send/understanding-needs-disabled-children-complex-needs.

Pinney, A. et al. (2023)
Putting together the data jigsaw: Linking administrative data sets on children with SEND - Learning from local area experience. University of Sunderland.
Available at: https://sure.sunderland.ac.uk/id/eprint/16347/.

Scott, N. et al. (2023)
Putting together the data jigsaw: The extent of the linking of administrative datasets on children by Integrated Care Boards (ICBs) in England. Responses to a Freedom of Information request. University of Sunderland.
Available at: https://sure.sunderland.ac.uk/id/eprint/17636/.

Sohal, K. et al. (2022)
'Connected Bradford: a Whole System Data Linkage Accelerator', Wellcome Open Research, 7, p. 26.
Available at: https://doi.org/10.12688/wellcomeopenres.17526.2.

Strand, S. and Lindorff, A. (2021)
'Ethnic Disproportionality in the Identification of High-Incidence Special Educational Needs: A National Longitudinal Study Ages 5 to 11', Exceptional Children, 87(3), pp. 344–368.
Available at: https://doi.org/10.1177/0014402921990895.

The Labour Party (2024)
'Mission 5: Breaking down barriers to opportunity'.

UCL Institute of Education (2011)
'Removing Barriers to Achievement: Session 3 – Moderate Learning Difficulties'. University College London Institute of Education.

Wilcock, R., Elliot, H. and Symons, T. (2022)
Unlocking the potential of early years data, nesta.
Available at: https://www.nesta.org.uk/blog/unlocking-the-potential-of-early-years-data/
(Accessed: 31 October 2025).

Appendices

Appendix 1. Additional detail on the intersection of learning disability terminology between the school census data and the paediatric data (2023/24 only).

Mild Intellectual Developmental Disability (IQ 50-70)

- In 2024, seven individuals had this health record. Their classification in the school census data is as follows:
  - 2 had a "Moderate learning difficulty"
  - 1 had a "Severe learning difficulty"
  - 1 had a "Specific learning difficulty"
    - Also had "learning difficulties" record in health data, but no mention of a 'specific' learning difficulty/disability in the health data
  - 3 had other SEN needs
    - Hearing impairment, Speech Language and Communication Needs (SLCN)

Moderate Intellectual Developmental Disability (IQ 35-49)

- In 2024, 125 individuals had this health record. Their classification in the school census data is as follows:
  - 35 had a "Moderate learning difficulty"
  - 27 had a "Severe learning difficulty"
  - 8 had a "Specific learning difficulty"
    - None had a 'specific' learning difficulty/disability in the health data, but one had "developmental academic disorder"
  - Many other SEN needs
    - 47 had autism
    - 20 had SLCN
    - Others include SEMH, physical disability, etc.

Severe Intellectual Developmental Disability (IQ 20-34)

- In 2024, 205 individuals had this record. Their classification in the school census data is as follows:
  - 95 had a "Severe learning difficulty"
  - 21 had a "Profound & multiple learning difficulty"
  - 4 had a "Moderate learning difficulty"
  - Many other SEN needs
  - 108 had autism
  - Others include SLCN, SEMH, physical disability etc.

Profound Intellectual Developmental Disability (IQ <20)

- In 2024, 13 individuals had this record. Their classification in the school census data is as follows:
  - 9 had a "Profound & multiple learning difficulty"
  - 2 had a "Severe learning difficulty"
  - Some had other SEN needs
    - 4 had SLCN
    - 4 had a physical disability

Developmental Academic Disorder

Crucially, individuals with this health record sometimes (N=30) have no School Census classification at all (contrary to all above diagnoses)

- In 2024, 300 individuals had this health record
  - 25 had a "Moderate learning difficulty"
  - 27 had a "Severe learning difficulty"
  - 11 had a "Profound & multiple learning difficulty"
  - 11 had a "Specific learning difficulty"
  - Many had other SEN needs
    - 125 had autism
    - 112 had SLCN

Appendix 2. Logistic mixed-effect models using paediatric record of a learning disability as an outcome measure – Omitted findings from objective 4

2021/22

Table 12. Logistic mixed-effects model predicting a learning disability identified in the paediatric data (2021/22) using demographic predictors and school as a random effect (N = 39,029)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.03	0.01	0.02 - 0.05	<0.001
Age	0.86	0.01	0.83 - 0.89	<0.001
Gender (Male)	1.76	0.19	1.42 - 2.17	<0.001
FSM Eligibility (Eligible)	2.05	0.28	1.57 - 2.67	<0.001
Asian (vs White)	1.00	0.21	0.66 - 1.51	0.989
Black (vs White)	0.56	0.24	0.24 - 1.31	0.179
Mixed (vs White)	0.81	0.23	0.46 - 1.42	0.464
Other or unknown (vs White)	0.95	0.22	0.61 - 1.48	0.809
Gender × FSM eligibility	0.96	0.16	0.70 - 1.32	0.818
Random Effects				
σ2	3.29			
τ00 School	1.29			
ICC	0.28			
N School	114			
Observations	39029			
Marginal R^2 / Conditional R^2	0.097 / 0.352			
AIC	6233.957			

Figure 7. Predicted probabilities of having a learning disability identified in the paediatric data by age.

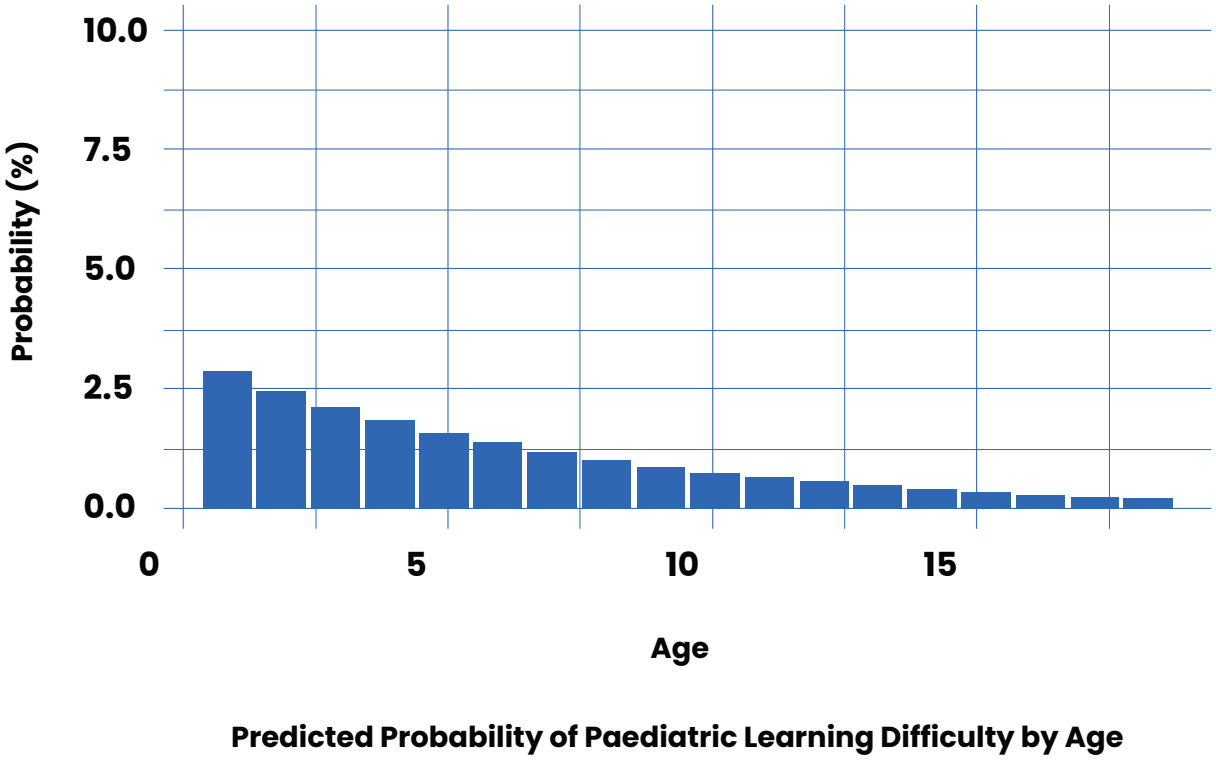


Table 13. Logistic mixed-effects model predicting a learning disability identified in the paediatric data (2022/23) using demographic predictors and school as a random effect (N = 39,108)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.03	0.01	0.02 - 0.04	<0.001
Age	0.89	0.01	0.86 - 0.91	<0.001
Gender (Male)	1.82	0.18	1.50 - 2.21	<0.001
FSM Eligibility (Eligible)	2.17	0.27	1.70 - 2.76	<0.001
Asian (vs White)	0.96	0.19	0.66 - 1.41	0.846
Black (vs White)	0.48	0.14	0.27 - 0.86	0.013
Mixed (vs White)	0.60	0.17	0.34 - 1.05	0.074
Other or unknown (vs White)	0.44	0.11	0.26 - 0.72	0.001
Gender × FSM eligibility	0.85	0.13	0.63 - 1.13	0.263
Random Effects				
σ2	3.29			
τ00 School	1.21			
ICC	0.27			
N School	115			
Observations	39108			
Marginal R^2 / Conditional R^2	0.080 / 0.327			
AIC	7294.124			

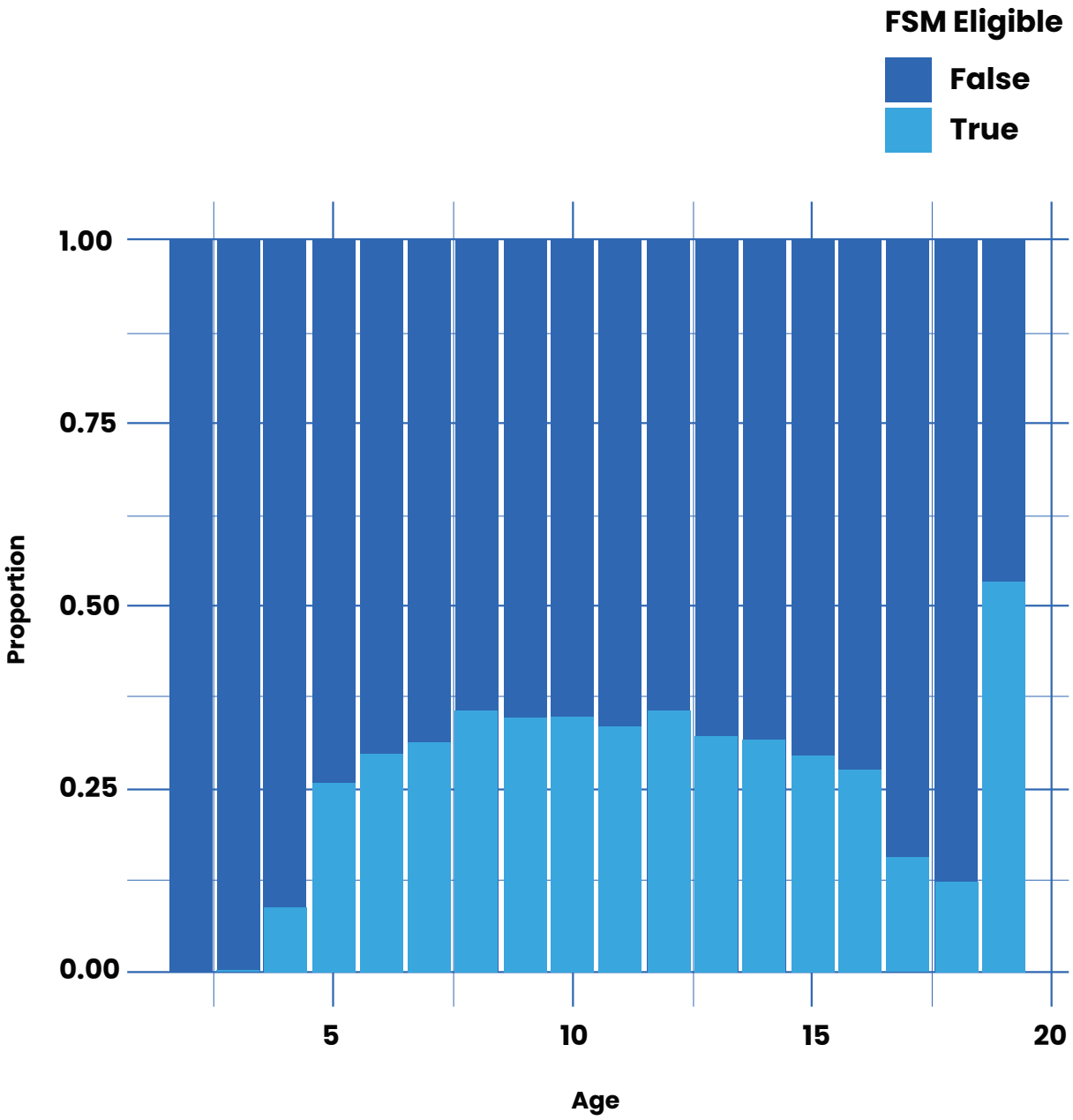
Table 14. Logistic mixed-effects model predicting a learning disability identified in the paediatric data (2023/24) using demographic predictors and school as a random effect (N = 39,088)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.03	0.01	0.02 - 0.04	<0.001
Age	0.90	0.01	0.87 - 0.92	<0.001
Gender (Male)	1.82	0.18	1.50 - 2.21	<0.001
FSM Eligibility (Eligible)	2.12	0.26	1.67 - 2.69	<0.001
Asian (vs White)	1.12	0.20	0.79 - 1.60	0.512
Black (vs White)	0.38	0.10	0.23 - 0.64	<0.001
Mixed (vs White)	0.64	0.17	0.37 - 1.09	0.099
Other or unknown (vs White)	0.64	0.14	0.41 - 0.99	0.043
Gender × FSM eligibility	0.80	0.12	0.60 - 1.06	0.126

Random Effects

σ2	3.29
τ00 School	1.24
ICC	0.27
N School	114
Observations	39088
Marginal R^2 / Conditional R^2	0.073 / 0.326
AIC	7448.361

Appendix 3. Stacked bar plot showing the relationship between age and FSM eligibility in 2022.



Appendix 4. Logistic mixed-effect models using paediatric record of a learning disability as an outcome measure – Omitted findings from objective 5

For this objective, mixed-effects general linear models were used again, however, while the demographic characteristic predictors were retained (except for ethnicity, as these small subgroups contributed to a lack of model convergence when clinical features were added), as was the random effect of school, new predictors related to clinical features were added. While this study has focused on the learning difficulties that can take the form of a child's primary or secondary SEN needs, there are other needs that may also be able to predict the presence of a paediatric learning disability outcome. As explained in section 5.1, and in Appendix 1, it appears that many children who have a learning disability according to the paediatric data do not have a learning difficulty classification in their school census records, but instead have another type of SEN need recorded. The most prominent of these needs were autism, SLCN, and SEMH. These were added to the model as predictors, as was the severity of a child's learning difficulty classification if they had one.

2021/22

Table 15. Logistic mixed-effects model predicting a learning disability identified in the paediatric data (2021/22) using demographic and clinical predictors and school as a random effect (N = 39,088)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.03	0.00	0.02 - 0.04	<0.001
Age	0.82	0.01	0.80 - 0.85	<0.001
Gender (Male)	1.27	0.11	1.07 - 1.51	0.006
FSM Eligibility (Eligible)	1.44	0.13	1.21 - 1.71	<0.001
Autism in School Census	7.35	0.78	5.98 - 9.05	<0.001
SLCN in School Census	5.10	0.49	4.23 - 6.14	<0.001
SEMH in School Census	2.69	0.36	2.07 - 3.49	<0.001
LD Severity (Moderate)	6.28	0.74	4.98 - 7.91	<0.001
LD Severity (Severe)	20.18	5.15	12.23 - 33.29	<0.001
LD Severity (Profound and Multiple)	37.59	16.78	15.67 - 90.19	<0.001
Random Effects				
σ2	3.29			
τ00 School	0.57			
ICC	0.15			
N School	114			
Observations	39029			
Marginal R^2 / Conditional R^2	0.256 / 0.365			
AIC	5408.954			

In Table 15, all demographic and clinical characteristics were significant predictors of paediatric learning disability identification. Gender and FSM eligibility remain strong predictors, however after the inclusion of clinical features, they are not as strong as in previous models (see Tables 9-14).

Having an SEN need in the school census other than a learning difficulty was a strong predictor of whether a child had a paediatric a learning disability identified. Children with autism as a school census SEN need were over 7 times more likely to have an identified learning disability in the paediatric data than those who did not. Similar effects were also observed for SLCN (over 5 times more likely) and SEHM (over 2 times more likely).

Notably, having autism as an SEN need is actually a stronger predictor of a learning disability identified in the paediatric data than having a 'Moderate learning difficulty' as an SEN need, although having this type of learning difficulty still increased the odds of an identified learning disability by over 5 times (compared to children without a recorded learning difficulty in the school census).

As expected, the stronger the severity of learning difficulty classification, the higher the odds that a learning disability was identified and documented in the paediatric data, although the strength of the effects are extreme (20 times higher odds for severe learning difficulties and 37 times higher odds for profound and multiple learning difficulties). Some caution should be observed when interpreting learning difficulty severity effects, as the 'Profound and multiple' level of severity contains a small number of children, resulting in high confidence interval ranges and large standard error estimates.

Once these clinical features were added to the model, there was a considerable increase in marginal R^2 compared to the models fitted in Tables 12, 13 and 14, even without the inclusion of ethnicity as a predictor. The demographic and clinical characteristics accounted for around 25% of the variability in a learning disability identified in the paediatric data (compared to around 9% in the demographic-only model in Tables 12-14), and once the random effect of school is taken into account, the overall conditional R^2 suggests that the fixed and random effects account for around 36% of variability in this outcome measure.

2022/23

Table 16. Logistic mixed-effects model predicting a learning disability identified in the paediatric data (2021/22) using demographic and clinical predictors and school as a random effect (N = 39,108)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.03	0.00	0.02 - 0.03	<0.001
Age	0.85	0.01	0.82 - 0.87	<0.001
Gender (Male)	1.23	0.10	1.05 - 1.44	0.010
FSM Eligibility (Eligible)	1.51	0.12	1.28 - 1.76	<0.001
Autism in School Census	7.98	0.74	6.65 - 9.59	<0.001
SLCN in School Census	5.55	0.48	4.69 - 6.57	<0.001
SEMH in School Census	2.50	0.30	1.98 - 3.16	<0.001
LD Severity (Moderate)	5.85	0.65	4.72 - 7.27	<0.001
LD Severity (Severe)	20.66	5.28	12.52 - 34.09	<0.001
LD Severity (Profound and Multiple)	57.05	24.59	24.51 - 132.77	<0.001
Random Effects				
σ2	3.29			
τ00 School	0.51			
ICC	0.14			
N School	115			
Observations	39108			
Marginal R^2 / Conditional R^2	0.245 / 0.347			
AIC	6233.460			

For the 2022/23 data, the significance, strength and direction of the gender and FSM eligibility effects are similar to the 2021/22 model. There is also a high level of similarity between the two academic years of data regarding the predictive effects of autism, SLCN and SEMH classifications on identification of a learning disability in the paediatric data. The only substantial differences between the model outputs in 2021/22 and 2022/23 is that the odds ratio of the 'Profound and multiple learning difficulties' predictor is considerably higher in the 2022/23 model. Children with this learning difficulty classification are 57 times more likely to have a learning disability identified in the paediatric data than those who had no learning difficulty, an increase from the odds ratio of around 36 in the previous model.

It is important to acknowledge that this model produced a convergence warning during fitting, primarily due to small subgroup sizes for higher levels of learning difficulty severity. The original model was retained for its clinical relevance, but findings related to severity should be interpreted with caution (see Appendix 5 for details).

2023/24

Table 17. Logistic mixed-effects model predicting a learning disability identified in the paediatric data (2023/24) using demographic and clinical predictors and school as a random effect (N = 39,088)

Predictors	Odds Ratios	std. Error	CI	p
(Intercept)	0.13	0.02	0.09 - 0.18	<0.001
Age	0.87	0.01	0.84 - 0.89	<0.001
Gender (Male)	1.18	0.09	1.01 - 1.38	0.034
FSM Eligibility (Eligible)	1.48	0.12	1.26 - 1.73	<0.001
Autism in School Census	8.34	0.74	7.01 - 9.92	<0.001
SLCN in School Census	5.12	0.43	4.34 - 6.04	<0.001
SEMH in School Census	2.16	0.26	1.71 - 2.73	<0.001
LD Severity (Moderate)	6.00	0.65	4.85 - 7.42	<0.001
LD Severity (Severe)	21.86	5.56	13.27 - 36.00	<0.001
LD Severity (Profound and Multiple)	142.36	63.54	59.35 - 341.44	<0.001
Random Effects				
σ2	3.29			
τ00 School	0.46			
ICC	0.12			
N School	114			
Observations	39088			
Marginal R^2 / Conditional R^2	0.243 / 0.335			
AIC	6333.974			

The gender, FSM eligibility, autism, SLCN and SEMH main effects in the model fitted to the 2023/24 data are consistent with the previous two academic years. The strength and direction of these effects are consistent, and there is still a strong association between having the SEN need of autism in the school census and having a learning disability identified in the paediatric data.

While fitting this model, the convergence warning referred to in the previous model (and explored in Appendix 5) was not produced, however there is a high amount of variability in some of the small subgroups. The odds ratio for the 'Profound and multiple' subgroup increased dramatically to over 140, compared to children without a learning difficulty classification. Both the standard error and confidence intervals have more than doubled, and while the direction of the effect is likely reliable, the strength of the effect is extreme, and should be regarded with appropriate caution. Despite this, the marginal R2 and conditional R2 remain similar to previous years, reinforcing the stability of the model.

Appendix 5. Notes on model convergence in Table 16

The final models for Tables produced a convergence warning (max|grad| = 0.0085; tolerance = 0.002), meaning that the optimization algorithm did not fully meet the strict convergence criteria. Removing the learning difficulty severity terms resolved this convergence warning, however due to their clinical importance to objective 5, the detail they provide, and their contribution to the conditional R2, the original model was retained. Diagnostic checks (DHARMa residual plots) suggested that the model fit was acceptable, and parameter estimates were stable across specifications. Results were retained and interpreted with caution, and the sparsity of data in certain learning difficulty severity subgroups was acknowledged.

University of
Sunderland