

The Education & Child Health Insights from Linked Data (ECHILD) Database

Key messages from stakeholders

August 2021

1. Introduction

Collaboration across stakeholder groups is essential to ensure that research using public sector data is truly in the interests of those it hopes to benefit. Stakeholders include the intended users of research findings (for example, policymakers, service providers and public advocacy groups). All these groups have an interest in knowing whether the focus and methods of research using public sector data are ethical, robust and useful, and that any potential negative consequences have been considered and mitigated.

On 29 April 2021, a range of stakeholders came together to discuss the Education and Child Health Insights from Linked Data (ECHILD) Database, a new linked data resource (see Box 1) for research in the public interest. The purpose of the event was to provide an opportunity for stakeholders to ask questions about and share their views on the ECHILD Database, and to suggest priority areas for its use in future research for the public benefit. The event consisted of short talks from the event panel (representatives of the ECHILD project partners) and a facilitated discussion with stakeholders. Stakeholders from government, the third sector and children's services with expertise in health, education and social care were invited. A full list of attendees is given in Appendix 1. This report covers the key messages arising from discussions at the stakeholder event.

Box 1: Overview of the ECHILD Database¹

The ECHILD Database brings together information about health, education and social care for all children in England for the first time. It includes de-identified, linked records for around 14 million children. The ECHILD Database links two existing data sources:

- **The National Pupil Database (NPD)** – which includes information about students at schools and colleges in England and about contact with children's social care services.
- **Hospital Episode Statistics (HES)** – which includes information on hospital admissions, A&E attendances and outpatient appointments in NHS hospitals in England.

The de-identified ECHILD Database will soon be made available to external researchers via the Office for National Statistics (ONS) [Secure Research Service \(SRS\)](#), to facilitate research to improve policymaking for children's health, education and wellbeing across a range of health and social science disciplines. Researchers will need to be [approved](#) and submit a successful application to access the data.

2. The research value of the ECHILD Database

There was much discussion about the unique aspects of the ECHILD Database as a research resource at the stakeholder event. The key strengths were seen to be its ability to cut across domains that are deeply interrelated and its coverage of the whole population. By bringing together information from across the domains of health, education and social care, the ECHILD Database allows researchers to assess, at a population level, how children's needs are being met by services in England and how service provision and outcomes vary across the country. This type of research will provide the evidence needed for policymakers and service providers to better understand how to support all children to thrive.

Dorian Kennedy (Department of Health and Social Care, a partner in the ECHILD project) emphasised the importance of different parts of government not working in isolation, and the potential of the ECHILD Database for *"breaking down those traditional silos we work within so that in developing policy we look at not just the impact of an intervention in one silo, but to the child as a whole, and their family as a whole, and I think that's absolutely vital."* Michael Chapman (NHS Digital) echoed these comments, stressing that education and health are in inextricably linked. He commented, *"I feel we have a duty to use that data in a safe, transparent, ethical fashion to improve health and education... to make policy decisions that are based on data and based on large, credible datasets rather than gut feeling or instinct or small datasets. So, this is an absolutely enormous opportunity."*

"The case for this type of work has never been stronger."

Gary Connell, Department for Education

Gary Connell (Department for Education) agreed, noting that, as a result of the Covid-19 pandemic, *"the case for this type of work has never been stronger."* He explained that the ECHILD Database will enable government departments to *"improve our evidence base so that we are able to design and make policies to improve outcomes for all children, based on data rather than on our expectation"*.

The ECHILD Database includes linked data for around 14 million children, providing a far greater sample size than research studies that collect data, such as surveys. This facilitates research to benefit groups of children with less common conditions or experiences. Christine Farquharson, ECHILD Co-Investigator from the Institute for Fiscal Studies, stressed that understanding the intersections between health and education is *"something that's been done very well with survey datasets and with smaller sources of data, but with this administrative data we'll be able to be much more precise about specific kinds of conditions that are perhaps a little bit too rare to pick up in survey data."*

Similarly, Becky Allen of Teacher Tapp described how research using the ECHILD Database could benefit schools, teachers and pupils by providing evidence about how to best support children with conditions or needs that are atypical.

The ECHILD Database: at a glance



The ECHILD Database is **DE-IDENTIFIED**



It does not contain:

- ✗ Name
- ✗ Address
- ✗ Postcode
- ✗ Date of birth
- ✗ NHS number
- ✗ Pupil number

Strengths for research



Linked data for

14.7 million pupils

Safe-guarded at every step

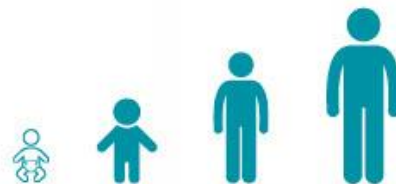


- Data
- Access
- Users
- Projects
- Outputs

Can help answer important questions

How does children's health affect their education?

Which groups of children need more support in school?



Information from birth to age 24

For more information see:
www.ucl.ac.uk/child-health/echild

3. The importance of data security

Stakeholders raised the importance of data security in research using linked data. Data security has been a central consideration when creating the ECHILD Database and will continue to be crucial when making the data available for wider use by accredited researchers. As noted by Dorian Kennedy (Department of Health and Social Care), *“the first thing we want with ECHILD is the safe and appropriate use of data.”*

The linked ECHILD Database will be made available for wider research via the Office for National Statistics (ONS) Secure Research Service (SRS), a secure research environment. ONS representative Bill South described how the ECHILD Database is securely held within the SRS in line with [the ‘Five Safes’ model](#) (Figure 1).

“The first thing we want with ECHILD is the safe and appropriate use of data.”

Dorian Kennedy, Department of Health and Social Care

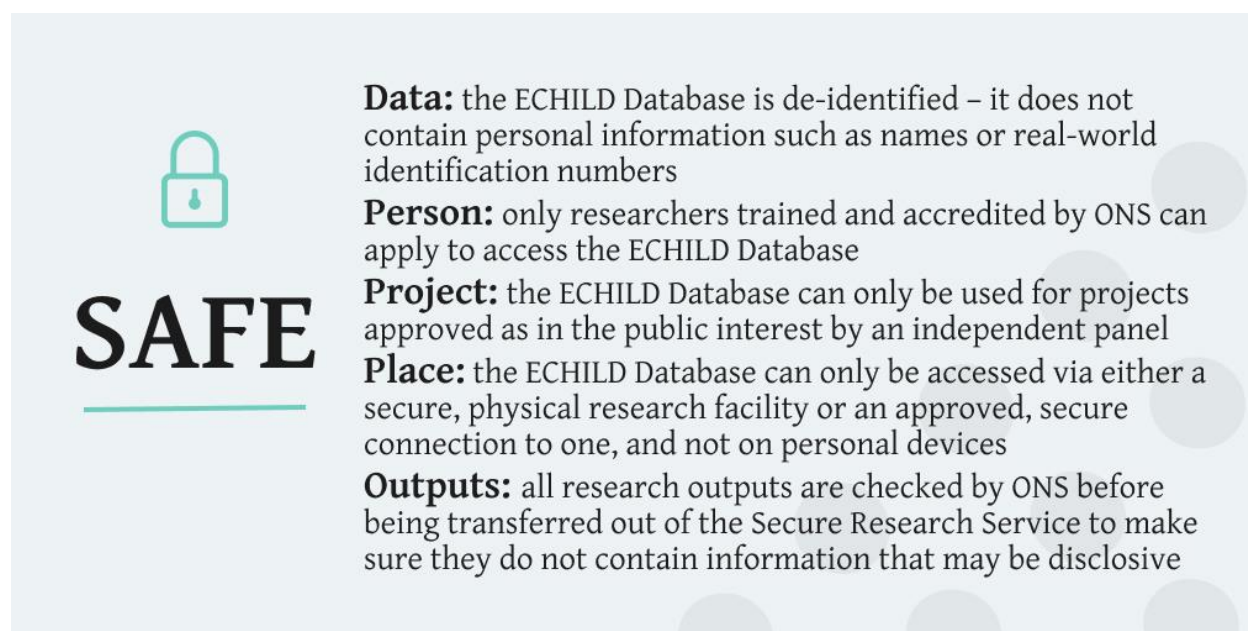


Figure 1: The ‘Five Safes’

Importantly, the ECHILD Database will not be made available for operational use; for example, to make decisions about an individual child. Dr Emma Gordon, Director of ADR UK (Administrative Data Research UK), emphasised that the de-identified ECHILD Database is a research database, only to be made available to researchers exploring patterns and trends in children’s health and education experiences on a population-wide level (for example, looking at differences between groups of children). This ensures the data is only used for purposes in the public benefit, and that the privacy of individuals is protected. Emma also explained that *“the legislation that allows us to use this data for research specifically prohibits the re-identification of people, and you’d be breaking the law if you attempted to do that.”* Michael Chapman of NHS Digital added, *“The research teams*

are not interested in individuals; they're interested in groups of children, and so we designed the system to meet that need whilst minimising disclosure."

Stakeholders also questioned whether commercial organisations would be able to use the ECHILD Database. Emma Gordon (ADR UK) explained that research projects using de-identified administrative data, such as the ECHILD Database, must be approved by [UK Statistics Authority's Research Accreditation Panel](#) (UKSA RAP), and that all projects, whether proposed by accredited researchers from academic, third sector or commercial organisations, must demonstrate a public benefit. *"The big test is that the methodology has to be good; the researchers have to know what they're doing and evidence that; and every single project has to pass the public good test."*

Before it can be approved by the UKSA RAP, data owners must also be satisfied that any proposed research will have a public benefit. Any proposed research using the ECHILD Database would need to show that it promotes children's education and wellbeing, or benefits the provision of health and social care services. Michael Chapman (NHS Digital) added, *"The legislation under which we work is very clear that the data we collect and that we would make available can only be used for the benefit of health and social care. But it doesn't speak to who is creating that benefit... When the conditions are right, we also do release data to commercial organisations, but the test there is: is the work going to benefit health and social care?"*

Examples of public benefit from research using de-identified health and social care data were gathered from members of the public in a recent series of workshops held as part of the [Putting Good into Practice](#) project. They included, for example: understanding population-level health and care needs; finding cures, new treatments and therapies to improve health outcomes; planning for the future of health and social care; and evaluating health and care provision and making long-term improvements.¹

¹ The 'Putting Good into Practice' project is co-funded by the National Data Guardian for Health and Social Care, Understanding Patient Data and UK Research and Innovation's Sciencewise programme. [Read the full public dialogue report.](#)

4. The importance of public engagement and involvement

Central to discussions throughout the stakeholder event was the importance of embedding children and young people's voices in the way the ECHILD Database is made available and used by researchers. Engaging with children and young people around the use of their data for research is essential to ensure data is used ethically and appropriately, and with full consideration of the needs and interests of those the research hopes to benefit.

Stakeholders discussed the importance of engaging with intended users of the research (such as school leaders and teachers) to ensure that research findings are useful to them and they can understand and apply them to the immediate benefit of children and young people. The importance of engagement between stakeholders when developing research questions to ensure findings are based upon shared needs and interests and are therefore useful to all was also

“It's about transparency, and it's about ensuring that we involve the people that we're working for at every stage.”

Michael Chapman, NHS Digital

highlighted. Stakeholders described a diverse range of questions of interest that could be explored in future research using the ECHILD Database (Appendix 2).

Jennifer Preston of Generation R (a national network of Young People's Advisory Groups) stressed that it's essential to establish a process for

involvement of children and young people in shaping research using the ECHILD Database, *“especially those who don't normally get a voice.”* Michael Chapman of NHS Digital went on to explain that, to tackle any underlying concerns around the use of the ECHILD Database, it's essential to ensure that *“the people whose data we're using – the children and young adults – are aware, involved and willing to support... It's about transparency, and it's about ensuring that we involve the people that we're working for at every stage.”*

Engagement has been central to the work already carried out as part of the ECHILD project, which has involved a range of patient, pupil, parent and other groups to understand different views on the use of linked data for research, and get valuable feedback on plans for research using the ECHILD Database. As of July 2021, the ECHILD project has had seven engagement events involving 87 people.² The ADR England Children & Young People Representative Panel will enable further users of the ECHILD Database to ensure the needs and interests of children and young people are properly considered in their work. The Panel is made up of representatives from a variety of third sector organisations, voluntary and community groups and professional practices – all working directly with or on behalf of children and young people. The Panel's input and advice will help researchers frame their work around the most pressing issues faced by children and young people, as well as to help facilitate direct engagement with children and young people in the research.

² Further information about the engagement activities of the ECHILD project can be found on the [ECHILD website](#).

5. The future expansion of the ECHILD Database

Emma Gordon highlighted that, “as we move out of the critical and into the recovery phase of the pandemic, the need for insights based on health data linked to education and other forms of administrative data will only grow.” Stakeholders expressed an interest in expanding the range of datasets included in the ECHILD Database to broaden the range of research questions it could address. The main datasets of interest for future expansion of the ECHILD Database were:

Primary care data

Primary care is often the first point of contact with health services, and data from this sector – such as from GP practices – would add even greater depth to the ECHILD Database. It would, for example, enable a deeper understanding of the impact of reduced primary care services on vulnerable children’s lives during the Covid-19 pandemic. Stakeholders highlighted that not all children with ill health go to hospital, and primary care data would enable researchers to better understand the healthcare interactions of children with conditions that are often managed in the community, such as autism.

Community Services Data Set (CSDS)

Discussion also arose around the importance of the [Community Services Data Set \(CSDS\)](#), and how it could enrich the ECHILD Database by providing information about children who have regular contact with health services commissioned by, but not provided directly by, the NHS. This includes services provided in settings such as health centres, Sure Start centres, day care facilities, schools or community centres, mobile facilities, or a patient's own home.

Longitudinal Education Outcomes (LEO)

The [Longitudinal Education Outcomes \(LEO\)](#) dataset brings together information from the Department for Education with employment, benefits and earnings information from the Department for Work and Pensions (DWP) and Her Majesty’s Revenue and Customs (HMRC).³ Stakeholders suggested adding LEO to the ECHILD Database would enable exploration of health and education on other, longer-term outcomes in adulthood.

Data from local authorities

It was noted that Local Authority data, particularly from children’s social care services, could be useful for understanding interactions not captured by national data. For example, Susan Cooke (NSPCC) explained the importance of having social care data for the first five years, because “children's early experiences can be key determinants of future mental health, physical health and other outcomes. From [NSPCC's] perspective, we know under-fives are particularly vulnerable to abuse and neglect; they're over-represented in the care system and in serious case reviews.” Because of the way national data are collected, it is only possible to link health and education data to social care data for children in contact with children’s social care services during their school-going years. The inclusion of Local Authority data could enable more precise insights into early care interactions which cannot be fully explored in the current ECHILD Database.

³ [Official Statistics: Graduate outcomes for all subjects by university \(Gov.uk\)](#)

Acknowledgements

The ECHILD project is led by [University College London](#) in collaboration with the [London School of Hygiene & Tropical Medicine](#) and the [Institute for Fiscal Studies](#), in partnership with [NHS Digital](#) and the [Department for Education](#), working with the [Office for National Statistics \(ONS\)](#). It is funded by [ADR UK \(Administrative Data Research UK\)](#) and the National Institute for Health Research (NIHR) Policy Research Programme, and supported by the NIHR Great Ormond Street Hospital Biomedical Research Centre and Health Data Research UK (HDRUK).

Authors

Elizabeth Waind, ADR UK Strategic Hub, elizabeth.waind@ukri.org

Louise Mc Grath-Lone, UCL Institute of Health Informatics, l.mcgrath-lone@ucl.ac.uk

Visit the [ECHILD website](#)

Visit the [ADR UK website](#)



[@ucl echild](#)
[@ADR UK](#)



Economic
and Social
Research Council



Appendix 1: List of attendees, ordered alphabetically by organisation⁴

Chair: [Claire Bolderson](#), News Journalist and Documentary Maker

<i>ECHILD project partners</i>	Representatives
Administrative Data Research UK	Emma Gordon, Director
Department of Health and Social Care	Dorian Kennedy, Deputy Director for Children, Families and Communities
Department for Education	Gary Connell, Head of Platform Operations and National Pupil Database (NPD)
Institute of Fiscal Studies	Christine Farquharson, ECHILD Co-Investigator (Co-I)
NHS Digital	Michael Chapman, Director of Research and Clinical Trials
Office for National Statistics	Bill South, Acting Deputy Director, Research Services and Data Access division
University College London	Ruth Gilbert, ECHILD Principal Investigator & Katie Harron, ECHILD Co-I
<i>External stakeholders</i>	Representatives
Children's Commissioner for England	Haroon Chowdry, Director of Evidence & Emma Nelson, Head of Statistics and Data Collection
Contact	Amanda Elliott, Strategic Health Lead
Council for Disabled Children	Amanda Allard, Assistant Director: Health
Down's Syndrome Association	Carol Boys, CEO
GenerationR Alliance	Jennifer Preston, PPI & Engagement Priority Lead
MENCAP	Cath Lunt, Research & Evaluation Manager, Disabled Children's Partnership
NASEN (National Association of Special Educational Needs)	Adam Boddison, Chief Executive & Chair of Whole School SEND
NSPCC	Susan Cooke, Head of Research & Evidence
Parentkind (formally PTA UK)	Kirsty Yates, Research Officer
Redthread	John Poyton OBE, CEO
Riverside School , Special Educational Needs school	Martin Doyle, Headteacher
SCOPE	Hannah Dobbin, Policy Manager, Children & Young People
Specialist Children & Young People's Services (SCYPS)	Jill Ellis, Consultant Community Paediatrician
Teacher Tapp	Becky Allen, Chief Analyst
useMYdata	Chris Carrigan, Expert Data User

⁴ Invitations were also extended to other organisations who did not respond or could not attend the event.

Appendix 2: Areas of interest for future research using the ECHILD Database identified by stakeholders

School attendance and health

- Are children who are often absent from school sicker than those who aren't?
- What are the characteristics, health and otherwise, of vulnerable children who regularly don't attend school?

Special educational needs (SEN) and health

- To what extent is demand for Education, Health and Care Plans (EHCPs) accounted for by the prevalence of various medical conditions or events, for example, premature birth?
- What trends in medical diagnosis can the ECHILD Database identify in relation to specific needs, such as Social, Emotional and Mental Health, or Autistic Spectrum Disorder?
- What is the variation in health outcomes for pupils receiving special educational needs (SEN) support and those in Alternative Provision according to types of educational provision and settings? How does this map to type of need and background characteristics?
- What is a common pathway for children with an EHCP and Attention Deficit and Hyperactivity Disorder (or other specified medical condition) by local authority?

Vulnerability and disadvantage

- Can exclusion from school be used as a valid indicator of a child being vulnerable in other ways?
- Do the disadvantages in health and education that appear in the poorest areas continue when families move away from such areas of deprivation? For example, does asthma and educational attainment improve if family housing conditions improve?
- Do the early health disadvantages that appear in areas of deprivation continue to show themselves in school achievements at ages 7, 12 and in GCSEs, regardless of any major health conditions?

Knife crime

- What are the characteristics of young people who have been admitted into hospital with knife crime associated injuries?