

Education and Child Health Insights from Linked Data



Katie Harron

k.harron@ucl.ac.uk

The ECHILD Database

National Pupil Database

Hospital Episode Statistics



The ECHILD Database is
DE-IDENTIFIED

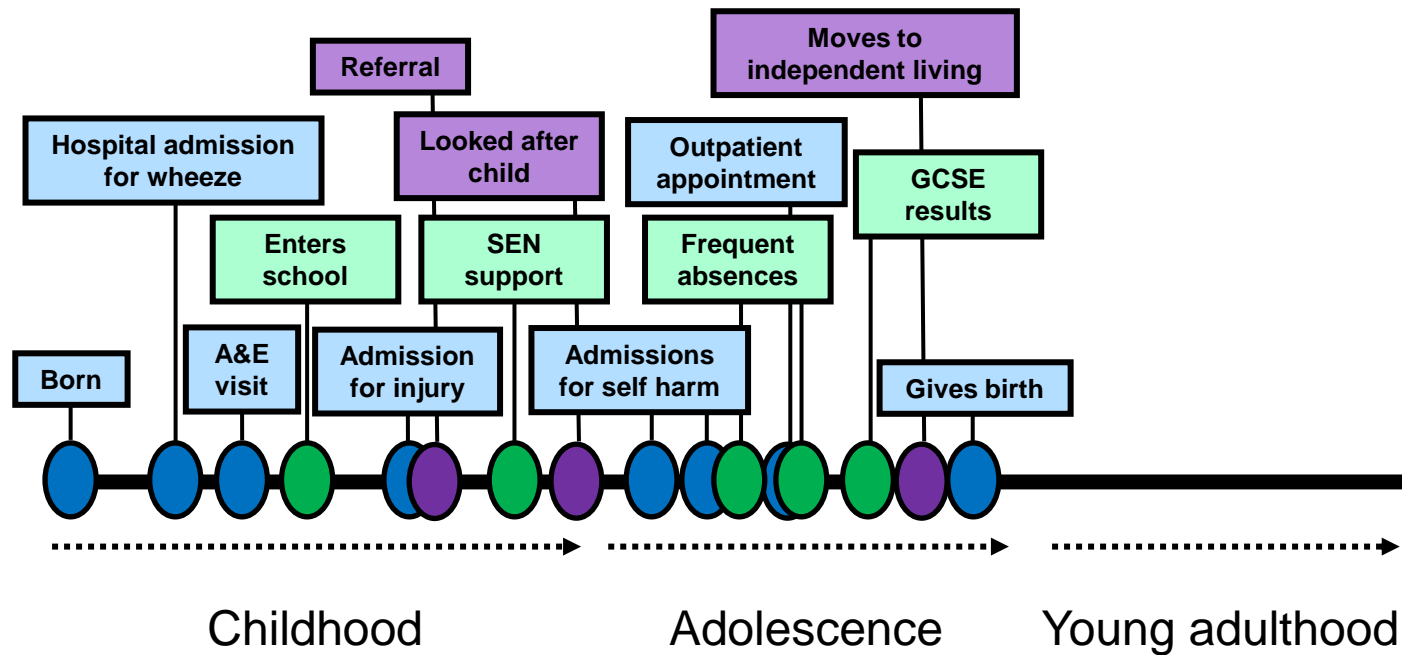


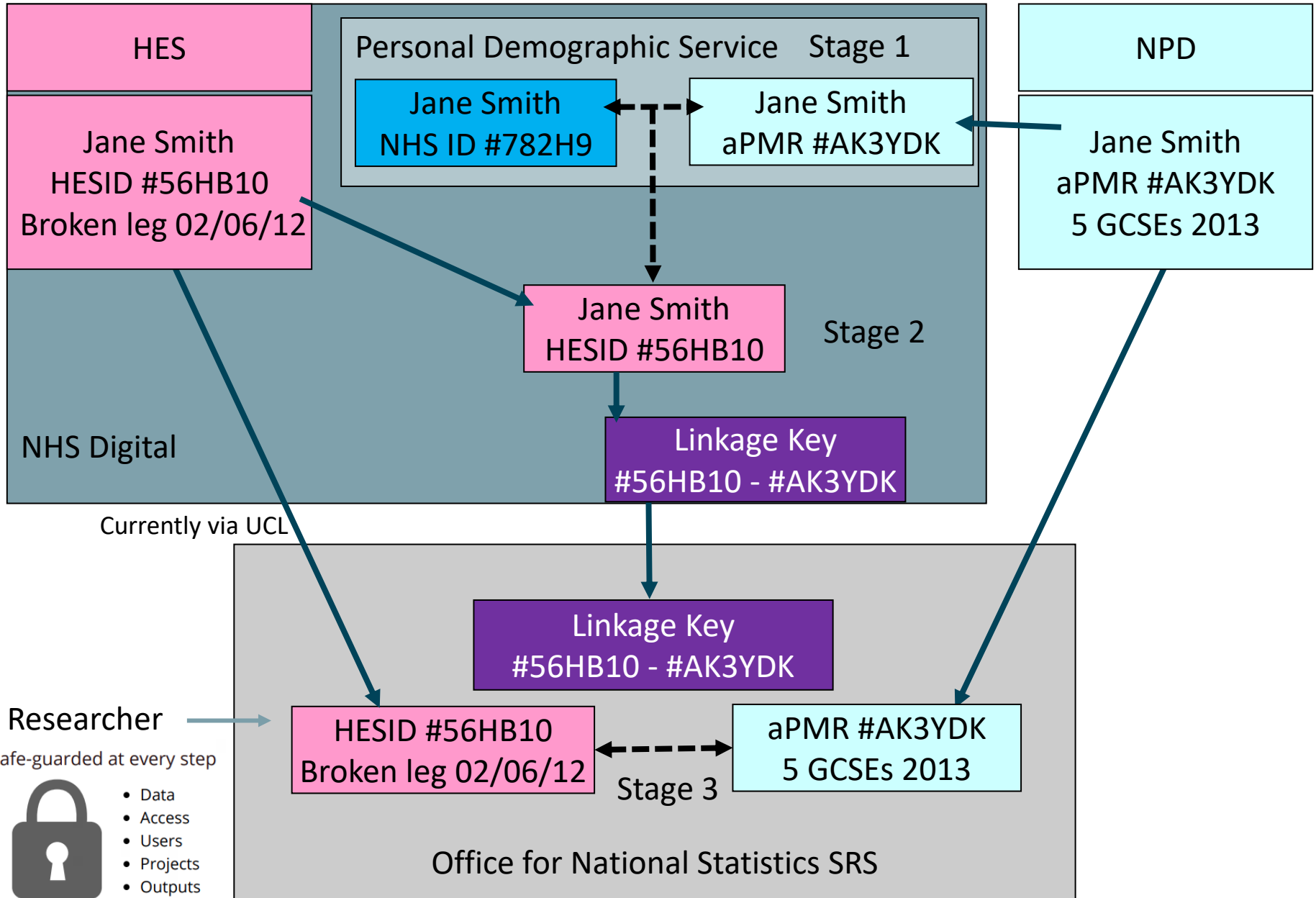
Linked data for

**14.7
million
pupils**



Information from birth to age 24





Linkage process

Stage 1 (Link NPD to Personal Demographic Service)

- Full names, postcode, DoB, sex
- 8 steps

Stage 2 (Link PDS to Hospital Episode Statistics)

- NHS no, postcode, DoB, sex
- 7 steps

Stage 3 (Merge Hospital Episode Statistics to NPD)

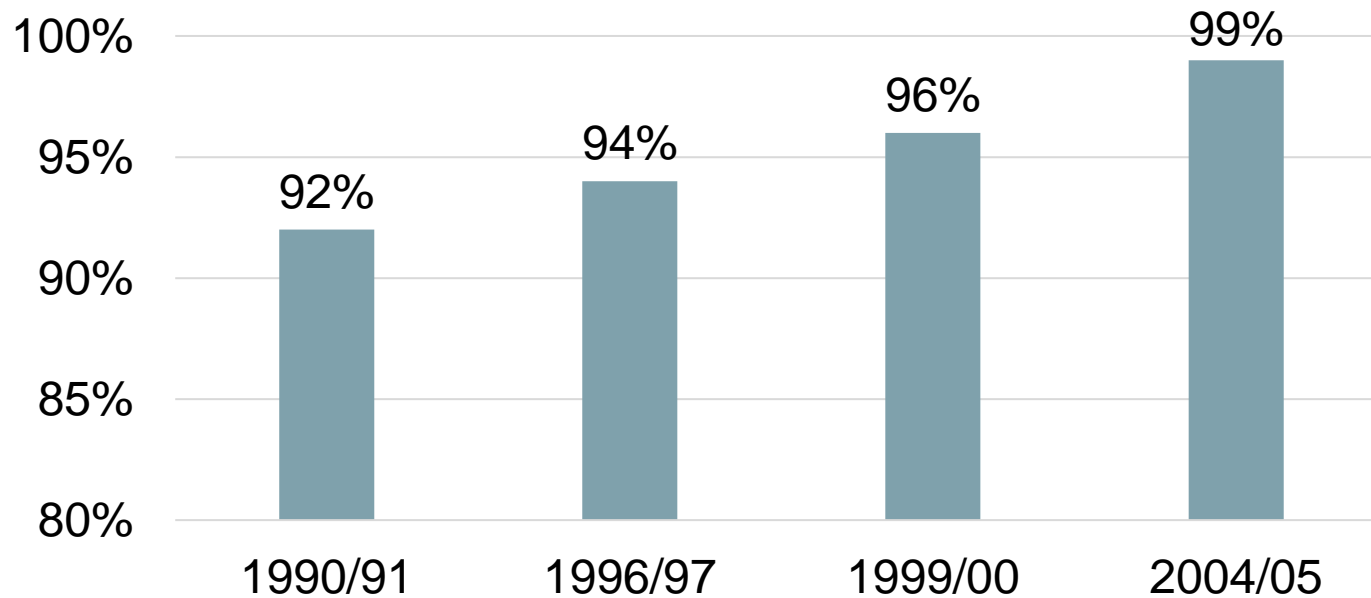
- Pseudonymised HESID-aPMR

Linkage evaluation

1. Development and evaluation of linkage algorithm
 - 4 x 1-year cohorts for births 1990/1, 1996/7, 1999/00, 2004/5
 - Linked on Spring census only

2. Application of algorithm to ALL years (births 1995 onwards) and using **all** NPD modules

Overall match % NPD-HES



Later linkage steps captured more ethnic minorities
and more children living in deprived areas

Stage 1 (NPD → Personal Demographic Service)

- >90% of NPD records link to PDS at step 1
- BUT Steps 2-8 are important:
 - 17% of records linked in step 1 are non-white ethnicity, vs 38% in steps 2-8

Step	First name	Surname	Date of birth	Gender	Postcode
1	Exact	Exact	Exact	Exact	Exact

Stage 2 (PDS to Hospital Episode Statistics)

- **98-99%** of PDS records link to HES at steps 1 or 2
- BUT steps 3-7 are important:
17% of records linked in steps 1-2 are non- white ethnicity vs 41% in steps 3-7

Step	NHS Number	Date of birth	Gender	Postcode
1	Exact	Exact	Exact	Exact
2	Exact	Exact	Exact	

Linking ALL years (births 1995 onwards) using all NPD modules

Can we link more children if we add identifiers from additional NPD modules to our linkage spine?

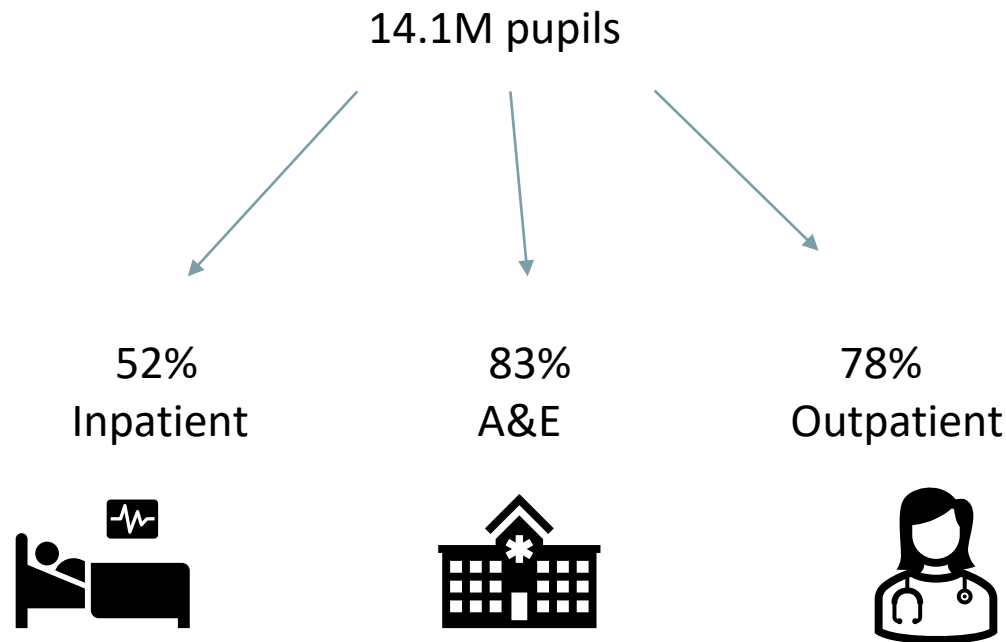
Number of unique pupils in each NPD module – (14,721,934 total unique pupils)



Linking ALL years (births 1995 onwards) and using all NPD modules

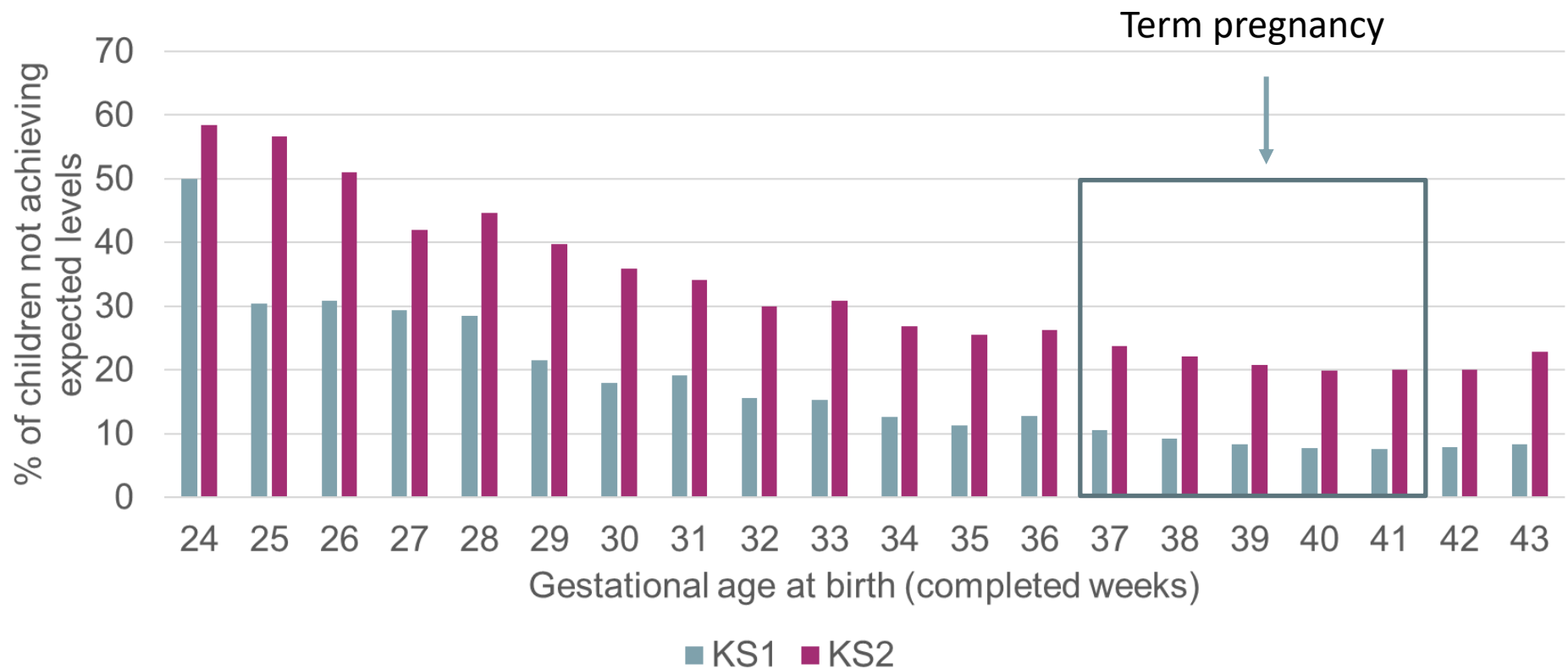


Using only identifiers from Spring census, 84% of NPD records link to HES
 Using identifiers from all modules increases this to 96%



School assessments and gestational age

N=452,013 pupils born in 2004/5 and linked to a HES birth record



Summary

- Linkage rates between NPD-HES are high: 92.0%-98.5%
- **Linkage algorithms:**
 - Non-white ethnic groups and more deprived were more likely to be missed
 - Solution: use a multi-step approach at both linkage stages
- **Linkage bias:**
 - Poorer linkage rates for non-white and disadvantaged pupils could underestimate health problems in certain groups
 - Solution: create adjustments/weights– treating unlinked groups as missing

Next steps

1. Evaluate linkage for all modules of NPD and by year
2. Ongoing evaluation needed as systems change
3. Generate meta-data for wider research use
4. De-identified linkage spine facilitates additional linkages to enhance ECHILD without identifiers

ich.echild@ucl.ac.uk

[@ucl_echild](https://twitter.com/ucl_echild)



Government data providers:

- NHS Digital: Garry Coleman and team
- Department for Education: Gary Connell and team
- ONS SRS team

Academic team: Ruth Gilbert (PI), Ruth Blackburn, Louise Mc Grath-Lone, David Etoori, Katie Harron, Nicolás Libuy, Erin Walker, Matthew Lilliman, Max Verfuerden, Pia Hardelid, Russell Viner, Jenny Woodman, Linda Wijlaars, Bianca De Stavola, Matthew Jay (all UCL), & Christine Farquharson (IFS) Imran Tahir (IFS), Jan van der Meulen (LSHTM) and Min Hae Park (LSHTM).

This study/project is funded by the National Institute for Health Research (NIHR) Policy Research Programme. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.



Data resource profiles

ECHILD:

https://www.ucl.ac.uk/child-health/sites/child-health/files/echild_user_guide_v1.1.0.pdf

Health:

HES - <https://pubmed.ncbi.nlm.nih.gov/28338941/>

Education:

NPD - <https://ijpds.org/article/view/1101>

Social Care:

CLA - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5005948/>

CIIN - <https://bmjopen.bmj.com/content/9/2/e023771>

Family Justice:

Cafcass - <https://ijpds.org/article/view/1159>

CFJDS: <https://www.ucl.ac.uk/child-health/sites/child-health/files/research-ppp-legal-epi-jay-et-al-moj-datashare-full.pdf>

FJO data scoping: <http://wp.lancs.ac.uk/observatory-scoping-study/files/2017/10/FJO-NATIONAL-DATA-SCOPING-FINAL.pdf>