

Matching the NPD to Longitudinal Studies

Clare Baker (G7 Statistician)

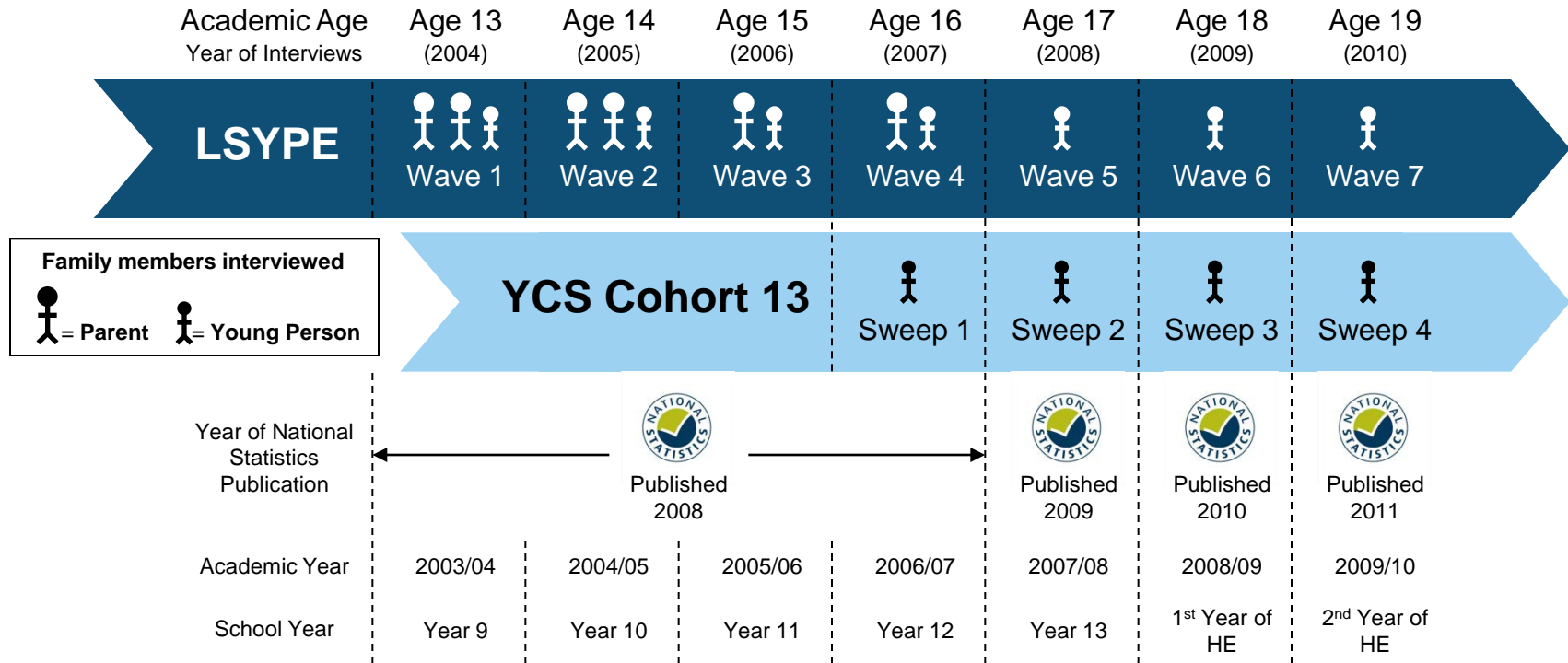
Helen Wood (Senior Researcher)

Alicia Heptinstall (Assistant Statistician)

**Infrastructure, Funding and Longitudinal Analysis Division
Department for Education**

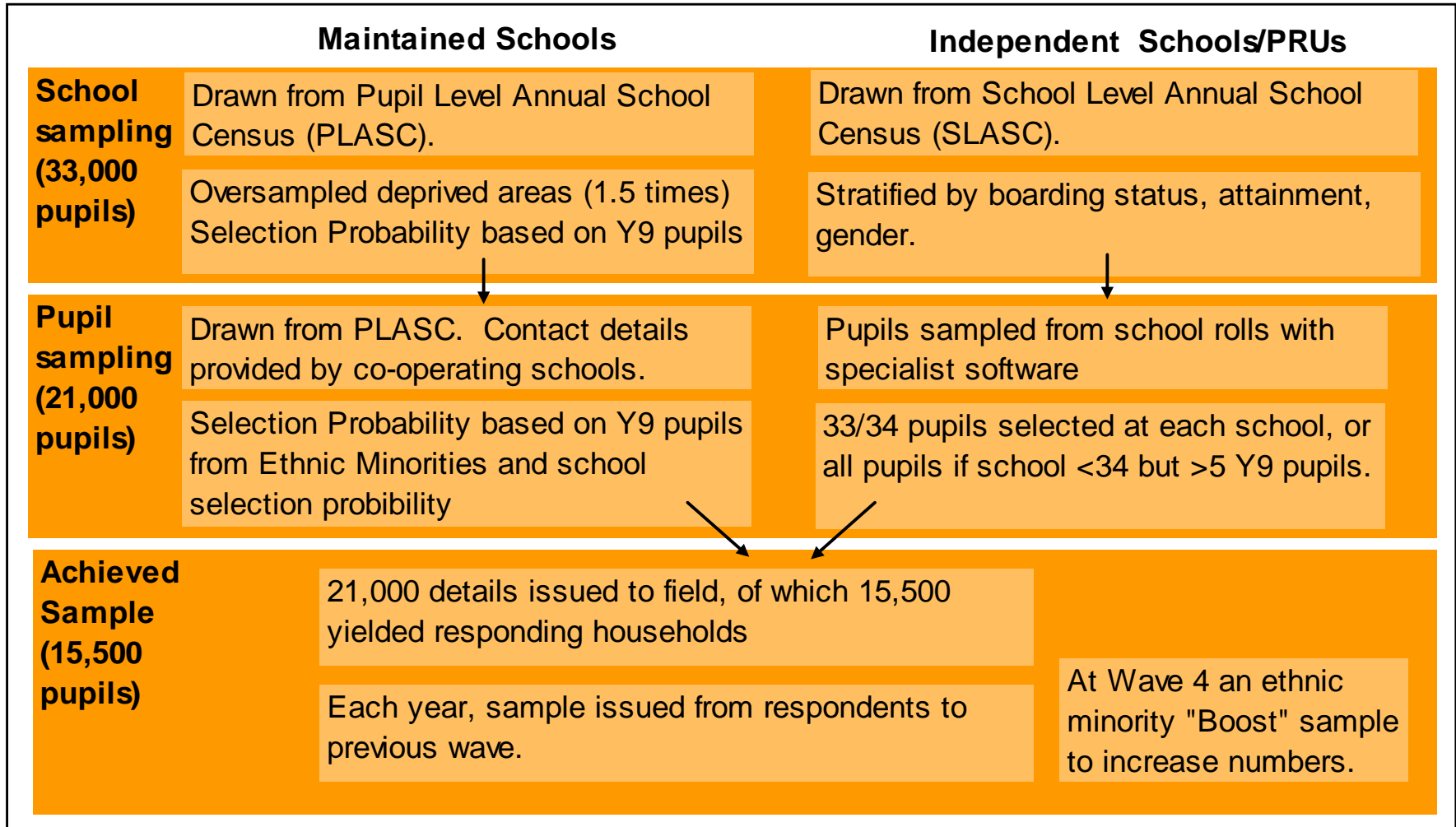
team.longitudinal@education.gsi.gov.uk

Longitudinal Study of Young People in England (LSYPE) and Youth Cohort Study (YCS)



- Initial sample size: LSYPE - 15,500 young people (YCS Cohort 13 - 7,500)
- Average 87% response rate Waves 1-7
- Mixed mode interviews: internet, telephone and face to face

LSYPE Sample design



- This sample design meant that for LSYPE we were able to directly link to NPD (i.e. without the need for 'fuzzy matching')

LSYPE content

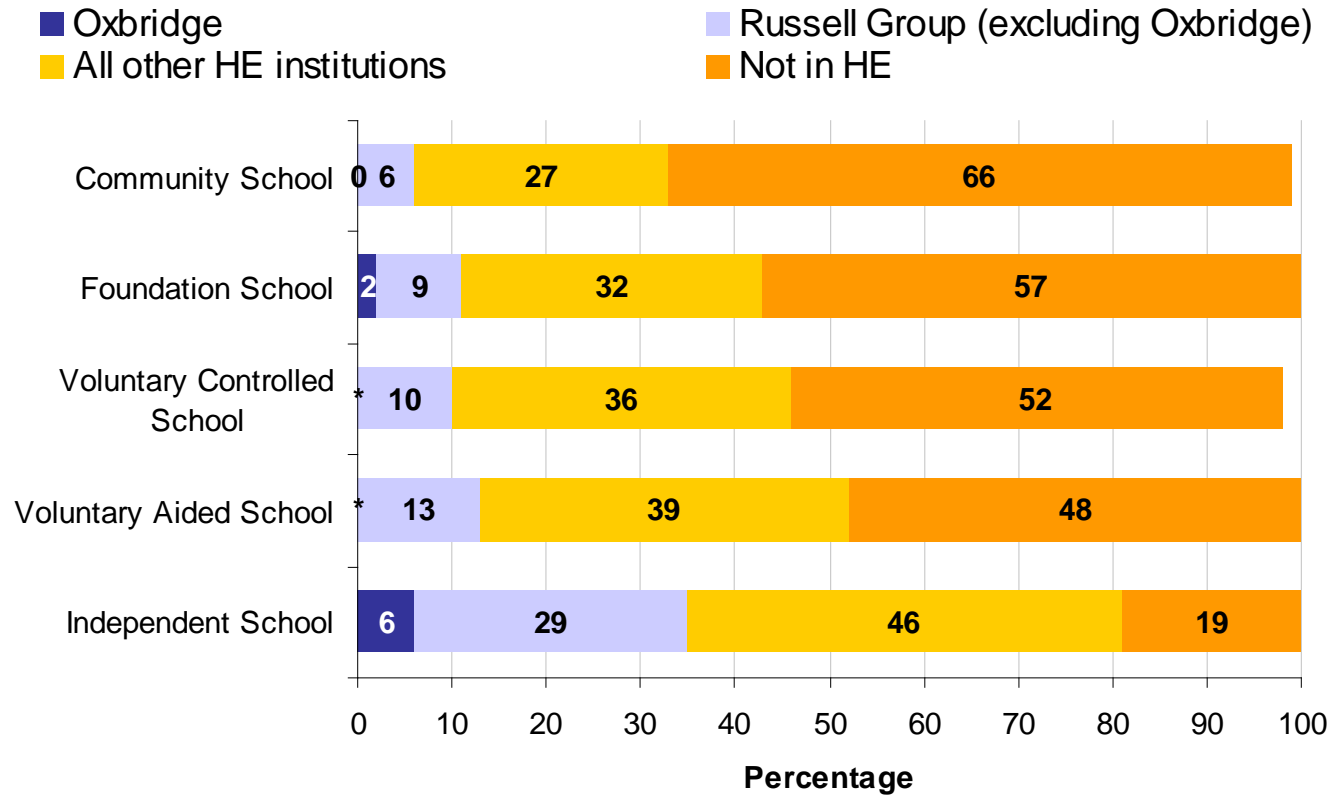
“Main” Parent	“Second” Parent	Young Person
<ul style="list-style-type: none">▪ Employment Histories including income▪ Relationship histories (household composition)▪ Demographics - ethnicity, gender, disability▪ Aspirations for the young person▪ Attitudes to school and education▪ Contact with services▪ Family Cohesion		<ul style="list-style-type: none">▪ Educational experiences including bullying and truancy▪ Non-school activities: positive & risky▪ Health, disability & SEN▪ Caring responsibilities (inc. children)▪ Relationship with parents▪ Aspirations▪ Relationships and sexual activity▪ Further Education▪ Higher Education choices & debt▪ Employment and benefits▪ Attitudes & experiences re local community

NPD data which can be matched to LSYPE

- **School level data:** Admissions policy; Institution type; Specialist school type; Gender of entry; Total number of pupils; Ethnic composition; Special Educational Need (SEN) composition; % eligible for Free School Meals (FSM); % whose first language is not English; Denomination; Sixth form flag; School historical Key Stage 2 prior attainment; School historical attainment at Key Stage 3 and 4; School absence rate.
- **Geographic markers:** Local Area District (LAD); Government Office Region (GOR); Lower Layer Super Output Area (LSOA); Urban/rural indicator; Income Deprivation Affecting Children Index (IDACI); Index of Multiple Deprivation (IMD); Local Authority (LA); Local Learning and Skills Council Area (LLSC).
- **Pupil level achievement data:** Exams taken at Key Stage 2 & 3 including marks; Level achieved in English/Science/Mathematics at Key Stage 2 & 3; Number of entries in different qualification types at Key Stage 4; Number of qualifications achieved at each grade at Key Stage 4; Grade achieved in a variety of subjects at Key Stage 4; Achieved Level 1/2 at Key Stage 4.

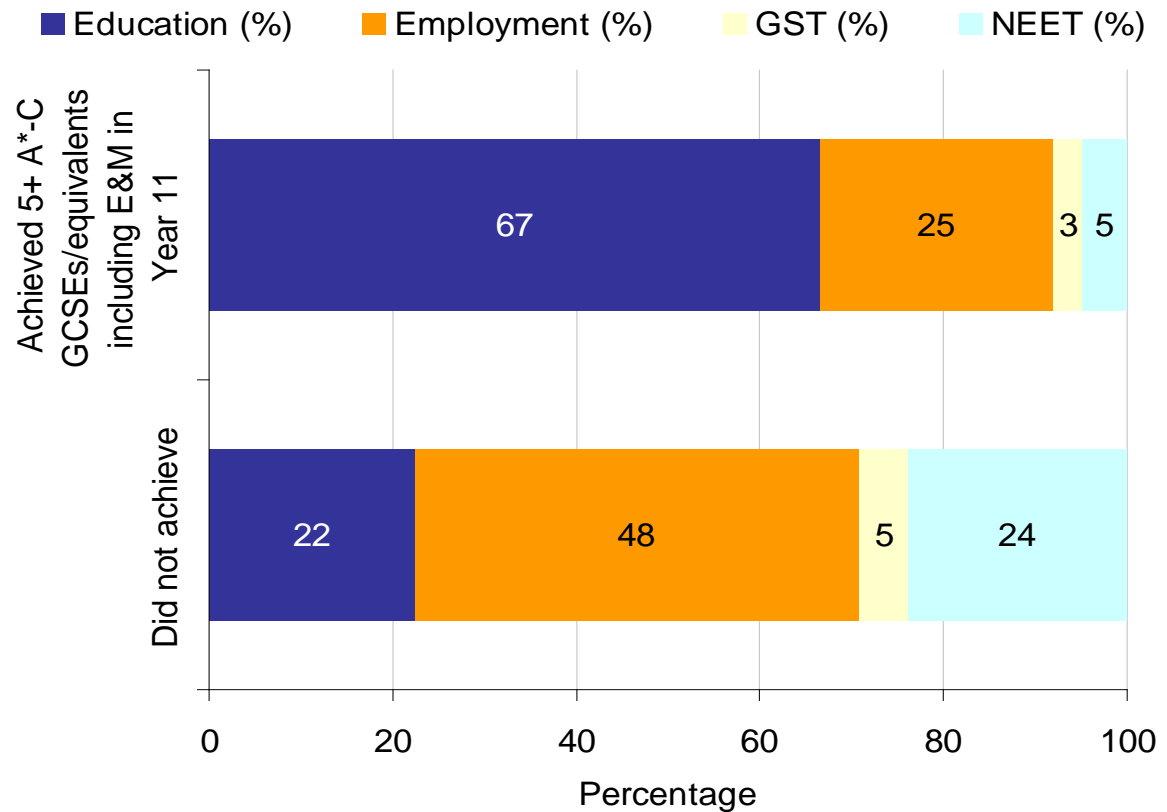
Example analyses using linked LSYPE and NPD data (1)

HE institution type at age 19 by school type in Year 9



Example analyses using linked LSYPE and NPD data (2)

Main Activity at age 19 by achievement in Year 11



LSYPE Data - accessing the linked data

- **First port of call should always be UKDA (www.esds.ac.uk) and our interactive website iLSYPE (www.education.gov.uk/ilstype).** **Some aggregate variables are already widely available through these routes.** For example: SEN, FSM, KS2/KS3 average point score, Achieved 5 or more GCSE's A*-C.
- **For more information see:**
<https://www.education.gov.uk/ilstype/workspaces/public/datasets/npd>
- If the data you want is not on those sources, it may be available direct from the team, subject to the completion of a confidentiality agreement.

LSYPE Data Linkage – our approach

- **Consent** – respondents informed in the first Wave of the Study of intention to link to NPD.
- **All or nothing approach** - if refused consent to linking in Wave 1 then they were not included in the study at all.
- **Reminders-** Respondents were not asked to re-consent to linking in following waves. Instead they were informed in advance letters that the linking would take place at each wave and were given contact details to find out further information.
- **Respondent website** – updated annually as another source of information on what happens with linked data and how respondents can go about withdrawing consent for linking.
- **We have since learned a lot of lessons from using this approach we will implement in the future, including other longitudinal studies where we are planning to link.**

Handling and sharing LSYPE/NPD linked data – some issues to consider:

- LS Team only hold some NPD data, NPD team provide sensitive data
- NPD data can only be linked in for LSYPE respondents (not drop outs)
- Can only provide NPD data if we are satisfied that requests meet the requirements of the Prescribed Persons regulations of the Education Act
- Data you request should be proportionate to your needs, not excessive
- Each request is created as a bespoke dataset
- Important to think about data protection throughout the whole process

Lessons learned

- Over the lifecourse of LSYPE there has been a major change in the way we view data and the responsibilities we have to those who provide it
- Be very clear from the outset on your data policies
- Take into account ethics as well as legal requirements- and future uses of the data
- Be aware of the tensions between data transparency and data protection
- Impact of getting it wrong is very concerning

Other longitudinal surveys linking to NPD

- ***Understanding Society***- NPD data extracts will be linked within the next month- now need to figure out how they can be widely shared securely

Our experiences from Understanding Society and LSYPE are helping to inform DfE colleagues for other longitudinal studies where NPD is linked in including:

- **MCS**
- **ALSPAC**

Thank you for your time

Any questions?

Contact the Longitudinal Surveys Team:

team.longitudinal@education.gsi.gov.uk