Inclusive data: better data and evidence to understand the lives, circumstances and needs of children and young people from the Gypsy, Traveller and Roma communities

Project overview report

Polly Vizard[1] and Polina Obolenskaya[2] with Chelsie Dolling[3]

Specialist advisors:
Sarah Mann[4] and Margaret Greenfields[5]

1 Associate Director, Centre for Analysis of Social Exclusion, LSE
2 Visiting Fellow, Centre for Analysis of Social Exclusion, LSE and Research Fellow, City, University of London
3 Training Manager, Friends Families and Travellers
4 Chief Executive Officer, Friends, Families and Travellers
5 Professor of Social Policy, Anglia Ruskin University and Visiting Professor (Social Policy and Community Engagement) Buckinghamshire New University

CASE Report 149

December 2023

This research report provides open access to the findings of the project 'Testing and demonstrating the value of the Growing Up in England (GUiE) dataset: Gypsy, Traveller and Roma children case study'. The research was funded supported by ADR UK (Administrative Data Research UK), an Economic and Social Research Council (ESRC) investment (part of UK Research and Innovation) [Grant number: 106391.R.000.1002.1128]. Supplementary funding for engagement outputs was received from the LSE knowledge exchange and impact support fund. Applications to use the GUiE dataset were made to the ONS Research Assessment Panel (RAP) and to the Department for Education. Analysis of the GUiE dataset was undertaken under Digital Economy (2017) regulated conditions in the ONS Secure Research Service (SRS). Ethics approval for the research was provided by the LSE Research Ethics Committee (reference numbers: 20005 and 94267). Ethics review of the research was also undertaken by the National Statisticians Data Ethics Committee (see: NSDEC Minutes and Agenda – April 2021 – UK Statistics Authority).

Acknowledgements

Sarah Mann, Director of Friends, Families and Travellers, and Margaret Greenfields, Professor of Professor of Social Policy, Anglia Ruskin University and Visiting Professor (Social Policy and Community Engagement) Buckinghamshire New University, were members of our project Advisory Group and regularly met with us to discuss, scope and plan the work and their input has helped to scope, shape and sensitivity test the research and our outputs. We are also grateful to other members of our Advisory Group for their input, support and scrutiny, including Karen Irving (Ministry of Housing, Communities and Local Government, lead analyst with a focus on Gypsy, Traveller and Roma), Rebecca Williams (Gypsy, Traveller and Roma lead, Cabinet Office Race Disparity Unit) Sanne Velthuis (University of Manchester), Stephanie Thomson (University of Aberdeen) and ONS (represented variously by Diana Airimitoaie, Rachel Mullis, Simon Heckenmueller, Yoli Bartlett, Jason Powell-Bavester and Clare Melson).

The research project ran from January 2021 to December 2022. In late 2021, the project was extended and expanded to include two engagement workshops to hear directly from children and young people from the communities and their families around the issues raised to the project. At this point, Friends, Families and Travellers became formal project partners and the engagement workshops were co-produced by the LSE research

team and Friends, Families and Travellers. The workshops were organised and facilitated by Chelsie Dowling. We are also extremely grateful to Michelle Gavin, Lucy Thomas and Sami McLaren at Friends, Families and Travellers for their input, guidance and note taking on the day. We would also like to extend our gratitude to the participants at the two direct engagement workshops that were held in July 2022. The views, perspectives and experiences highlighted at the workshops are written up in this report and have shaped and informed this research report and the interpretation of the statistical findings we present.

Other outputs from the project include a key-findings brief which provides a shorter summary of the findings in this research report and two key accessible outputs (an animated video and a digital comic) which have been tailored for onward sharing with children and young people from the Gypsy, Traveller and Roma communities and their families. We are hugely grateful to Gemma Challenger (artist) and a community narrator for their commitment and input into the animated video. At Nifty Fox Creative, Laura Evans-Hill, Mike and Simona Hodoňová for their commitment and creativity in developing a set of accessible outputs in different formats that bring the research findings to life.

The research process has straddled several years and over this period we have liaised with the many ONS staff that have worked on and contributed to the development of the GUiE dataset including in relation to early access arrangements, design and quality assurance and information gaps. Input, support, discussions, guidance and clarifications at different project stages have been provided by Emma Nelson, Diana Airimitoaie, Rachel Mullis, Simon Heckenmueller, Tracy Paul-Sharp, Yoli Bartlett, Gregory Goddard, Jason Powell-Bavester and Clare Melson. At the point at which the research team gained access to the GUiE dataset, we were the first external researchers to use it, and a formal process of meetings and clarifications was arranged between ADR UK and ONS in summer 2021 and the team have fed recommendations into the forward development of the dataset from an early stage. Additionally, we have benefited from the broader discussions, clarifications and information sharing as part of regular ONS GUiE technical and strategic groups during 2022 and early 2023 from colleagues across ONS including linkage related clarification from Jonathon Swan (ONS Data Engineering). Supplementary tables using data that we were not able to access ourselves for 100% Census and residuals analysis were produced by ONS colleagues. We are especially grateful for their work in producing this analysis. We are also grateful to ONS SRS staff for their help and support with statistical disclosure, and for input and support from Gary Connell and Chris Douglas (Department for

Education) for their help in relation to data access and information gaps in relation to the education side of the GUiE dataset. Elisha Duddle (Department for Education) also provided useful clarifications relating to the content of Wave 1 GUiE (YPMAD) attainment data.

Finally, we are finally grateful to Teeah Blake, Rosie French, Shayda Kashef (ADR UK) and to Karen Powell (ESRC UKRI) for their support and commitment to this project at different stages of this project; for comments and input from Tania Burchardt of the Centre for Analysis of Social Exclusion; and for ethics review comments from the Chair of the LSE Research Ethics Committee and the National Statisticians Data Ethics Committee NSDEC, which have shaped the research design.

Data acknowledgement and disclaimer

This work contains statistical data from ONS which is Crown Copyright. The work is based on the analysis of Wave 1 of the Growing Up in England (GUIE) dataset [Office for National Statistics] (2021): Secure Research Service Access. The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The work was produced using statistical data from ONS which is Crown Copyright. The views expressed are solely those of the authors and do not reflect the views of the Office for National Statistics or the wider UK Government. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. The work uses research datasets which may not exactly reproduce National Statistics aggregates. The analysis was carried out in the Secure Research Service, part of the Office for National Statistics.

[December 2023]

Contents

List of figures

List of tables

1. Introduction

This research report assesses the potential of the newly linked Growing Up in England (GUiE) dataset for challenging data exclusion and building up new evidence on the lives, circumstances and needs of children and young people from the Gypsy, Traveller and Roma communities. Tackling data exclusion and building up inclusive data infrastructure and evidence on smaller or highly disadvantaged or marginalised groups (sometimes described as 'vulnerable', 'hard to reach' or 'seldom heard' groups) that in the past have been systematically missing from, or invisible within, social statistics and quantitative analysis, is critical from an equalities and human rights perspective. In the wake of the COVID-19 pandemic, there is also increased high level recognition of the importance of challenging data exclusion and building up inclusive data infrastructure and evidence on the needs of groups that have in the past been under-represented in national statistics as a basis for effective public policy making. Gypsy, Traveller and Roma people are one of the groups that are most affected by severe data gaps, lack of robust quantitative evidence and lack of visibility and recognition within official statistics and reports, public service delivery and government policy. The report reviews this broader landscape; assesses the strengths and limitations of GUiE as one of a new generation of large scale longitudinal administrative data-population census linkages in developing more inclusive data infrastructure and evidence; and presents new experimental statistical findings that draw on Wave 1 GUiE to provide new insights and understanding in relation to the educational pathways and lifepaths of children and young people from the communities. Additionally, we report feedback from a direct engagement exercise with young people from the Gypsy, Traveller and Roma communities and their families about the issues raised by the project.

1.1 The Growing Up in England (GUiE) dataset

The background to this report is the creation of the Growing Up in England (GUiE) dataset. This is a new large-scale longitudinal administrative data population-census linkage that brings together administrative records from the education system and children's social care services with information about individual and household characteristics and circumstances from the 2011 Census. The GUiE dataset was created as a collaboration between the ONS, Department for Education and ADR UK. This followed on from the identification of data and evidence gaps in the 2017 Childhood Vulnerability in England Review and the establishment

of the Data for Children Partnership, a strategic partnership between ADR UK, the Office of the Children's Commissioner for England and others. The GUiE data linkage was undertaken by ONS and the GUiE Wave 1 and Wave 2 datasets, together with elements of Wave 3, are now available via the ONS Secure Research Service for UKSA accredited researchers to access.

1.2 Research aims

The project was from the outset conceptualised as an early exploratory project that aimed to assess the 'value added' of the GUiE linkage and its potential of building up new evidence on the lives, circumstances and needs of children and young people from the Gypsy, Traveller and Roma communities including by producing an initial set of experimental findings and setting out recommendations on the forward development of the GUiE data infrastructure. Our original research proposal specified three main objectives:

- To assess the potential of Wave 1 of the linked Growing Up in England dataset for the purposes of Gypsy, Traveller and Roma research, including by providing feedback on unmatched cases.
- To conduct initial analysis of the linked Growing Up in England data with a view to demonstrating its value added in building up new evidence, knowledge and understanding of the Gypsy, Traveller and Roma group.
- To shape the development of the Growing Up in England dataset, in order to maximise its potential for the analysis of the Gypsy, Traveller and Roma group going forward.

Additionally, as we note below (under 'ethics and engagement'), following input from the National Statistics Data Ethics Advisory Committee (NSDEC) and additional support from ADR UK and the LSE Knowledge Exchange and Impact fund, the project was expanded mid-way to incorporate:

- A direct engagement exercise with children and young people from the Gypsy, Traveller and Roma communities and their families (with two direct engagement workshops focusing on the issues raised by our project co-organised with Friends, Families and Travellers and held in July 2022;

- The development of a suite of accessible project outputs, including an animation designed for onward sharing with workshop participants. These outputs were co-produced with Friends, Families and Travellers and creative designers Nifty Fox and a community artist and narrator.

In practice, the project has also had much more of a 'trailblazer' role than we originally envisaged with the project contributing to broader GUiE development by revealing and raising issues, barriers and information gaps and feeding recommendations into ONS and ADR UK processes throughout the lifetime of the project. The team were the first academic researchers to access GUiE and we have addressed multiple issues relating to access, legal gateways, ethics, quality assurance and information gaps during the lifetime of the project. This has included a process of meetings and correspondences with ONS teams; raising issues and highlighting barriers through ADR UK quarterly monitoring processes and participation in the ADR England Data Linkage Steering Group; a formal system of quality assurance feedback and meetings with ONS put into place with support from ADR UK; regular participation in the ONS GUiE technical and steering groups once they were created; providing comments on ONS and ADR UK GUiE related materials; and arranging meetings and corresponding with Department for Education, ONS GUiE researchers and other stakeholders. Many of the issues we have raised have been clarified or resolved, or partially clarified and resolved, through these processes as well as through the publication of GUiE dataset metadata and User Guide (ONS n.d.b.); the wave of GUiE documentation and findings that were published in Autumn 2022 (ONS cdef); the release of Wave 2 GUiE; the release of elements of Wave 3 GUiE and the imminent release of other components of Wave 3; and the release of the AEDE spine duplicate flag dataset. A record of issues we have addressed and followed up during the lifetime of the project and some key clarifications that have informed our own analysis and which we believe will be useful for other GUiE researchers is included this report. Additionally, our recommendations relating to the forward development of GUiE are included in the conclusions of this report.

1.3 Ethics and engagement

The project has benefited from the advice, guidance and input of the national charity Friends, Families and Travellers (gypsy-traveller.org) (FFT). FFT are a national charity who work on behalf of all Gypsies, Travellers and Roma regardless of ethnicity, culture or background. The

charity aims to end racism and discrimination against Gypsies, Travellers and Roma and to protect the right to pursue a nomadic way of life, and provides advice and consultancy, promotes health and wellbeing, contributes to research and policy making and deliver trainings (including specific training on cultural awareness) amongst other activities. Friends, Families and Travellers has participated in our project Advisory Group and a series of additional follow-up meetings throughout the lifetime of the project. Additionally, when the project was expanded to include a direct engagement exercise, FFT became a formal funded project partner and co-organised and facilitated the direct engagement workshops. We are also particularly grateful for specialist advice and input from Professor Margaret Greenfields who was a member of our project Advisory Group and has regularly met with us to discuss, plan and guide the research and has provided comments on our outputs.

To undertake the quantitative research exercise, we gained early access to the Wave 1 GUiE dataset in June 2021. As there was no established single legal gateway for researcher access to GUiE at that time accessing GUiE involved completing separate applications to both the Department for Education and the Research Accreditation Panel (RAP). Ethics approval for the project was obtained from the LSE Research Ethics Committee and in addition our project was referred by RAP for an ethics review by the National Statisticians Data Ethics Advisory Committee (NSDEC). NSDEC provides additional ethical scrutiny of selected administrative data research projects applying the six core principles of 'statistics for the public good' set out in Figure 1. The NSDEC Committee formally reviewed our project and raised several issues that we were required to address:

- The Committee requested further assurance that the researchers would engage with the Gypsy, Traveller and Roma communities, and that the Gypsy, Traveller and Roma community supports the public good of this work;
- The Committee recommended that the researchers consider that the 'education equals progress' assumption may not necessarily be shared by all;
- The Committee stated that the application would benefit from more detail on the whether there are any potential consequences for the Gypsy, Traveller and Roma communities and how these will be mitigated;
- The Committee questioned whether children who may be absent from education records for one year, but later reappear, would be excluded from the dataset;
- The Committee highlighted the importance of consideration of missing data, for example, as a result of self-identification of key

characteristics in both the census and the feasibility All Education Data for England (AEDE) data, and also missing data associated with exclusions and periods of absence from school.

Figure 1: Ethical principles applied by the National Statisticians Data Ethics Advisory Committee (NSDEC)

- The use of data has clear benefits for users and serves the public good.
- The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.
- The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.
- Data used and methods employed are consistent with legal requirements such as Data Protection Legislation[the Data Protection Act 2018, including the GDPR)], the Human Rights Act 1998, the Statistics and Registration Service Act 2007 and the common law duty of confidence.
- The views of the public are considered in light of the data used and the perceived benefits of the research.
- The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

These recommendations have been addressed in multiple ways through the lifetime of the project. At a very early stage, we fed recommendations on the importance of nonattainment education data (exclusions, absences, SEND, free school meals and so forth) into ONS and ADR UK workstreams on GUiE development. In line with the Committee's guidance, we provide context for the current study by reviewing qualitative literature including specialised studies, lived experience studies, issues highlighted in the analysis and advocacy of civil society in this report, and the high findings and recommendations of multiple high-level inquiries and reviews. We have also developed a good practice framework for assessing the strengths and limitations of administrative data linkages which we use to assess the inclusivity of the GUiE linkage, taking account of key concerns such as missingness and disproportionality. In applying this framework, we caveat the extent to which that GUiE dataset can be thought of as representative of the underlying population of Gypsy, Traveller and Roma young people. Additionally, building on the NSDEC principles of engagement, as well as ADR UK initiatives on public engagement, data use and the public good, and working directly with Friends, Families and Travellers, we co-organised

two direct engagement workshops that enabled us to hear directly from children and young people from the Gypsy, Traveller and Roma communities and their families about their views and experiences both in relation to their interactions with the education system, and their broader experiences relating to official data and data use. Finally, we have worked with Friends, Families and Travellers, creative designers Nifty Fox and a community artist and narrator on a suite of accessible outputs, including an animated video.

Overall, we believe that building on NSDEC input and the approach to administrative research in the public good being developed by ADR UK, we have developed a good practice mixed methods framework for undertaking our project, with ethical principles and engagement at its heart (Figure 2). Additional ethics approval for the engagement workshops and the suite of tailored engagement outputs was obtained from the LSE Research Ethics Committee.

Figure 2: Framework of good practice: ethics and engagement

Ethics review

- ***LSE full ethics review for quantitative work using GUiE***
- ***LSE full ethics review for qualitative engagement workshops***
- ***NSDEC ethics review***

Engagement

- ***ADR UK level engagement***
  - ADR UK public dialogue
  - Inclusion of a representative civil society organisation in the ADR England Children and Young People Representative Panel
- ***Dataset level engagement***
  - Consideration of the GUiE linkage by the ADR England Children and Young People Representative Panel
- ***Project specific engagement***
  - Civil society representative organisation (Friends, Families and Travellers) guidance and advice through the project Advisory Group
  - Friends, Families and Travellers co-production of direct engagement exercise
  - Accessible outputs

1.4 Limitations, caveats and ethnicity classifications

There are several important limitations of our analysis. In particular, we underline that the statistical findings using GUiE that we present in this report are of an experimental nature: our central aim in the report is to set out a preliminary set of findings on the feasibility of using the GUiE dataset to undertake analysis of the records of Gypsy, Traveller and Roma young people; to assess the limitations and potential of this approach; and to set out an initial set of findings. This is research we hope to build on and take forward in the future. Notably, our analysis in the report identifies several general limitations of GUiE including the fact that while young person level and some household level information census information was included in the Wave 1 GUiE link, the individual records of other household members are missing. Additionally, we identify several key limitations and caveats relating to the extent to which GUiE can be thought of as representative of the underlying population of Gypsy, Traveller and Roma children and young people including key issues around differential linkage rates. This restricts the extent to which generalisations can be drawn from our experimental statistical results. The concluding section in this report looks forward and sets out recommendations on how methodological improvements to the design and inclusivity of the GUiE data infrastructure could be made in the future.

It should also be noted that while assessments of GUiE disproportionality in the GUiE linkage (residuals analysis) and of any GUiE selection or compositional effects (through analysis of the 100% Census) were included in our original project specification, we were not permitted direct access to the GUiE residuals (unmatched cases) or to access 100% Census within our project. However, through a process established with support from ADR UK, ONS have undertaken an analysis of the feasibility AEDE residuals and the 100% Census for us and the findings from these exercises are included in this report. Additionally, it should be noted that the data in GUiE only goes up to 2014/15 and that one of our key recommendations on GUiE forward development is that the Wave 1 and Wave 2 data is updated, both in terms of its education content and in terms of linkage to the 2021 Census, to ensure that GUiE remains useful for public policy purposes in the 2020s. This would also support good practice in ethnicity identification, as a specific Roma identifier was included for the first time in the 2021 Census.

Our use of the term 'Gypsy, Traveller and Roma' in this report also has important limitations and caveats. In the UK context, recommended good practice is to understand this term as an umbrella category to refer to

multiple different and distinct subgroups that are recognised as ethnic minorities under the Equality Act (2010). This includes Romany Gypsies (both English and Welsh Romany Gypsies); Scottish Gypsy Travellers (an indigenous specific group); Irish Travellers (with specific Irish roots); Roma people (who are generally understood to be more recent migrants from Central and Eastern Europe). In addition, several other subgroups which are not being recognised as ethnic minorities under the Equality Act (2010) are often referred to under the umbrella term 'Gypsy, Traveller and Roma', including New Age Travellers, Boaters, Bargees and Showpeople (Friends, Families and Travellers n.d., Women and Equalities Committee 2019, Cabinet Office 2022, Traveller Movement 2022).

We argue that there are several good reasons for making use of the collective term 'Gypsy, Traveller and Roma'. Conceptually, the use of a collective term provides recognition that different subgroups can share characteristics including shared cultures, histories and values (such as a shared history of nomadism); have similar experiences of inequalities, disadvantage and discrimination; and face common challenges (Friends, Families and Travellers n.d., Women and Equalities Committee 2019, Cabinet Office 2022, Traveller Movement 2022). Additionally, use of the collective term 'Gypsy, Traveller and Roma' can be a pragmatic response to existing data limitations and imperfect ethnicity classifications. This is the case in the current project, as our analysis of the new Growing Up in England dataset, which combines administrative education data records for the period 2001/2-2014/15 with the detailed information on individual characteristics and household circumstances from the 2011 Census, is limited by the imperfect ethnicity classifications that are available in the original data sources.

However, in using the collective term 'Gypsy, Traveller and Roma', it is also important to recognise that no single definition of the term 'Gypsy, Traveller and Roma' is universally accepted and that the use of the collective term 'Gypsy, Traveller and Roma' to describe people from different complex and interconnecting ethnicities, identities and communities can conceal considerable heterogeneity in terms of subgroups with distinct cultural traditions, histories and needs. For some policy purposes, more fine-grained disaggregated categories are more appropriate as the used of combined classifications can conceal important differences which may be critically important for policy purposes. For example, Roma people are more likely to be recent migrants and to have English as a second language. Additionally, while the term Gypsy, Traveller and Roma is often used to describe people with a share history of nomadism, it is important to recognise the majority of Gypsy and Traveller

people in the UK today live in settled (bricks and mortar) accommodation and that no Roma people are known to be living nomadically (Friends, Families and Travellers n.d., Women and Equalities Committee 2019, Cabinet Office 2022, Traveller Movement 2022).

Looking forward, the inclusion of a Roma identifier in the 2021 Census will improve the potential for disaggregated analysis in the 2020s. As we highlight in this report, several recent high-level inquires and reviews have underlined the importance of consistent and harmonised ethnicity classifications across data collections building on the Census 2021 ethnicity 18+1 ethnicity collection as well as the importance of more fine-grained categories in ethnicity monitoring. Our recommendations at the end of this report include the development of an updated GUiE dataset build that links to the 2021 Census as well as the adoption of consistent and harmonised ethnicity categories in administrative education data.

1.5 Organisation of this report

Following on from this introduction, the report is organised as follows:

- Section 2 provides background to and context for our research project by examining the nature of data exclusion and examining the aims and objectives of the inclusive data agenda. We begin by setting the scene in terms of the persistence of systematic gaps in data and evidence on the lives, experiences and needs of people from the Gypsy, Traveller and Roma. We next examine findings from several recent high-level inquiries and initiatives that have identified the importance of challenging data exclusion and building up more inclusive data infrastructure and evidence as a basis for equalities, human rights and public service delivery in the 2020s. Finally, we introduce the idea of administrative data-population census linkages and provide initial comments on the potential role that they can play as part of broader efforts to tackle data exclusion and build up inclusive data infrastructure and robust evidence on smaller or high risk (sometimes known as 'vulnerable' 'seldom heard' or 'hard to reach') groups that in the past have been systematically 'missing from', or 'invisible within', social statistics and quantitative analysis.
- Section 3 provides further context for our GUiE study by reviewing what we know about educational disadvantage and broader multidimensional inequalities that impact on Gypsy, Traveller and Roma young people and their families from existing evidence. Building on the recommendations of the NSDEC Committee concerning our report highlighted in the introduction (section 1), we review key specialist

studies (including qualitative studies) in this field and provide an overview of what can be learnt in terms of key issues, barriers, good practice and policy solutions from existing civil society organisations that represent Gypsy, Traveller and Roma people. Additionally, we highlight relevant findings from recent high level inquires and reviews and identify some of the key public policy developments that have particular relevance to the current study.

- Section 4 provides details of the analytical framework that we use in this report to assess the strengths and limitations of administrative data linkages in tackling data exclusion and building up inclusive data infrastructure and evidence. We begin by identifying the major potential strengths of administrative data linkages (including large scale longitudinal administrative data-population linkages such as the newly linked Growing Up in England dataset) in challenging data exclusion and building up inclusive data infrastructure and evidence. Next, we examine some of the key limitations and weaknesses that impact on administrative data linkages, including key issues of coverage, bias, error and disproportionality, and identify good practice recommendations for addressing these concerns. Finally, building on good practice recommendations set out by ONS, UKSA and others relating to assessing quality and inclusivity, we outline a good practice analytical framework for assessing the quality and inclusivity of the Growing Up in England dataset that we apply in subsequent sections of this report.
- Section 5 provides a general introduction to the Growing Up in England (GUiE) dataset and a general assessment of its potential, strengths and limitations. We begin with a discussion of the background to and aims and purposes of the data linkage and an overview of the underlying data sources that it aims to bring together and integrate. We then provide a summary of the evolution of the GUiE build and of published information on content, design and quality drawing on a series of ONS outputs. Next, we provide a record of some of the key issues relating to content, design and quality that we have followed up during the lifetime of our project and of the key clarifications that we have received. These are important to record as they inform our assessments of the GUiE Wave 1 analysis and our approach to the research exercise we report in section 7 and may also be useful for other researchers. Finally, drawing on the above, we set out our working summary description of GUiE Waves 1–3.
- Section 6 provides further details of the strengths and limitations of the Wave 1 Growing Up in England (GUiE) dataset for undertaking research on the lives, outcomes and needs of Gypsy, Traveller and Roma children and young people. We begin by highlighting the rich fine-grained

ethnicity information that is available in Wave 1 GUiE as a major strength and breakthrough in terms of supporting the identification of the records of Gypsy, Traveller and Roma young people at scale. We then address a series of caveats and limitations relating to missingness, invisibility and inclusivity and which impact on the extent to which the dataset can be viewed as being representative of the underlying population of Gypsy, Traveller and Roma young people. Next, we provide detailed statistical assessments of the extent to which the records of Gypsy, Traveller and Roma young people are successfully matched into the Wave 1 dataset on an equal basis and the extent to which the linkage with the 2011 Census as associated with deprivation-related compositional effects. Finally, we draw attention to the importance of interpreting statistical findings on Gypsy, Traveller and Roma young people produced using the GUiE dataset in the light of these caveats and limitations.

- Section 7 sets out findings from an initial research exercise that uses the Growing Up in England (GUiE) dataset to build up new experimental evidence on the lives, circumstances and needs of young people from the Gypsy, Traveller and Roma communities. In the research exercise, we use the Wave 1 GUiE dataset to identify the records of substantially larger numbers of Gypsy, Traveller and Roma young people than has been possible in most previous research. Our methods for maximising the identification of the records of Gypsy, Traveller and Roma young people within the GUiE Wave 1 data exploit the large-scale and longitudinal nature of the dataset as well as the potential for pooling ethnicity information sourced from the 2011 Census and administrative education data systems. Applying these methods, we use the education side of the GUiE Wave 1 dataset to follow the educational journeys of three academic year cohorts of young people over four years, from the academic year they turned 15 to the academic year that they turned 19. We report on the educational pathways of the Gypsy, Traveller and Roma young people within the cohorts as they transition through 14-16 and 16-18 education and training through any route (schools, further education, vocational education, work-based learning or apprenticeships) and on their highest educational attainment (Level 1, Level 2 or Level 3 qualifications) by ages 17 and 19, comparing these to the pathways and attainment of their peers. Additionally, drawing on the rich information about individual characteristics and household circumstances recorded in the 2011 Census, we use the GUiE Wave 1 data to examine the multidimensional child disadvantage characteristics of the Gypsy, Traveller and Roma children identified in GUiE Wave 1 (observed at ages 14 and 15) and to assess the strength of the associations between different indicators of multidimensional child

disadvantage and highest educational attainment by age 19 attained by any route for Gypsy, Traveller and Roma and other young people.

- Section 8 provides a write up of the feedback from two direct engagement workshops with young people aged 14-19 from the Gypsy, Traveller and Roma communities and their families that were co-organised with Friends, Families and Travellers in July 2022. The section begins with some background relating to the broader public engagement exercises being undertaken by ADR UK and partners around administrative data and the public good and the GUiE data linkage. We then set out methods (including the process for obtaining informed consent). Next, we provide a record of the feedback from the first part of the workshop, which focused on experiences of the education system (schools, colleges and training) and home education. We then move on to the second part of the workshop, providing a record of experiences and view-points in relation to official data collection and statistics, including issues around the disclosure of Gypsy, Traveller and Roma identity. We finally reflect on how the findings from the engagement workshops have fed into the broader project and draw together a list of suggestions and policy recommendations, drawing on the feedback from participants.
- Finally, section 9 draws together the key findings and recommendations from this report and considers the policy implications of our analysis. We begin with a summary of key findings. We then highlight our general recommendations on the further development of the GUiE dataset and on data inclusivity going forward. Next, we address the policy implications of our analysis. Finally, we set out our future research plans for extending our research on the life-paths, circumstances and needs of young people from the Gypsy, Traveller and Roma communities.

2. The inclusive data agenda: challenging data exclusion and building up more inclusive data infrastructure and evidence for the 2020s

This section provides background to and context for the research project by highlighting the phenomenon of data exclusion and the potential role that the new generation of administrative data linkages can play in building up more inclusive data infrastructure and evidence for the 2020s. We begin by setting the scene in terms of the persistence of systematic gaps in data and evidence on the lives, circumstances and needs of young people from the Gypsy, Traveller and Roma communities. We next examine findings from several recent high-level inquiries and initiatives that have identified the importance of challenging data exclusion and building up more inclusive data infrastructure and evidence as a basis for equalities, human rights and public service delivery going forward. Finally, we begin to consider the specific contribution that large-scale administrative data-population census can play in tackling data exclusion and making breakthroughs in building up evidence on smaller or highly disadvantaged or marginalised groups that in the past have been systematically missing from, or invisible within, social statistics and quantitative analysis.

2.1 Data exclusion and systematic gaps in data and evidence on the lives, experiences and needs of people from the Gypsy, Traveller and Roma communities

'Data exclusion' is the phenomenon whereby some of the most disadvantaged and marginalised groups of people in the UK remain systematically 'missing' from, or 'invisible' within, national data infrastructure, statistics and evidence.

In a previous research report (Child poverty and multidimensional disadvantage: tackling data exclusion and extending the evidence base on 'missing' and 'invisible' groups) we highlighted that data exclusion is a critically important concern for at least three key reasons. First, data exclusion can result in failures of understanding and measurement (for example, underestimating the extent of child poverty because relevant population sub-groups are not included within official estimates). Second, data exclusion can result in failures of monitoring and reporting (when the position of sub-groups is not addressed within official statistics, analysis and reports). Third, data exclusion can result in failures of public policy and the needs of the relevant sub-groups not being addressed or given inadequate recognition within public policies and frameworks for public

service delivery (Vizard et al 2018). In this way, data exclusion contributes to the paradox whereby some of the most highly disadvantaged and marginalised subpopulations that are most in need are least likely to be adequately supported by high-quality public services (the so-called 'inverse care law' - for relevant recent findings, see, for example, Gordon 2012, Sutton Trust 2017 and Vizard et al 2018). Concerns about data exclusion and their implications for public policy were underlined by a series of child vulnerability assessments undertaken by the Children's Commissioner's Office in the years running up to the COVID-19 pandemic. These put the spotlight on the data and evidence gaps relating to a long list of high-risk groups of children and young people, underlining the critical importance of a concerted national statistics response in terms of developing improved data infrastructure and evidence as a basis for effective public policy relating to children (CCO 2017, 2019).

Gypsy, Traveller and Roma people are one of the key groups that are most affected by severe data gaps, lack of robust quantitative evidence and understanding, and lack of visibility and recognition within official statistics and reports, public service delivery and government policy. The extent, importance and implications of systematic data and evidence gaps affecting the Gypsy, Traveller and Roma communities and the public policy consequences of these gaps has been regularly highlighted in civil society organisation research and advocacy and equality and human rights monitoring and reporting over many years and has been recognised in key high-level inquiries and initiatives discussed below (for example: Friends, Families and Travellers 2022; Traveller Movement 2022; Equality and Human Rights Commission 2010, 2015, 2018; Women and Equalities Committee 2019; Cabinet Office Race Disparities Audit (2017); UKSA Inclusive Data Taskforce (2021a); Children's Commissioners Office (2017, 2019). As Friends, Families and Travellers (2022) note, lack of reliable statistics relating to the communities and the issue of Gypsy, Traveller and Roma exclusion from and invisibility in mainstream datasets mean that their needs cannot be effectively identified, understood or addressed within public services provision and delivery.

In our previous research project, we reviewed the well-known limitations of traditional social survey analysis that underlie the problem of data exclusion including the limited coverage of sampling frames, the absence of relevant identifiers and low population fraction / numbers (Vizard et al 2018; c.f. section 4.1). In the current report, we look to the new generation of large-scale longitudinal administrative data -population census linkages in challenging data exclusion and building up more inclusive data infrastructure and evidence. However, it is important to state at the outset

that we recognise that administrative data is not a panacea and that there are also major gaps in and limitations of administrative data infrastructure which can have major consequences for coverage, analysis and interpretation. Indeed, civil society organisations representing Gypsy, Traveller and Roma people regularly highlight the exclusion and invisibility of Gypsy, Traveller and Roma people in routine national administrative data systems and the consequential limitations of official statistics, evidence and policy based on these data sources.

A classic example is the historic failure of the NHS Register to include a Gypsy, Traveller or Roma identifier. This key gap in NHS ethnicity monitoring systems was highlighted repeatedly in civil society analysis and advocacy during the 2010s but was not addressed, meaning that major gaps in reliable statistical data and evidence on health outcomes for Gypsy and Irish Traveller people remained a key concern during the second decade of the 21st century. In early 2020, the consequences for health monitoring and policy - together with the systematic health access and protection gaps affecting the communities - were exposed when the COVID-19 pandemic struck. For example, in 2021, a cross-government report on the ethnic dimension of COVID-19 disparities concluded that "vaccination estimates for Gypsy, Roma and Irish Traveller populations are not available as NHS systems do not have these categories available to choose" (Cabinet Office Race Disparity Unit 2021).

In contrast, as we discuss in detail in this report, Wave 1 Growing Up in England (GUiE) brings together and integrates two of the best data sources that we have for identifying the records of Gypsy, Traveller and Roma young people: education data drawn from administrative data systems and population census data. However, it is critical to be crystal clear from the outset that while each of these data sources has strong potential for challenging data exclusion and making breakthroughs in Gypsy, Traveller and Roma research relatively speaking, they nevertheless have multiple limitations including in relation to coverage and ethnicity identification and that ethnicity classifications adopted in administrative education systems and the 2011 Census both fall short of recognised good practice in important respects. These limitations are discussed in detail in this report in sections 6.2.1-6.2.3. The process of creating a linked dataset raises additional key concerns in terms of inclusivity, and the limitations of GUiE from this point of view are discussed in detail in section 6.2.4.

Moreover, as we highlighted in our previous research report (Vizard et al 2018), data exclusion can be driven by broader factors including underlying patterns of inequalities, discrimination and unequal social recognition. These can impact on trust and the willingness of Gypsy,

Traveller and Roma people to disclose their identity in the context of both social survey and official administrative data collections. The issue of non-disclosure of Gypsy, Traveller and Roma identity was identified in the key findings of the Women and Equalities Committee Report on tackling Gypsy, Traveller and Roma inequalities (discussed below) and is a key theme in existing research and civil society organisation analysis (discussed in section 3.2) and in the feedback from the engagement workshops undertaken in the context of this project (discussed in section 8 below).

Additionally, while the 'big data' revolution opens-up new opportunities for addressing data exclusion, it also raises new ethical concerns related to data use, privacy and confidentiality, consent, discrimination and other potential harms. Building up the evidence base on missing and invisible groups cannot therefore be undertaken in a statistical vacuum and requires a range of broader initiatives including advocacy, engagement, social activism and broader measures to challenge discrimination and to protect and promote information rights.

2.2 Inclusive data infrastructure and evidence for the 2020s: findings and recommendations from high-level inquiries and reviews

On a positive note, the importance of addressing data exclusion and building up more inclusive data infrastructure and evidence has been moving up the national statistics and policy agenda. Prior to the COVID-19 pandemic, the extent of systematic data and evidence gaps was recognised in several initiatives including in a major report on tackling inequalities facing the Gypsy, Traveller and Roma communities by the Women and Equalities Select Committee; in the findings of the Race Disparities Audit; and in a series of child vulnerability assessments undertaken by the Office of the Children's Commissioner for England. The pandemic itself resulted in increased recognition of the importance of key data gaps including the absence of a Gypsy, Traveller and Roma identifier in the NHS dictionary (Race Disparity Unit 2021). Additionally, in 2021, the UK Statistics Authority Inclusive Data Taskforce addressed the importance of building up inclusive national data infrastructure and evidence that captures and reflects the needs of all groups as a basis for effective government policies, interventions and support. The inclusive data agenda subsequently emerged as a strategic priority for national statistics (UKSA 2021ab, Diamond 2021b).

2.2.1 Findings of the Women and Equalities Committee (2019) report on tackling Gypsy, Traveller and Roma inequalities

The extent of the data and evidence gaps affecting the Gypsy, Traveller and Roma communities was brought to the fore by a major report on tackling inequalities facing the communities by the Women and Equalities Committee in 2019. The importance of better data and evidence was a key theme in oral and written evidence submitted to the Committee with implications for the identification of needs and service commissioning and delivery nationally and locally. The Committee found that most public bodies do not provide the option of disclosing Gypsy, Traveller or Roma identity and that the communities are rarely considered in policies and strategies. This 'conspicuous lack of due regard' for the needs of Gypsy, Traveller and Roma people resulted in the communities 'being poorly served by policy-makers and public services', contributing in turn to the 'very poor outcomes' of Gypsy, Traveller and Roma people across critical areas of life including in relation to health, education, employment and hate crimes; to broader feelings of exclusion and lack of trust; and ultimately to severe discrimination. Overall, the Committee concluded that lack of consistent data collection on Gypsy, Roma and Traveller people means that public bodies are failing to tackle inequalities that are clearly evidenced in academic research and in work being carried out by grassroots organisations (Women and Equalities Committee 2019).

These broad conclusions were underpinned by three sets of key findings on data and evidence gaps. First, the Committee's report laid bare the extent of information gaps relating to the most basic facts including the absence of reliable population estimates of the number of Gypsy, Roma and Traveller people who live in the UK. The inclusion of a Gypsy and Traveller identifier for the first time within the 2011 Census was recognised as an important advance and pointed to a total headcount of 58,000 Gypsy and Traveller people in England and Wales, with an additional 4000 recorded in Scotland. Additionally, education data was recognised as an important basis for identification of Gypsy, Traveller and Roma children. However, both census and education identifiers were recognised as being limited in important respects and as underestimating the population of Gypsy, Traveller and Roma population. The Committee concluded that the true population of Gypsy, Traveller and Roma is unknown but that best estimates pointed to around 100,000-300,000 Gypsy / Traveller people and up to 200,000 Roma people living in UK (Women and Equalities Committee 2019, drawing on Council of Europe and University of Salford estimates).

Second, while highlighting the systematic nature of data and evidence gaps, the Committee's analysis nevertheless identified variable practice

across public services. Health was identified in the Committee's report as an area of particularly poor practice and one of the Committee's major recommendations was that Gypsy, Irish Traveller and Roma categories be added to the NHS data dictionary as a matter of urgency. Some areas of relative good practice were also identified, including in the context of population census data and education services, and the Committee recommended that good practice be rolled out across all public services. The urgency of 'swift action' to enable Gypsy, Traveller and Roma people to identify themselves to public services so that their needs can be taken into account was underlined (Women and Equalities Committee 2019). A follow up assessment identified lack of focus and specificity in the Government response to the Committee's 2019 recommendations. At this stage, it was suggested that the Race Disparity Audit should review all government and public datasets and ensure that all departments are using as a minimum the 2011 census ethnicity classifications and that the Government reports back on progress on health funding and data collection.

Third, the Select Committee report highlighted the importance of broader factors that are of important for understanding the nature and scope of the data and evidence gaps affecting the Gypsy, Traveller and Roma communities. This includes possible reluctance to disclose Gypsy, Traveller and Roma identity, with the Committee report highlighting witness evidence that Gypsy, Traveller and Roma people may be reluctant to self-identity, even where the option is available, as there is a widespread fear that disclosing identity will result in hostile treatment and discrimination. Further, the Committee suggested that this reluctance to self-identity is best understood in the context of broader distrust of government and official data collections, and the broader historical context of persecution and discrimination, as well as ongoing experiences of hostile treatment, discrimination and inequalities. The Committee's recommendations here included outreach and engagement work with grassroots Gypsy, Roma and Traveller organisations to formulate a wide-ranging campaign to explain the importance of data collections and to encourage self-disclosure (Women and Equalities Committee 2019). The importance of addressing participation in data collection is also highlighted in the Inclusive Data Taskforce, which identified distrust in government and government statistics as a key barrier to inclusive data for the Gypsy, Traveller and Roma communities. The Taskforce also emphasised the importance of trust going forward, including reassurances around privacy, data confidentiality and anonymity (UKSA 2021a).

2.2.2 The Race Disparities Audit (2017)

The Women and Equalities Committee's analysis drew in part of the findings of the previous Cabinet Office *Race Disparities Audit* published in 2017. The Race Disparities Unit provided an in-depth audit of existing data sources for different ethnic groups within administrative data sets across government and public services and in social survey data collections. It resulted in the disclosure and release of data where this was available and the creation of a new website bringing together disaggregated data about the experiences and outcomes of people from different ethnic groups across different critical areas of life including in relation to health, education, employment, violence and the police and criminal justice system ('Ethnicity facts and figures'). The Audit highlighted data gaps for Gypsy, Traveller and Roma people in most areas, especially in relation to health and the criminal justice system, as well as evidence of exceptionally poor outcomes where data did exist. This included the lowest educational outcomes and progress, and the highest rates of exclusions and absences, of any ethnic group; as well as extremely poor rates of post-16 participation (Cabinet Office 2018). An assessment of 130 published datasets undertaken as part of the audit found that only 27 included relevant identifiers (detailed census categories or Gypsy / Roma and Travellers of Irish Heritage classifications), with key gaps and inconsistent recording classifications even in areas of 'relative' good practice' (Women and Equalities Committee 2019, Cabinet Office 2018; Cabinet Office n.d.).

In follow up work, Cabinet Office (2022) provided an updated overview of the availability of statistics and evidence on the lives, experiences, circumstances, needs of people from the Gypsy, Traveller and Roma ethnic groups across different areas of life, with higher education as well as the NHS data dictionary highlighted as key gaps. Recommendations on improving data availability and quality emphasised the importance of moving forward based on harmonised consistent classifications and categories across datasets, including in the context of the NHS Data Dictionary. The review also specifically highlighted the role of data linkage as a way of producing more robust data for the Gypsy, Traveller or Roma groups, and noted that this might improve the quality of ethnicity analysis if an ethnicity classification that is known to be more reliable is linked from another dataset, citing the linkage of the Census to Hospital Episodes Statistics (HES) data and death registrations. The ethnicity classifications for Gypsy, Traveller and Roma groups are not included in the HES data, and are not collected in death registrations, and linkage information

provided a way of providing some information on death rates for the Gypsy and Irish Traveller groups in the context of COVID-19.

2.2.3 The Children's Commissioner's Office Child Vulnerability Initiative

As noted above, during the years running up to the COVID-19 pandemic, a series of child vulnerability assessments undertaken by the Children's Commissioner's Office (CCO 2017, 2018, 2019) also highlighted the critical importance of delivering substantial improvements to the coverage and quality of data relating to the identification, needs and outcomes of vulnerable children. An initial review identified 71 different operational definitions and uses of the term 'vulnerable children' that were being used across different government departments and emphasised the importance of an overarching holistic analytical framework for identifying and measuring children's needs as a basis for developing an accurate and valid estimate of the scale of aggregate need and as a foundation for delivering an adequate social and public policy response. The Child Vulnerability Framework was produced as a starting point for taking this agenda forward and included a list of 70 vulnerable groups falling with the scope of seven domain of vulnerability (or multiple vulnerability domains):

- Children receiving statutory care, support or interventions.
- Children known to have experienced specific personal harm.
- Children with a disability, ill-health or developmental difficulties.
- Children in households or families with characteristics that indicate higher potential likelihood of current and future harm.
- Children who are vulnerable by virtue of their identity or nationality
- Children at risk in relation to activity or institutions outside the home.
- Children caring for others.

The CCO assessments undertaken in 2017, 2018 and 2019 included a comprehensive review of existing social survey and administrative data sources that are available as a basis for building up a national picture of the numbers of children that fall within each group and category of vulnerability and on needs and outcomes at the aggregate level. Based on the data sources available at the time, best estimates suggest that just prior to the pandemic an estimated 2.3 million children were living with risk because of their family background (for example, living in households

where there is domestic violence or substance abuse, or living in a household that is classified as being in absolute poverty). Of this total, an estimated 829,000 children were 'invisible' in the sense of not being known to services and therefore not getting any support and another 761,000 children, around a third, were known to services but with unclear levels of support. The remaining 669,000 children were being helped through a formal, national programme of support such as the Troubled Families programme and different forms of children's social care, with around 128,000 were identified as receiving most intensive statutory support such as being in care or on a child protection plan (CCO 2019).

Overall, the CCO child vulnerability assessments highlighted the extent of data gaps relating to the numbers, needs and outcomes of vulnerable children and the importance of a national statistics response to build up a more comprehensive national data infrastructure as a basis for social and public policy in this field. The Data for Children Partnership, a strategic partnership between ADR UK, the Office of the Children's Commissioner England and ONS, was established to address this challenge. The Growing Up in England (GUiE) dataset was taken forward within this broader context as a collaboration between the ONS, Department for Education and ADR UK.

2.2.4 UK Statistics Authority Inclusive Data Taskforce

The UK Statistics Authority (UKSA) Inclusive Data Taskforce was established in October 2020 to address how to make a step-change in the inclusivity of UK data and evidence following on from the publication of the UK Statistics strategy 'Statistics for the Public Good' (UKSA 2020) and in the wake of the COVID19. The Taskforce report, published in 2021, set out a series of recommendations with the goal of ensuring that "data and evidence across the UK is reflective and inclusive of all" so that "everyone in society counts and is counted" and "no one is left behind". Gypsy, Traveller and Roma were identified as one of the 'harder to reach' groups for whom better data infrastructure and evidence was required (UKSA 2021a).

Overall, the Taskforce findings highlighted that addressing evidence and data gaps and building up better and more inclusive data infrastructure is likely to involve a combination of different methods and strategies - both quantitative and qualitative. In relation to quantitative methodologies, potential techniques include the inclusion of relevant and improved identifiers in social survey, cohort and administrative data collections; over-sampling of target populations in social surveys; and the development of specialist parallel studies with alternative recruitment and retention

methodologies. However, qualitative and voice-based approaches, including structured and semi-structured interview-based methods as well as a range of different engagement, voice and listening initiatives designed to elicit broader and deeper information, experiences and perspectives, are all identified as playing an important role. Additionally, broader measures to maximise participation such as community engagement initiatives and initiatives to build up trust are recommended.

Specific recommendations to systematically build up an inclusive national data infrastructure included:

- Data producers should review the representativeness of key surveys and administrative datasets (initially by benchmarking against the UK 2021/22 Censuses) and take swift action to address identified issues, particularly as they relate to historically under-represented populations or more marginalised groups. Such a review should ensure that key surveys and datasets are using consistent measures and that inconsistencies are not distorting comparability.
- Data producers should ensure sufficient granularity of data to enable meaningful disaggregation. The use of meta-categories which disguise heterogeneity between and within groups as well as categories with which people may not identify should be avoided, with a move towards classifications that are recognisable and meaningful to those from specific ethnic groups. Ethnicity standards should be harmonised across datasets.
- Comparative studies should be undertaken across the UK to examine 'what works' in promoting inclusive data collection. Data producers should use targeted oversampling of under-represented groups as an approach to address specific gaps in knowledge that result from small sample sizes, as well as linking administrative data to survey data or other administrative sources, and improved use of longitudinal and panel data collection. Where data is linked, the linking process should be demonstrated to be of appropriate quality and accuracy and safeguards to protect respondent confidentiality and anonymity should be in place.
- In order to enhance understanding of the experiences of relevant groups and populations, data producers should also explore more varied, innovative and flexible approaches to data collection and combination, including the use of qualitative research, ethnographic research to understand lived experiences and voice-based approaches.
- For surveys currently collecting information about children, data producers should consider what information can be collected directly from children themselves, using appropriate instruments and diverse

forms of data collection (for example, pictures and diaries), drawing upon best practice in data collection and ethical approaches. Data should be collected to reflect more marginalised children (for example, disabled children, children of prisoners, Gypsy, Roma and Traveller children, looked after children, refugees and unaccompanied migrants) as a priority.

- There should be specific exercises to build up trust, including through appropriate guarantees of consent, confidentiality and anonymity. Additionally, UK data and evidence should be equally accessible to all, with tailored outputs where appropriate.

The publication of the Inclusive Taskforce report was accompanied by a statement from the National Statistician (UKSA 2021b). This re-iterated the importance of building up more inclusive data and evidence on groups that have previously been under-represented in UK social statistics. Additionally, it highlighted the broader context of COVID-19 and the importance of 'joined up' government polices - and of 'joined up data' to support this – going forward. It also highlighted the potential for transforming national data infrastructure brought about by rapid technological change and the expansion of 'big data', coupled with the new legal landscape created by the Digital Economy Act (2017) (ONS 2021a, UKSA 2021ab, Diamond 2021ab, UKSA 2023).

2.3 Looking forward: the potential of large-scale administrative data linkages

In the current study, we focus on the specific contribution that administrative data linkages can play within these broader efforts, as a basis for challenging data exclusion and building up new inclusive evidence on the lives, circumstances and needs of Gypsy, Traveller and Roma young people. In terms of terminology, the term *administrative data* is generally used to refer to statistical information that is routinely collected by government and public services and organisations such as schools, the NHS, the benefits system, the police and criminal justice system, as part of public administration (for example, by local, regional and central government, public services and organisations such as schools, health services, social care services, tax and benefit administration bodies, the police and criminal justice system). Administrative data *linkage* is the process of joining together or combining an administrative dataset with another dataset, for example to a social survey dataset, another administrative dataset or a population census. The process of data linkage

involves matching records across different datasets and hence assessing whether the records belong to the same entity (for example, the same person, household or firm) (ONS 2021a). Different methods for undertaking data linkage are available and some of these are discussed in this report. Internationally, administrative data linkages are increasingly undertaken so that the information and insights from different datasets can be combined and as we identify below, there are multiple general clear strengths of administrative data linkages from the point of view of data inclusivity.

At the current time, a substantial expansion in the administrative data landscape and of the availability of administrative datasets, including linked administrative datasets, for research purposes is ongoing. Following on from the 2011 Census, there was a stated high level strategic aim to transform national population statistics with more use of administrative data that is routinely collected for operational administrative purposes coupled with less reliance on social survey data (ONS 2013, 2019, 2020b). Additionally, rapid technological, computational and methodological advances are opening up new opportunities for the development of 'big' data infrastructure and analysis, including new methodologies for de-identification and large-scale data linkages, while the enactment of the Digital Economy Act 2017 provided a new legal framework for public authorities to share personal data for defined purposes for the public good while upholding individual rights to privacy (ONS 2021a, UKSA 2021ab, Diamond 2021ab, UKSA 2023).

Office for Statistics Regulation (n.d.a) identifies there are two high level opposing narratives relating to these developments. The first is a positive narrative around the untapped potential of administrative data to promote knowledge and understanding in new ways; the new research opportunities opened-up by technology and 'big data' and algorithms; and new opportunities for challenging data exclusion and building up new evidence on highly disadvantaged groups that were previously 'missing' from, or 'invisible within' national statistics and evidence. However, a second negative narrative focuses on critique, limitations and potential harms. Key concerns arise for example in relation to bureaucratic power and interests; the potential for political manipulation, including where the availability of administrative data is driven by performance targets or the need for high level ministerial sign-off of research datasets. Other potential limitations relate to the quality, reliability and validity of administrative data. For example, inconsistent and poor recording practices, with poor and variable quality underlying data recording systems resulting from fragmented policy silos, pressurised public sector workforces and limited resource. Broader

ethical concerns have been raised around privacy, consent, the potential for discrimination and other harms, including in relation to new technologies and predictive algorithms as well as bias and learning bias that could result when machine learning approaches are applied using imperfect data (Office for Statistics Regulation n.d.a; c.f. UKSA 2021ab; House of Commons Committee on Public Accounts Committee 2019). Moreover, there is an essential trade-off between maintaining privacy and data quality; arriving at an appropriate balance is not a purely statistical decision but raises fundamental issues of law, ethics and the public good (ONS 2021a).

In later sections of this report, we provide a detailed assessment of the strengths and limitations of the newly linked Growing Up in England dataset and the role it can play in challenging data exclusion and building up inclusive evidence on young people from the Gypsy, Traveller and Roma communities. Before turning to this exercise, we review some of the previous key specialist quantitative and qualitative studies that address educational disadvantage and the broader multidimensional inequalities experienced by Gypsy, Traveller and Roma young people. Additionally, we provide an overview of key issues, barriers, concerns and good practice identified in civil society organisation analysis and advocacy and the broader policy context of this study.

3. Data and qualitative evidence review: what do we already know about educational disadvantage and the broader multidimensional inequalities that impact on the Gypsy, Traveller and Roma communities?

This section provides additional background and context for our GUiE project by reviewing existing evidence on educational disadvantage and the broader multidimensional inequalities that impact on Gypsy, Traveller and Roma young people and their families. Building on the recommendations of the NSDEC Committee highlighted in the introduction of this report, we provide an overview of previous specialist Gypsy, Traveller and Roma research in this field (both quantitative and qualitative) and of what can be learnt in terms of key issues, barriers, good practice and policy solutions from civil society organisations that represent Gypsy, Traveller and Roma people. Additionally, we highlight relevant findings from recent high level inquires and reviews.

3.1 Evidence on Gypsy, Traveller and Roma educational disadvantage

The extent to which children and young people from the communities experience low educational outcomes across a wide range of educational indicators and across all stages of education, as well as poor transitions into training and employment, is a central and recurring theme in a small but discrete body of research evidence. This includes specialist studies (e.g. Cemlyn et al 2009, Mulcahy et al 2017, Greenfields 2019); findings reported in equality and human rights monitoring exercises (e.g. EHRC 2010, 2015, 2016, 2018); findings identified in civil society organisation analysis and advocacy (e.g. Traveller Movement 2017, 2019, 2020a, 2021, Friends, Families and Travellers 2022b); and findings included in reviews, high level inquiries and parliamentary committees (Race Disparities Audit 2017 – see Cabinet Office 2018; Women and Equalities Committee 2019; House of Commons Education Committee 2022ab).

3.1.1 Official published statistics

Table 1 provides an overview of official published statistics on educational outcomes by ethnicity. The indicators covered in the table include: school readiness; key stage 2 attainment; key stage 4 attainment (GCSE attainment 8); post-16 participation and sustained destinations

(education, employment or apprenticeships); A level attainment; permanent and fixed term exclusions; persistent absences; SEND; and highest educational attainment at age 19. From the available official data, it is evident that children and young people from the Gypsy, Traveller and Roma communities have worse outcomes than all other ethnic groups across most educational indicators and stages of education including early years development, primary and secondary school, and post-16 participation. Young people from the communities also experience the highest exclusion rates of any ethnic group, more persistent absence and higher prevalence of special educational needs and disabilities.

The analysis of educational outcomes in the current study focuses on highest educational attainment at age 19. This indicator evaluates qualifications attained through any route (school, college or training) and of any type (GCSE, A level or vocational or work-based qualifications) by Level, distinguishing between Level 1 qualifications (one to four GCSEs at any grade) or equivalent; Level 2 qualifications (five or more GCSES grades A* to C or equivalent; Level 3 qualifications (two or more A levels or equivalent); and apprenticeships. The published statistics on highest educational qualification at age 19 included in Table 1 show that just before the COVID-19 pandemic in 2018/19, amongst the general population, 81% of young people have secured Level 2 qualifications by age 19 and 57% had secured Level 3 qualifications. However, amongst Gypsy/Roma young people, rates of Level 2 attainment at age 19 stood at 23.9% and Level 3 at 9%, while for young Travellers of Irish heritage, rates were 24% and 14% respectively.

Note that low numbers can result in suppression of Gypsy, Traveller and Roma breakdowns in official educational statistics for some indicators for reasons of both disclosure risk (that is, to protect individual confidentiality and information rights) and statistical reliability (where numbers are deemed too low for reliable inferences to be drawn). In Table 1, particularly low numbers are evident in relation to the proportion of A level students attaining A and A* grades. Here, a flag highlights that low numbers of students have been in the identified 2018/19 academic year cohort for both the Gypsy/Roma and Traveller of Irish Heritage categories (30 and 9 respectively) and that the statistical reliability of the statistics reported may be limited as a result (with 0 percent identified as attaining 3 A or A* grades within both categories).

Statistics on higher education are *not* provided in Table 1. However, Mulcahy et al (2017) report that 3% to 4% of Gypsies, Roma and Travellers aged 18-30 accessed HE in 2014, whereas 43% of 18–30-year-olds in the national population did so. Figures cited by the House of Commons

Education Committee suggest marginally higher participation rates (see Figure 3).

Table 1: Overview of published statistics on educational outcomes that identify Gypsy, Traveller or Roma ethnic identities

Ethnicity	**Percentage of 4 to 5 year olds who achieved the early years development goals in 2018/19[1]**	**Percentage of pupils meeting standards in reading, writing and maths in year 6 when they were 10/11 years (2018/19 school year) [2]**		**GCSE and equivalent entries and achievements of pupils at the end of key stage 4, 2018/19 (revised)[3]**				**Permanent exclusions rate (all state-funded schools, England)[4]**			**Absences: overall absence rate and persistent absence rate by ethnicity (2020/21 school year)[5]**		**Total number of pupils SEN Support and Statement or EHC in England 2021/22[6]** (includes state-funded nursery, primary, secondary and special schools, non-maintained special schools and pupil referral units. Does not include independent schools or the small number of general hospital schools)		
				English and Maths GCSEs		Progress 8									
		Expected standard	Higher than expected	Number of pupils at the end of key stage 4	Percentage6 of pupils who achieved 9-5	Number of pupils included	Average Progress 8 score	2009/10	2014/15	2018/19	Percentage	School sessions	Total	SEN Support	Statement or EHC
All	71	65	11	542,56	42.	512,91	-	0.1	0.	0.	4.6	117,183,52	_	_	_
Asian	69	69	13	58,111	51.	52,766	0.47	-	-	-	4.7	13,139,879	_	_	_
Bangladeshi	67	70	13	9,877	50.	9,289	0.47	0.1	0.	0.	5	2,138,679	151,613	15,125	6,242
Indian	78	77	19	15,429	64.	13,508	0.71	0	0	0	3.2	2,611,253	289,179	18,084	6,850
Pakistani	64	62	9	23,502	41.	22,020	0.24	0.1	0.	0.	6	6,606,468	380,781	41,762	13,881
Asian other	69	71	15	9,303	60.	7,949	0.66	0.1	0.	0.	3.8	1,783,479	170,572	13,238	5,952
Black	68	64	9	31,175	37.	27,191	0.13	-	-	-	3.8	5,323,173			
Black African	68	67	10	19,807	42.	16,861	0.33	0.2	0.	0.	3.3	3,142,694	338,614	35,604	14,967
Black Caribbean	68	56	6	7,493	26.	7,146	-	0.6	0.	0.	5.4	1,378,936	82,402	13,277	4,444

Ethnicity	Percentage of 4 to 5 year olds who achieved the early years development goals in 2018/19[1]	Percentage of pupils meeting standards in reading, writing and maths in year 6 when they were 10/11 years (2018/19 school year) [2]		GCSE and equivalent entries and achievements of pupils at the end of key stage 4, 2018/19 (revised)[3]				Permanent exclusions rate (all state-funded schools, England)[4]			Absences: overall absence rate and persistent absence rate by ethnicity (2020/21 school year)[5]		Total number of pupils SEN Support and Statement or EHC in England 2021/22[6] (includes state-funded nursery, primary, secondary and special schools, non-maintained special schools and pupil referral units. Does not include independent schools or the small number of general hospital schools)		
				English and Maths GCSEs		Progress 8									
		Expected standard	Higher than expected	Number of pupils at the end of key stage 4	Percentage6 of pupils who achieved 9-5	Number of pupils included	Average Progress 8 score	2009/10	2014/15	2018/19	Percentage	School sessions	Total	SEN Support	Statement or EHC
Black other	66	60	7	3,875	33.	3,184	0.08	0.5	0.	0.	4.2	801,543	65,501	8,085	3,387
Chinese	76	80	28	2,006	76.	1,697	0.86	0	0	0	2.1	238,438	48,396	2,422	1,100
Mixed	72	68	12	27,018	43.	25,175	0	-	-	-	5	7,949,675	-	-	-
Mixed White/Asian	75	64	17	6,144	55.	5,734	0.22	0.1	0.	0.	4.3	1,641,691	134,821	12,985	4,158
Mixed White/Black African	71	65	11	3,262	41.	2,968	0.04	0.3	0.	0.	4.7	1,036,764	76,262	9,205	3,159
Mixed White/Black	69	73	7	7,504	31	7,322	-	0.4	0.	0.	6.3	2,506,216	133,504	21,603	6,490
Mixed other	73	19	14	10,108	47	9,151	0.14	0.2	0.	0.	4.6	2,765,004	213,597	23,601	8,534
White	72	26	10	406,70	42.	391,72	-	-	-	-	4.6	85,728,791	-	-	-
White British	73	63	10	375,41	42.	368,32	-	0.1	0.	0.	4.6	76,695,613	5,379,74	730,65	220,68
White Irish	74	73	16	1,688	54.	1,561	0.13	0.2	0.	0.	5.5	352,601	21,495	2,851	901
Gypsy/Roma	34	19	1	1,243	6	896	-	0.9	1.	1.	15	1,225,310	27,359	6,040	1,258

Ethnicity	Percentage of 4 to 5 year olds who achieved the early years development goals in 2018/19[1]	Percentage of pupils meeting standards in reading, writing and maths in year 6 when they were 10/11 years (2018/19 school year) [2]		GCSE and equivalent entries and achievements of pupils at the end of key stage 4, 2018/19 (revised)[3]				Permanent exclusions rate (all state-funded schools, England)[4]			Absences: overall absence rate and persistent absence rate by ethnicity (2020/21 school year)[5]		Total number of pupils SEN Support and Statement or EHC in England 2021/22[6] (includes state-funded nursery, primary, secondary and special schools, non-maintained special schools and pupil referral units. Does not include independent schools or the small number of general hospital schools)		
				English and Maths GCSEs		Progress 8									
		Expected standard	Higher than expected	Number of pupils at the end of key stage 4	Percentage6 of pupils who achieved 9-5	Number of pupils included	Average Progress 8 score	2009/10	2014/15	2018/19	Percentage	School sessions	Total	SEN Support	Statement or EHC
Traveller of Irish Heritage	39	26	1	151	13.	141	-	1.2	1.	0.	19.	361,920	6,903	1,675	393
White other	66	63	11	28,214	41.	20,803	0.43	0.1	0.	0.	4.2	7,093,347	575,540	54,250	15,944
Other	63	61	10	9,607	43.	7,442	0.5	0.1	0.	0.	4.6	2,352,976	185,355	18,440	5,991
Unknown	-	49	8	-	-	-	-	-	-	-	-	-	-	-	-
Unclassified				7,943	35.	6,913	-	-	-	-	-	-	136,371	15,397	6,107

Source

1. Data source: Development goals for 4 to 5 year olds - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)
2. Data source: Reading, writing and maths results for 10 to 11 year olds - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)
3. Data source: Education Data Division, Department for Education, GCSE and equivalent entries and achievements of pupils at the end of key stage 4 by pupil characteristics 2019_KS4_revised_national_characteristics_tables.xlsx (live.com)
4. CASE SPDO data and charts SPDO indicator exercise indicators.xlsx (live.com)

5. Data source: Absence from school - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)
6. Data source: Special educational needs by ethnicity Special educational needs in England, Academic year 2021/22 – Explore education statistics – GOV.UK (explore-education-statistics.service.gov.uk)

	Destinations of those who completed KS4 in academic year 2016/17 after two terms (in 2017/18)[1]					Key stage 4 national level destinations data (England, **2016/17**, state funded mainstream schools - percentages)[2]				Percentage of pupils going into sustained education, apprenticeships or employment after key stage 4 (**2017/18**)[3]						Key stage 4 destinations data (England, **2018/19**, state funded mainstream schools - percentages)[4]				Average point score and equivalent grade by ethnicity (A level students only) 2020/21[5]		Students getting 3 A grades or better at A Level (**2018/19** academic year)[6]		Level 2 and 3 attainment at age 19 by ethnicity (academic year **2018/19**)[7]	
Ethnicity	Education	Apprenticeships	Employment	No sustained education/employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Education, apprenticeships, or employment	Education	Apprenticeships	Employment	No sustained education/ employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Average point score	Equivalent grade	%	Students in cohort	Percentage who attained Level 2	Percentage who attained Level 3
All	87	4	3	5	1					94	87	4	3	5	1					41.6	B	13	218,398	81.8	57.1
Asian										96	94	1	1	3	1					39.95	B	11	24,942	86.2	68.7
Bangladeshi						93	2	1	4	96	95	1	1	3	1	95	1	1	4	38.64	B	7.8	4,194	86.6	68.8
Indian						96	1	1	2	97	96	1	1	2	1	96	1	1	2	42.79	B+	15.5	8,222	91.8	77.6
Pakistani						92	1	1	4	94	92	1	1	4	1	92	1	1	4	37.33	B-	7.3	7,788	81.1	61.1
Asian other						94	1	1	3	96	95	0	1	3	1	95		1	3	40.46	B	11.8	4,738		
Black										95	93	1	1	4	1	92	1	1	4	37.06	B-	5.5	10,475		

Ethnicity	Destinations of those who completed KS4 in academic year 2016/17 after two terms (in 2017/18)[1]					Key stage 4 national level destinations data (England, **2016/17**, state funded mainstream schools - percentages)[2]				Percentage of pupils going into sustained education, apprenticeships or employment after key stage 4 (**2017/18**)[3]						Key stage 4 destinations data (England, **2018/19**, state funded mainstream schools - percentages)[4]				Average point score and equivalent grade by ethnicity (A level students only) 2020/21[5]		Students getting 3 A grades or better at A Level (**2018/19** academic year)[6]		Level 2 and 3 attainment at age 19 by ethnicity (academic year **2018/19**)[7]	
	Education	Apprenticeships	Employment	No sustained education/employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Education, apprenticeships, or employment	Education	Apprenticeships	Employment	No sustained education/ employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Average point score	Equivalent grade	%	Students in cohort	Percentage who attained Level 2	Percentage who attained Level 3
Black African						93	1	1	3	96	94	1	1	3	1	94	1	1	3	37.32	B-	6.1	7,520	86.7	69.3
Black Caribbean						89	2	2	6	93	89	2	2	6	1	90	1	2	6	35.97	B-	3.4	1,839	78.1	54.3
Black other						90	2	1	5	93	90	2	1	5	1	92	1	1	4	36.79	B-	5.4	1,116		
Chinese						98			1	98	97	1	0	1	1	98			1	45.77	A-	25.7	1,480	95.2	64.2
Mixed										93	88	2	2	6	1					40.57	B	11.2	8,982		
Mixed White/Asian						90	3	2	5	94	90	2	3	5	1	90	2	2	4	42.35	B+	15.3	2,511	86	66.9
Mixed White/Black African						89	3	2	5	93	89	2	2	6	71	89	2	2	6	39.29	B	8.3	1,080	83.4	60.8

Ethnicity	Destinations of those who completed KS4 in academic year 2016/17 after two terms (in 2017/18)[1]					Key stage 4 national level destinations data (England, **2016/17**, state funded mainstream schools - percentages)[2]				Percentage of pupils going into sustained education, apprenticeships or employment after key stage 4 (**2017/18**)[3]						Key stage 4 destinations data (England, **2018/19**, state funded mainstream schools - percentages)[4]				Average point score and equivalent grade by ethnicity (A level students only) 2020/21[5]		Students getting 3 A grades or better at A Level (**2018/19** academic year)[6]		Level 2 and 3 attainment at age 19 by ethnicity (academic year **2018/19**)[7]	
	Education	Apprenticeships	Employment	No sustained education/employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Education, apprenticeships, or employment	Education	Apprenticeships	Employment	No sustained education/ employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Average point score	Equivalent grade	%	Students in cohort	Percentage who attained Level 2	Percentage who attained Level 3
Mixed White/Black Caribbean						83	4	3	8	91	84	3	3	8	1	84	3	4	8	38.33	B-	6.2	1,812	76.2	49.7
Mixed other						89	2	2	5	94	90	2	2	5	1	90	2	2	5	40.63	B	11.8	3,579		
White										94	85	5	4	5	1					40.86	B	11	131,947	81.1	54.9
White British						85	6	4	5	94	85	6	4	5	1	85	5	4	5	40.82	B	11	121,533	81.3	54.7
White Irish						88	4	2	5	94	88	3	3	5	1	90	3	2	4	42.39	B+	13.8	816	85.1	66.5
Gypsy/Roma	56	2	8	23	10	56	3	7	24	67	56	2	8	23	10	54	2	10	25	31.98	C+	0	30	23.9	9
Traveller of Irish Heritage	58	10	5	22	6	55	8	11	20	72	58	10	5	22	6	58	2	5	25	38.18	B-	0	9	41.8	13.9
White other						89	2	2	5	92	89	1	2	5	3	89	1	2	5	41.18	B	11.5	9,559	80.7	58.3

Ethnicity	Destinations of those who completed KS4 in academic year 2016/17 after two terms (in 2017/18)[1]					Key stage 4 national level destinations data (England, **2016/17**, state funded mainstream schools - percentages)[2]				Percentage of pupils going into sustained education, apprenticeships or employment after key stage 4 (**2017/18**)[3]						Key stage 4 destinations data (England, **2018/19**, state funded mainstream schools - percentages)[4]				Average point score and equivalent grade by ethnicity (A level students only) 2020/21[5]		Students getting 3 A grades or better at A Level (**2018/19** academic year)[6]		Level 2 and 3 attainment at age 19 by ethnicity (academic year **2018/19**)[7]	
	Education	Apprenticeships	Employment	No sustained education/employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Education, apprenticeships, or employment	Education	Apprenticeships	Employment	No sustained education/ employment	Unknown	Sustained education destination	Sustained apprenticeships	Sustained employment destination	Not recorded as a sustained destination	Average point score	Equivalent grade	%	Students in cohort	Percentage who attained Level 2	Percentage who attained Level 3
Other						91	1	1	4	93	92	1	1	4	2	92	1	1	4	39.76	B	10.2	3,526		
Unknown										91	83	4	4	7	2					46.83	A-	23.4	37,046	79.8	55.7
Unclassified						83	4	4	8							83	4	4	8						

Sources:

1. Data source**:** https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/after-education/destinations-of-school-pupils-after-key-stage-4-usually-aged-16-years/latest.
2. Data source: downloaded from https://explore-education-statistics.service.gov.uk/data-tables/fast-track/e2bcdfef-7c24-4bc4-9971-51b1f3dc752f
   Note: State-funded mainstream schools include community schools, voluntary aided schools, voluntary controlled schools, foundation schools, academies, free schools, city technology colleges and further education colleges with provision for 14- to 16-year-olds. Includes pupils for whom ethnicity or first language was not obtained, refused or could not be determined.
3. Data source: Destinations of school pupils after GCSEs (and equivalent qualifications) - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)
   Note: The data shows the destinations of students who finished their GCSEs and other 14 to 16 study in state-funded mainstream schools in England.

4. Data source :https://explore-education-statistics.service.gov.uk/data-tables/fast-track/e2bcdfef-7c24-4bc4-9971-51b1f3dc752f
   State-funded mainstream schools include community schools, voluntary aided schools, voluntary controlled schools, foundation schools, academies, free schools, city technology colleges and further education colleges with provision for 14- to 16-year-olds. Includes pupils for whom ethnicity or first language was not obtained, refused or could not be determined.
5. Data source: Average score for students taking A levels and other qualifications - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)
6. Data source: Students getting 3 A grades or better at A level - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)
   Note: the reliability of these figures for the White Gypsy and Roma and Traveller of Irish Heritage ethnic groups are based on a small number of students and this affects the reliability of generalisations.
7. Source gov.uk Level 2 and 3 attainment age 16 to 25 https://explore-education-statistics.service.gov.uk/data-catalogue/level-2-and-3-attainment-by-young-people-aged-19/2019-20

3.1.2 Key studies on educational disadvantage

Research commissioned by the Equality and Human Rights Commission (Cemlyn et al 2009) reviewed official data and quantitative studies as well as studies that used qualitative methodologies including interviews, case studies and practitioner research to explore the educational experiences of Gypsy and Traveller young people, the views of parents and the attitudes, practices and strategies of teachers and other staff. This study identified key common themes in relation to the educational disadvantages experienced by Gypsy and Traveller young people including: enforced mobility and interrupted learning; consistent experiences of racist harassment and bullying; excessive exclusions linked to these experiences and to inadequate school responses; lack of validation of Gypsy and Traveller culture in schools; limited relevance of the curriculum for some pupils; cultural barriers that children have to negotiate between home and school; teachers' low expectations; and the impact of national targets on schools' readiness to admit children and young people from the communities. Issues of inclusion, recognition and identity have also been examined in recent literature (e.g. Greenfields 2019) as well as experiences of 'outsider' status amongst those from the communities who do progress and participate in higher education (Morgan et al 2023).

Mulcahy et al (2017) identified multiple barriers to Gypsy, Traveller and Roma progression through primary, secondary and 16-18 education into higher education. These include discrimination; cultural barriers (mobility; language and system knowledge; norms, aspirations and expectations and cultural identity); material barriers (poverty, inadequate housing and homelessness, lack of access to healthcare and special educational needs support); curriculum issues, bullying and self-exclusion from mainstream schooling; and specific barriers to accessing higher education (including perceptions of relevance and lack of role models as well as attitudes to finance and debt, with debt aversion and reluctance to take out loans playing a role). The Traveller Movement (2021) highlighted reluctance to take out loans including the Advanced Learner Loan as a barrier to participation in 19+ further education and lifelong learning.

Several studies report that Gypsy, Traveller and Roma children and young people are more likely to drop out of school and to not progress to 16-18, further or higher education. Wilkin, Derrington and Foster (2009) and Greenfields et al (2022) review the evidence on high drop-out rates in transitions from primary to secondary school. Both cite a study undertaken by Derrington and Kendall which followed forty-four 11 year olds longitudinally and found that only a third of the sample were still in school

at age 16 (the statutory leaving age at the time). Boys were found to be twice as likely to drop out as girls and year eight was identified as a particular window for declining participation (Wilkin, Derrington and Foster 2009; c.f. Derrington and Kendall 2004, 2007). Wilkin, Derrington and (2009) report that their findings of declining participation rates with age, particularly during secondary stages, are supported by broader data, noting that only a third of the number of Gypsy/Roma pupils in Key Stage 4 compared to Key Stage 1 in England; and less than a half of Travellers of Irish Heritage are recorded at Key Stage 4 compared to Key Stage 1. They also cite evidence from a Scottish study (published in 2000) that reported that many Traveller children had dropped out of school completely by the age of 14.

Greenfields et al (2022) cites evidence (published in 2014) of a steep decline in attendance at secondary schools in the Welsh context. Citing a more recent study, Greenfields et al (2022) notes that 61% of the Traveller Movement's (2020) sample of 44 young Irish Travellers aged between 15 and 25 dropped out of education before the statutory leaving age. Recent research by the Traveller Movement highlights year 11 as a particular trigger point, with higher numbers of Gypsy, Roma and Traveller pupils falling out of education before sitting their exams, with the start of the examination year being a key fall out point (Traveller Movement 2021). Recent research focussing on Gypsy, Traveller and Roma participation in further and higher education in the West Yorkshire context identified preferences for 'staying local' in the context of oversubscribed courses as a key barrier. Young people were generally identified as not being aware of what apprenticeships were and to lack information on apprenticeship routes and other practical-based courses. Key recommendations include the provision of tailored careers information and guidance and ensuring that apprenticeship providers engage with the Gypsy, Traveller and Roma communities (Greenfields et al 2022).

3.1.3 Key issues, barriers and concerns highlighted in civil society analysis and advocacy

Civil society representative organisations regularly raise concerns about the relatively high numbers of Gypsy, Traveller and Roma young people dropping out of school, being excluded or off-rolled and pursuing elective home education. Research by the Traveller Movement published in 2017 found that 70% (139 out of 199) of respondents to a survey reported experiencing discrimination in education (Traveller Movement 2017). Qualitative research amongst young Travellers in London published in 2019

highlighted concerns that children from the communities were leaving school after primary or secondary school and that unaddressed bullying was a major factor. This research was led by the Traveller Movement with input from the Advisory Council for the Education of Romany and other Travellers and Southwark Traveller's Action Group. Reporting on the experiences of forty-four 15- to 25-year-old Travellers in London, interviews were conducted by young peer reviewers who had been trained in qualitative research techniques. The research found that 61% of interviewees had left school straight after primary or during primary school. Bullying was identified as a major contributory factor in early school leaving, with 45% of girls and 29% of boys reporting bullying as a reason for leaving school. Of these, 40% mentioned bullying by other pupils and 67% said they had been bullied by teachers because of their Traveller identity. Other reasons for leaving school that were cited included lack of support from their school including in relation to special educational needs and disabilities (5%) and the family not being settled (4%) (Traveller Movement 2019a).

The Traveller Movement (2019b) identified key issues, solutions and good practice based on depth analysis of its educational casework and advocacy. This research highlighted seven key case studies where children and their families had been assisted by the Traveller Movement advocacy team in the context of education related matters. Presenting issues included: racist bullying; unmet special educational needs (SEN); exclusions; admissions; transport; discrimination; attendance; and elective home education. It found that discrimination affects a majority of families assisted; that school exclusions are often given as a first rather than a last resort and are frequently given to children with SEN; that most schools lack the resources to support children with SEN; that funding for school uniforms causes many families distress and financial hardship; and that racist bullying is rarely addressed appropriately. The research also highlighted several examples of good practice including where schools tackled bullying proactively, ensuring that children from the communities are not disproportionately targeted; taking pro-active measures to build up good relationships with children and parents from the communities; including Gypsy, Roma and Traveller history and culture in curricula; and having Gypsy, Traveller or Roma teaching assistants, interns and support staff than can have a role as community links and more generally in building up trust.

The Traveller Movement (2020) examines Gypsy, Roma and Traveller experiences in secondary education, identifying barriers and obstacles to secondary educational participation and highlighting solutions and good

practice, drawing on the Traveller Movement education casework and projects. Concerns were identified and solutions and recommendations proposed in relation to four key areas.

- **Bullying.** Ethnicity related bullying that largely goes unaddressed in schools was again reported as a key concern. This was found to damage educational experiences, negatively impacting on pupils' sense of belonging in schools and resulted in children taking matters into their own hands in some instances. Recommendations include inclusion of Gypsy, Roma and Traveller histories and cultures within the school curriculum; strong anti-racist bullying policies; the adoption of the term anti-Gypsyism to better define and understand the specific racism faced by Gypsy, Roma and Traveller people; and pro-active measures to promote re-engagement where gaps in schooling were associated with unaddressed bullying.
- **Punitive school behavioural policies that disproportionately impact on children from the communities.** Punitive school policies related to zero tolerance and the use of fixed term and permanent exclusions were identified as disproportionately impacting on children from the communities. Examples drawn from casework included exclusions being imposed in the context of relatively minor issues; exclusions being imposed on children who experience special educational needs and disabilities and require support; and lack of recognition of the circumstances of Gypsy, Traveller and Roma children and their families. Traveller children were also reported as being removed from school rolls or being put on a part-time timetable where there was a preference for additional support. Recommendations include less routine use of punitive behaviour policies, more use of restorative justice and empathy approaches and more of a focus on cultural awareness and sensitivity.
- **Re-engagement pathways.** Another concern related to a lack of clear and accessible re-engagement pathways for children who have missed periods of school or have left school early. This includes re-engagement pathways for pupils who do not have formal qualifications due to early school leaving. A particular gap was identified in educational services for pupils aged 14-16 who find that traditional academic qualifications are not a suitable fit, with limited opportunities for vocational qualifications for this age group, increasing the risk of drop out. Recommendations include funding being set aside for 14-16 alternative provision where gaps in services exist. Additionally, there should be accessible and designated routes back into education for those with little or no qualifications, including opportunities for post16 re-engagement with the further education system, following a sustained

period of absence or breaks in schooling. Casework analysis highlights that pro-active re-engagement measures can be successful in the context of school absences relating to bullying and to promote attendance and re-engagement in relation to curriculum concerns (including parental concerns in the context of relationship and sex education).

- **Access to information on elective home education.** A further concern was that parents had not been given the correct information and guidance from schools and local authorities when it comes to their educational rights and responsibilities, particularly around elective home education, and that this information is not always easily accessible to those with lower levels of literacy.

The Traveller Movement (2021) focuses on access to further education, vocational qualifications and apprenticeships. The importance of vocational options for young people aged 14-16 was again a key theme, with funded, discrete and meaningful alternative and vocational pathways at ages 14-16 identified as providing substantial potential for re-engagement with young people who have fallen out of secondary school or are at risk of doing so. Funding for pupils aged 14-16 to pursue vocational qualifications was highlighted as limited and discretionary and a key recommendation was that funding should be more consistent and substantially increased. A further recommendation was that pupil entitlement funding should be used more flexibility to promote re-engagement, with increased consideration of alternative routes back into education including further education, college and vocational options from an earlier age, where this maximises participation.

Additionally, the Traveller Movement (2021) highlights the importance of further and vocational education and training for young people from the communities at ages 16+, including as a re-engagement route for those with lower educational qualifications and with substantial gaps in their educational experience. The report identifies that some Gypsy, Traveller and Roma young people are prevented from participating in apprenticeships due to formal qualification requirements including maths and English requirements. Other barriers include loan schemes for those aged 19-25 which can be off-putting for Gypsy, Roma and Traveller people returning to education as an adult. Specific recommendations include the extension of post-16 GCSE maths and English 'catch up' windows from one to two years where appropriate, as one year is inadequate for those with substantial educational gaps. A clear and distinct pathway for those with vocational qualifications but not English or maths should also be introduced

and there is an urgent need for more opportunities to acquire maths and English qualifications as part of work and at college. There should also be specific targeting of Gypsy, Traveller and Roma young people in relation to traineeships and apprenticeships (Traveller Movement 2021).

In relation to A levels, specific recommendations in the Traveller Movement (2021) include: adequate support at GCSE level to ensure that Gypsy, Traveller and Roma young people are able to obtain the necessary grades to progress to A Level qualifications; additional funding to provide tutoring or mentoring for young people from Gypsy, Roma and Traveller backgrounds, especially in the context of missed education due to the COVID-19 pandemic; addressing low teacher expectations and raising the aspirations of Gypsy, Roma and Traveller young people in relation to A Levels as a valid educational pathway; and importance of mentoring, tutoring and support in the context of A level participation and attainment. There is in addition a need for post-16 support for individuals who experience special educational needs and disabilities across all learning pathways including in school, college and training settings (the Travellers Movement 2021).

More broadly, good practice guidelines for promoting an inclusive teaching and learning environment for Gypsy, Traveller and Roma young people have been developed by Friends, Families and Travellers. This includes materials and resources setting out how Gypsy, Traveller and Roma culture and history can be recognised and affirmed within the educational system and the educational curriculum and tools and information to tackle Gypsy, Traveller and Roma targeted bullying (FFT n.d.b). Guidelines, a school audit tool and an action plan developed by FFT and the anti-bullying alliance are available here and an associated learning module is available here. Other recent initiatives highlight the importance of a high level of commitment to Gypsy, Traveller and Roma inclusion via the 'pledge' for schools and universities. In the higher education context, good practice in widening participation and the importance of adopting a whole university approach via a 'the pledge' is discussed in Greenfields 2019 and Smith et al 2021. Other examples of good practice are cited in a review and research report for Department for Education (Wilkin et al 2009, Wilkin et al 2010) and on the Department for Education website (Department for Education 2014).

Key concerns have been raised in relation to disproportionate pandemic effects on children and young people from the communities and the growing impact of digital exclusion. Friends, Families and Travellers (2022b) reported that educational inequalities were exacerbated by COVID-19 and the pandemic also created new needs with the rise of digital exclusion.

Many Gypsy, Traveller and Roma families experienced huge challenges in accessing online schooling, particularly larger family groupings and extended families with smaller living spaces and those with limited literacy, resulting in disengagement from schooling altogether in some instances. While Government provided some funding to promote the educational outcomes of Gypsy, Traveller and Roma children, this was insufficient and only reached a limited number of schools and pupils. Vital funding to local authority Traveller Education Services, which were set up to improve relationships with the communities, address exclusion rates and attainment gaps, has also been cut; and most Traveller Education Services no longer exist.

3.2 Evidence on broader multidimensional inequalities

A small but discrete body of specialist literature provides evidence of the broader multidimensional inequalities that impact on Gypsy, Traveller and Roma young people in critical areas of life such as living standards, employment, accommodation, health and social care, physical security (including hate crime), legal security (including discrimination) and in relation to aspects of individual and social life, such as repeated experiences of hostile treatment and lack of treatment with dignity and respect. Below, we provide a brief overview of some of key recent studies in this field, including research commissioned and / or undertaken by civil society organisations and bodies such as the Equality and Human Rights Commission, the Office of the Children's Commissioner for England and the Office for National Statistics (ONS).

3.2.1 Key specialist studies

The key study commissioned by the Equality and Human Rights Commission cited above (Cemlyn et al 2009) provided an in-depth review of the available official statistics and existing specialist research on the inequalities experienced by Gypsy, Traveller and Roma people across different critical areas of life. This provided extensive evidence of poor outcomes and adverse experiences across multiple domains including in relation to economic inclusion and access to employment; lack of housing security, including lack of secure accommodation and sites, as well as poor quality sites and hazardous environments and the impact of cycles of evictions; poor experiences of accessing healthcare, education and social care; high rates of hate crime; adverse experiences of the legal and criminal justice systems and policing; and low rates of political

participation. The review also reported on the pervasive, corrosive and cumulative impact of repeated everyday experiences of racism and discrimination, hostile treatment and poor interactions with public services across entire lifespans and across generations, including in relation to high rates of anxiety, depression and self-destructive behaviour such as suicide and substance abuse.

Research by the Traveller Movement published in 2017 highlighted pervasive experiences of prejudice, racism and abuse in the everyday lives of Gypsy, Traveller and Roma people including experiences of stereotypes, prejudice, bullying, verbal abuse, racism and discrimination. These experiences occurred repeatedly in the context of interactions with the public, police and public services (including schools and hospitals) and in the context of work and employment. In the education context, hostile treatment and bullying was found to often go unaddressed and to include adverse treatment by teachers and staff as well as other pupils and families. Adverse experiences in education were found to have lifelong effects and to interact with negative experiences of interactions with authorities and other public services, resulting in a sense of hopelessness and a sense that, when help is needed, services will refuse help or no action will be taken (Traveller Movement 2017).

In-depth evidence on the health inequalities experienced by Gypsies and Travellers is reported in Parry et al (2007). Research by Friends, Families and Travellers identified that 74 out of 100 GP surgeries had broken NHS England guidance by refusing to register a nomadic patient between March and April 2021. The most common reasons for refused registration were inability to provide proof of identity, fixed address or to register online – representing noncompliance with NHS England guidelines stating that "there is no regulatory requirement to prove identity, address, immigration status or the provision of an NHS number in order to register" and that "[d]elivery of applications for patient registration may be by any means, including post and digital" (Friends, Families and Travellers 2021).

Civil society organisation advocacy and analysis undertaken in 2022 reports on recent progress and retrogression in relation to accommodation, health and educational inequalities, hate crime and hate speech (including the role of the media) and equal access to the European Union Settlement Scheme for Roma people (Friends, Families and Travellers 2022b). Accommodation is a key area where the situation was reported to have deteriorated, with a dramatic reduction of assessed need for more pitches following the introduction of restrictive criteria which excluded those unable to travel, including the elderly, those with a disability and those caring for them. Additionally, the situation in relation to GP registrations was

assessed as having deteriorated as digital by default registration gateways become more widespread in the wake of the pandemic. Educational inequalities were also reported to have been exacerbated by COVID-19, which created new needs with the rise of digital exclusion. Many Gypsy, Traveller and Roma families experienced huge challenges in accessing online schooling, particularly larger family groupings and extended families with smaller living spaces and those with limited literacy, resulting in disengagement from schooling altogether in some instances. Major concerns have also recently been identified in relation to the criminalisation of nomadic Gypsies and Travellers in the context of the Policing Act which was enacted in June 2022. Civil society organisations raised major objections that people who live on roadsides would be subject to fines, imprisonment and their homes being taken away (Friends, Families and Travellers 2022c).

The Children's Commissioner for England (2021, 2022) conducted the Big Ask survey, a national 'listening' exercise that engaged with around half a million children aged 4-17. Proactive measures were adopted to ensure the participation of children from different vulnerable and disadvantaged groups and responses included 2300 children and young people who self-identified as being from a Gypsy or Irish Traveller background. The reported results suggest that whilst the majority of children from the communities reported being 'happy' or 'very happy' with different dimensions of their life, on average they reported lower levels of happiness than other children, including in relation to 'life at school or college' (42% compared to other 56% of other children), mental health (53% compared to 58% for children from Black and Minority Ethnic groups and 56% of White British children) and family life (70% happy compared to 81% of White British and 78% of children from Black and Minority Ethnic groups), with greater valuation of the importance of having a family of their own in the future (36% compared to 26% in the overall sample). Analysis of open responses highlighted the importance of practical options lessons. The findings have fed into broader workstreams being taken forward by the Children's Commissioner for England including advocacy work around the integration of more practical options into the school curriculum and the adoption of targeted interventions for children who are struggling to attend school and who experience bullying or mental health needs, to maximise their participation.

3.2.2 The ONS lived experience study

An innovative qualitative research exercise was undertaken by ONS to follow up and implement the recommendations of the UK Statistics

Authority Inclusive Data Taskforce (UKSA 2021ab). ONS commissioned Derbyshire Gypsy Liaison Group to collaborate on the project examining the experiences, priorities and needs of Gypsy and Traveller communities in England and Wales. Over 50 community members and 30 central and local government officials from England and Wales participated in either in-depth interviews, focus group discussions, or both. These were analysed to identify themes and patterns relating to people's life experiences, focusing on areas such as homes, education, health, and justice. An open-ended, life history approach was adopted for interviews, to ensure they were participant-led, with follow up questions on specific issues.

- Participants described experiencing a range of health conditions, which, coupled with delayed healthcare seeking, and barriers to accessing healthcare, could create vulnerability to negative health outcomes among Gypsies and Travellers: environmental factors, such as site locations and standards, and occupational hazards; challenges registering with a GP without a fixed address; perceived discrimination and derogatory attitudes.
- Participants shared varied histories of access to education and educational attainment. This ranged from some who had never been to school, to others who had completed compulsory education, or gained higher level qualifications. While some spoke of enjoying their education, others described having faced numerous challenges, including perceived discriminatory behaviour from other students and teachers. Education and skills development were valued by participants, particularly for children, and were seen as important to employment prospects. Participants also cited perceived discrimination, inflexibility of the education system, and aspects of the curriculum seen as contradictory to Gypsy and Traveller values as reasons for withdrawing their children from mainstream education.
- Participants described barriers to employment, including a lack of skills, education or formal qualifications, and perceived discrimination from employers, colleagues, and the settled community. They spoke of facing difficulties in re-skilling from traditional occupations to new types of work, including a lack of technical skills. The introduction of new licensing requirements, for example for selling scrap metal, could also make continuing employment in occupations traditionally common among Gypsies and Travellers more difficult.
- Adverse experiences of the justice system, fear of authorities, sense of being unfairly treated and discriminated against, perceived disproportionality and a sense of injustice threaded through participants' accounts of experiences with the police, including in the

described use of force, presumption of criminality and frequency of arrests, denial of bail and imprisonment, unfair laws which criminalised Gypsy and Traveller people and their ways of life (regulations around collecting scrap metal). Some positive experiences of interactions with police were also shared.

- The nature of past experiences affected their decisions about whether to share or avoid sharing their ethnicity with others. People also described situations in which they felt "outed" because their ethnicity became known to others in ways over which they had no control.

Source: ONS (2022ab)

3.2.3 Our previous study on multidimensional child disadvantage

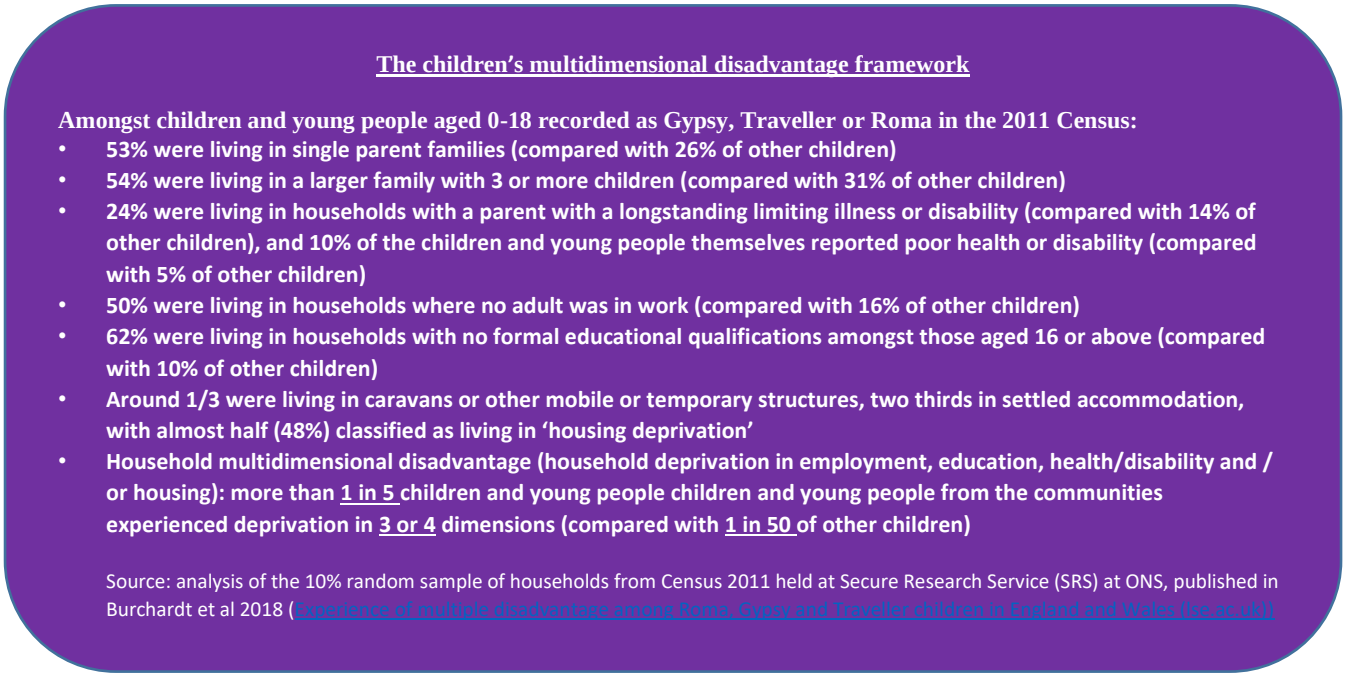
The current research project builds on our previous study that developed a multidimensional disadvantage framework for understanding the lives, circumstances and needs of children and young people from the Gypsy, Traveller and Roma communities using the 2011 Census (10% household sample secure microdata) (Vizard et al 2018, Burchardt et al 2018). The child multidimensional disadvantage framework assesses disadvantage across four dimensions - housing, household economic activity, education and health - and computes a multiple deprivation count.

The study identified that children and young people from the communities aged 0-18 were more likely than other children and young people of the same age to be living in single parent families (53% compared with 26% of other children); in families with three or more children (54% compared with 31% of other children); to be living with a parent who experiences a longstanding limiting illness or disability (24% compared with 14% of other children), or to report poor health or disability themselves (10% compared with 5% of other children); or to be living in a were living in household where no adult was in work (50% compared with 16% of other children). Around 1/3 of children and young people aged 0-18 from the communities were found to be living in caravans or other mobile or temporary structures, with two thirds in settled accommodation, and almost half (48%) classified as living in 'housing deprivation'.

Additionally, our previous study revealed that as many as 62% of children and young people from the Gypsy, Traveller and Roma communities experience household educational deprivation - where a young person has no educational resources within the household to draw on in the sense that no person aged 16 or above in the household has a formal educational qualification. This compared to 10% of other children and young people. Nearly a quarter (23%) of Gypsy, Traveller and Roma

children in England and Wales aged under 19 were found to be deprived in 3 or 4 dimensions, compared to just two per cent of other children. Conversely, only a small minority (15%) of Roma, Gypsy and Traveller children were found not to be deprived in any dimension, compared to the majority (67%) of all other children.

Figure 3: The children's multidimensional disadvantage framework



3.3 Findings from recent high-level inquiries and reviews

Official education data establishing the scale of the challenges facing children and young people from the Gypsy, Traveller and Roma communities is regularly and widely cited in specialist research on child and young people from the communities; in advocacy and analysis produced by civil society organisations that represent Gypsy, Traveller and Roma people; in equality and human rights monitoring by national and international bodies; and in the multiple recent high level inquires, committees and initiatives cited in this study. This includes the final report of the Race Disparity Audit 2017 (Cabinet Office 2018); the Women and Equalities Committee inquiry on tacking inequalities faced by the Gypsy, Traveller and Roma communities (2019); the Cabinet Office Gypsy, Roma and Irish Traveller ethnicity summary (2022); and, most recently, the House of Commons Education Committee (2022ab). Most recently, the scale of the educational challenges faced by young people from Gypsy, Traveller and Roma communities was highlighted by the House of

Commons Education Committee in 2022, which, along with many other high-level bodies that have considered this issue, concluded that the Government must take urgent concrete action to improve the educational attainment of Gypsy, Traveller and Roma children and young people. Some of the key findings of these bodies are summarised in Figure 3.

3.3.1 Women and Equalities Committee

The major report on tackling inequalities published by Women and Equalities Committee in 2019 cited in earlier sections of this paper similarly presented evidence on the extent and depth of the poor outcomes experienced by Gypsy, Traveller and Roma people across different critical areas of life. Summarising the existing evidence on what we know about the inequalities facing Gypsy, Roma and Traveller communities, the Committee concluded:

> "Gypsy, Roma and Traveller people have the worse outcomes of any ethnic group across a huge range of areas, including education, health, criminal justice and hate crime. Too often local authorities and public services fail to differentiate between different groups who have different needs. Our inquiry has found that, while many inequalities have existed for a long time, there has been a persistent failure by both national and local policy-makers to tackle them in any sustained way. This failure has led to services that are ill-equipped to support Gypsy, Roma and Traveller people to use services that they need and are entitled too" (Women and Equalities Committee 2019).

In relation to educational disadvantage, the Committee heard witness testimony and raised a series of key concerns around poor attendance, children missing from education and elective home education. The Committee heard evidence in relation to high drop-out rates, particularly from secondary education, with some children taken out of school as early as the end of primary school, some not attending and some never registering at a school at all. The Committee also heard evidence on a trend towards elective home education, noting that there were cases of successful and unsuccessful home education, with some children starting work at very young ages and others staying at home with no formal education. The assessment of the Committee was that while 'missingness' from school sometimes resulted from travelling because of needs of parents work, this was in a minority of cases, and broader reasons included bullying, not taking needs into account, the perception that school was not relevant; traditional gender roles; and concerns about the curriculum being inappropriate. The Committee also referred to additional 'push factors' including off rolling as well as high rates of exclusion.

The Committee put forward a raft of policy recommendations, from anti-discrimination and anti-bullying measures to inclusive schooling policies, cultural awareness, parental engagement and role models; pupil passports

for children with irregular educational histories due to frequent travelling, or moving in and out of school or moving from one local authority to another. Particular emphasis was put on right of the child to education and the duty of local authorities to ensure that children who are not in school are being 'suitably' educated, as well as evidence from the Children's Commissioner that the vast majority of councils felt that they did not have the powers to ensure that children are getting an education (Women and Equalities Committee 2019).

3.3.2 Equality and Human Rights Commission

The EHRC's equality and human rights monitoring reports have repeatedly highlighted key concerns relating to educational disadvantage and broader multidimensional inequalities impacting on the Gypsy, Traveller and Roma communities. The Commission's 2010 assessment cited evidence of lower life expectancy and high rates of miscarriages and still births amongst the Gypsy and Traveller communities; as well as the adverse effects of media stereotyping; and evidence of exceptionally low educational attainment and widening gaps. The Commission's 2015 assessment raised key concerns around the quality and location of sites for Gypsies and Travellers and difficulties with GP registration for those without fixed addresses or photographic ID. It raised further concerns around lowest attainment levels and identified that educational gaps between Gypsy and Traveller young people and other White children widened since its 2010 report. A follow up report in 2016 cited evidence of lower living standards and employment rates; poor health outcomes (including lower life expectancy, higher infant mortality rates, high maternal mortality rates, low child immunisation rates, higher prevalence of anxiety and depression, and low levels of GP registration); poor educational attainment; over-representation in prison; and widespread stigmatising treatment and experiences of harassment and hostile treatment. The Commission's 2018 assessment concluded that Gypsy, Roma and Travellers face multiple disadvantages across different areas of life including below-average results at school; difficulties accessing healthcare; worse health outcomes; and low standards of housing (EHRC 2010, 2015, 2016 and 2018).

Figure 4: Selected findings from high level inquiries and reviews

Race Disparity Audit 2017
"Pupils from Gypsy or Roma backgrounds and those from a Traveller or Irish Heritage background had the lowest attainment of all ethnic groups throughout their school years. As shown in Figure 4.1, at age 5, around a quarter of Gypsy and Roma pupils achieved a good level of development, making them around three times less likely to do so than average. At key stage 4 the disparity is wider; in 2015/16 the Attainment 8 score – an average of points scored for attainment in 8 GCSEs including English and Maths – for Gypsy and Roma pupils was 20 points compared with the English average of 50 points and 62 points for Chinese pupils. Gypsy and Roma pupils, and those from an Irish Traveller background, also made less progress compared with the average for pupils with similar prior attainment. They were also far less likely to stay in education after the age of 16 than pupils in any other ethnic group, with just 58% of Irish Traveller pupils and 62% or Gypsy and Roma pupils staying on in 2014/15, compared with 90% of White British pupils and 97% of Chinese pupils" (Cabinet Office 2018).

Women and Equalities Committee inquiry on tacking inequalities faced by the Gypsy, Traveller and Roma communities (2019)
"Gypsy and Traveller children leave school at a much earlier age than children in other ethnic groups, they have worse attainment standards than any other ethnic group from early-years onwards72 and only a handful are recorded as attending university in any given year (although this may be because they are choosing to hide their ethnicity).73 In addition, levels of both temporary and permanent exclusions are high and almost half of Gypsy/Roma students are classed as persistent non-attenders.74 After key stage 4 (usually aged 16), a quarter of Gypsy, Roma and Traveller children go into neither education nor employment." (Women and Equalities Committee 2019).

Cabinet Office Gypsy, Roma and Irish Traveller ethnicity summary (2022)
"Considering all data on pupil performance in state-funded mainstream schools in England … [a]t all key stages, Gypsy, Roma and Irish Traveller pupils' attainment was below the national average" … [including] … "the lowest percentages out of all ethnic groups" at the end of primary school (key stage 2 attainment); "the lowest percentage of all ethnic groups" attaining a grade 5 or above in GCSE (for White Gypsy or Roma pupils); "the least likely to stay on in education after GCSES (or equivalent qualifications)" or "to get at least 3 A grades levels" (Gypsy or Roma students); "the highest rates out of all ethnic groups" in relation to suspension; the "highest permanent exclusion rates" (for Gypsy or Roma students); and "the highest rates of overall absence and persistent absence" Cabinet Office (2021)

House of Commons Education Select Committee (2022ab)

- Published data on the development goals for 4–5-year-olds shows Gypsy, Roma and Irish Traveller pupils are least likely to meet the expected standard. On average, 71% of 4-5-year-olds met the expected standard in development in the 2018-2019 school year, compared with 34% of Gypsy/Roma pupils and 39% of Irish Traveller pupils.
- On average, children and young people from these communities have the lowest educational attainment. Gypsy/Roma pupils are almost three years behind White British pupils at GCSE level, and Gypsy/Roma and Irish Traveller pupils have the lowest Attainment 8 score out of all ethnicities with 22.7 and 30.7 out of 90 respectively. The average Attainment 8 score for pupils in England is 50.9 out of 90.
- In the 2018 to 2019 school year, Gypsy/Roma pupils had 21.26% temporary exclusions whilst Irish Traveller pupils had 14.63%; the overall temporary exclusions amongst all pupils was 5.36%. For permanent exclusions, Gypsy/Roma pupils had 0.39% whilst Irish Traveller pupils had 0.27%; the overall permanent exclusions was 0.10%.
- Barely more than 8% of Gypsy/Roma and just 20% of Irish Traveller pupils achieved a strong pass in English and maths GCSE, compared with a figure of 49.9% for all pupils.
- When it comes to accessing higher education by the age of 19, young people from Gypsy/Roma (5.2%) and Irish Traveller (7.6%) communities again come at the bottom of the table.

3.3.3 Education Select Committee

In 2022, as part of its broader work in relation to disadvantaged children and left behind groups, the House of Commons Education Committee held an inquiry on the educational challenges experienced by children and young people from Gypsy, Traveller and Roma backgrounds in 2022. The Committee's observations and reflections highlighted several key issues including:

- The importance of intervention at early ages in order to build up a stronger relationship between the Gypsy, Traveller and Roma communities and the education system, including through outreach services and the new and emerging network of Family Hubs (including the development of guidance on access and inclusion to ensure effective engagement and monitoring of progress).
- The promotion of parental engagement, including by addressing intergenerational illiteracy and introducing a targeted adult education programme to support parents with low literacy levels to engage with the education system; and by schools adopting pro-active measures to build trust amongst parents, including funded role model teaching assistants and support staff in targeted schools;
- The adoption of specific measures to address low educational attainment, including measures to address low levels of pupil premium receipt despite eligibility amongst children and young people from the communities; the provision of targeted careers advice to inform on career opportunities; the identification of incidents of bullying and racism; and the implementation of the recommendations in the Timpson Review of School Exclusion by identifying and addressing the needs of groups with protected characteristics; and specific monitoring of the impact of new measures to address persistent and severe absences within schools on the attendance of Gypsy, Traveller and Roma pupils.
- The Committee also took evidence relating to the relatively high rates of withdrawal from mainstream education to elective home education and concerns about inadequate support. Here, the Committee repeated previous recommendations relating to the importance of a statutory register and accurate records of children outside school and of powers to provide support where required.

(Source: House of Commons Education Committee 2022ab).

4. Our analytical framework for assessing the 'value added' of administrative data linkages in tackling data exclusion and building up more inclusive data infrastructure and evidence

This section provides details of our approach to assessing the 'value added' of administrative data linkages in tackling data exclusion and building up more inclusive data infrastructure and evidence. We begin by highlighting the major strengths of administrative data linkages for building up evidence on smaller and / or highly disadvantaged groups (sometimes known as vulnerable, seldom heard or hard to reach groups) that have in the past typically been missing from, or invisible within, standard social survey analysis. We then identify some key limitations and weaknesses of administrative data linkages that are identified in the emerging body of literature on administrative data research - including critical issues of coverage, bias, error and disproportionality - and identify good practice recommendations for addressing, assessing and communicating these concerns. Finally, building on good practice identified by ONS, UKSA and others, we summarise the analytical framework we use in this report to assess the 'value added' of the Growing Up in England dataset for undertaking research on the lives, circumstances and needs of Gypsy, Traveller and Roma children and young people in subsequent sections of this report.

4.1 Strengths of administrative data linkages in tackling data exclusion and building up more inclusive data infrastructure and evidence

We suggested in section 2.3 that administrative data linkages can play a key role in challenging data exclusion and building up more inclusive data infrastructure and evidence in the 2020s. The Growing Up in England dataset is an example of a specific type of administrative data linkage – namely, a large-scale longitudinal administrative-population census data linkage. Here, we highlight the major strengths of large-scale longitudinal administrative-population census data linkages in building up data infrastructure that more fully represents smaller and highly disadvantaged groups that have in the past been typically missing from, or invisible within, standard social survey analysis.

4.1.1 'Value added' of administrative data

In a previous study (Vizard et al 2018), we identified three key reasons why standard social surveys typically result in data exclusion. First, the limitations of standard sample frames mean that there is typically only limited coverage, or no coverage, of relevant subpopulations. Second, the absence of, or the limitations of, identifiers that are included in social survey instruments can fail to identify relevant subpopulations. Third, low population fractions can mean that, even if the relevant sub-populations are covered by a sampling frame and are identified within the variables collected in the social survey, low numbers mean that findings relating to sub-populations of interest cannot be reported because sample size is small to support reliable and valid statistical analysis or because the statistical findings have to be suppressed for disclosure reasons. These factors mean that in the past Gypsy, Traveller and Roma people have typically been missing from, or invisible within, standard social survey analysis.

In contrast, three key strengths of administrative data for building up evidence on groups that have in the past been typically missing from, or invisible within, social survey analysis are identified in literature on administrative data research (e.g. Harron et al 2017). The first key strength relates to scale: administrative datasets can help to overcome the limitations of small sample size as administrative records are often collected across large populations or sub-populations. For example, the administrative education records that are routinely collected by the Department for Education and which constitute an important component of the Growing Up in England dataset include the records of all children and young people in schools, alternative provision and colleges. Second, administrative records often routinely include relevant markers of vulnerability, disadvantage and needs that are used for operational policy purposes, such as markers of special educational needs and free school meals status within administrative education data systems. These identifiers can be used in administrative data research as a means of identifying relevant vulnerable sub-populations. Third, some public services specifically target relevant disadvantaged sub-populations and administrative records from the relevant data systems can provide information about these groups and identifiers for the sub-populations of interest which are rarely available in social surveys. For example, Wave 2 of the GUiE linkage draws on administrative records collected by local authority children's services for operational purposes and which cover children and young people who fall within scope of the children's services administrative categories 'children in need' and 'looked after children'

(administrative categories which are underpinned by children protection system and the Children's Act 1989).

4.1.2 'Value added' of administrative data linkages

As noted in section 2.3, administrative data *linkage* is the process of joining together or combining an administrative dataset with another dataset, for example to a social survey dataset, another administrative dataset or a population census. The process of data linkage involves matching records across different datasets and hence assessing whether the records belong to the same entity (for example, the same person, household or firm) (ONS 2021a). Administrative data linkages have key strengths in enabling the information gathered in different administrative data systems to be combined and integrated. For example, recent administrative data linkages combine and integrate data from the education, health, children's social care and police and criminal justice systems, enabling the information and insights from these different administrative systems to be used up in a 'joined-up' way.

A recent Cabinet Office review of data gaps affecting Gypsy, Traveller and Roma people (Cabinet Office 2022) highlights several advantages that can result from the use of linked administrative data. This analysis notes that in addition to benefits resulting from the availability of relevant identifiers and increased sample size compared to standard social survey analysis, the use of administrative data eliminates sampling variability which is a source of uncertainty in statistical estimates. Additionally, linking administrative datasets can provide a means of addressing missing ethnicity data. For example, where Gypsy, Traveller or Roma identity is not recorded in a particular administrative data system such as the health system, using a linked administrative dataset may enable ethnicity information gaps to be addressed. Additionally, the use of linked administrative datasets may make it possible to produce higher quality data for the Gypsy, Traveller and Roma groups, or indeed any ethnic group, if the ethnicity classification in a linked dataset is known to be more reliable. While data linkage does not always increase the sample size or the number of records available in the linked dataset, this might be the case if records that have missing ethnicity information are replaced by a known ethnicity classification from a linked dataset. In this study, we explore this approach, demonstrating how the ethnicity information made available in the Wave 1 GUiE linkage can be pooled to improve identification of Gypsy, Traveller and Roma children and young people compared to the use of the 2011 Census and education records separately.

Additionally, *longitudinal* linkages of administrative records have the added strength of enabling children's interactions with public services to be followed over time. For example, combining local authority administrative records can result in large national administrative datasets while longitudinal linkages of administrative records mean that lifepaths and interactions with public services can be tracked. Longitudinal linkages of administrative education, further education, higher education and administrative employment and earnings data provide new insights on educational progression and the transitions of multiple cohorts of young people into careers. Similarly, longitudinal linkages of children's social care administrative records can provide new data on the lifepaths of vulnerable children and their interactions with children's social care services over time.

4.1.3 'Value added' of administrative data-population census linkages

Nevertheless, while administrative data has the major strengths for research on groups that have typically been underrepresented in national statistics discussed above, one key limitation of working with administrate data relates to the lack of information on the family background and household circumstances of service users. Large scale linkages of administrative datasets and population-censuses are increasingly being explored as a solution to this limitation and enable information from administrative datasets to be linked to be 'joined up' with the in-depth and rich information about individuals and their families and households that is recorded in censuses on a population wide basis.

In England, a compulsory population census has taken place every ten years. The most census was undertaken in March 2021 and a new generation of large-scale administrative data – population census linkages are planned for the upcoming period. As we set out in section 5, the Growing Up in England (GUiE) dataset is an example of a large-scale administrative data-population census linkage, and links young people's records from administrative education and social care data systems to information about their personal characteristics and household circumstances recorded in the 2011 Census.

In recent years, there have been many discussions about the potential of administrative data as an *alternative* data source to the ten yearly population census (e.g. ONS 2013, ONS 2023cd). However, our analysis in this report highlights that the 'value added' of administrative data for challenging data exclusion and building up more inclusive data and

evidence may be maximised when administrative data and population census data are *integrated* and *combined*.

4.2 Limitations and quality issues in administrative data research

While recognising the major role that large scale longitudinal linkages can play in challenging data exclusion, it is also critical to acknowledge that multiple quality issues can arise in relation to both administrative data and data linkages. As a basic reference point, Office for Statistics Regulation (OSR) has developed standards of quality assurance for administrative data and the conditions for characterising statistics as national statistics (including case studies relating to examination statistics and other indicators) (OSR n.d.a.b.). Additionally, a series of UKSA and ONS reviews of the future of data linkages put central emphasis on the importance of identifying, communicating and understanding underlying data and linkage quality, as a basis for the meaningful, reliable and valid use of administrative data (and linked administrative data) for statistical and research purposes.

4.2.1 Limitations of underlying administrative data systems

Administrative data are produced as a result of the routine operational functions of administrative and policy systems. Therefore, administrative data can only be as reliable and valid as the recording practices in these underlying administrative and policy systems and there are a range of issues that can impact on the quality, reliability and validity of administrative data records. Underlying administrative and policy systems may themselves be fragmented and can operate at different levels (for example, local, regional and central) with varying degrees of integration and consistency. Key issues highlighted in the literature include missing records; inaccurate records; variable recording practices (for example, where administrative recording practices differ across local authorities or policy or practice areas); recording errors; and inconsistent operational definitions, classifications and identifiers.

Current emphasis on administrative data linkage is being driven forward in part by increased emphasis on 'joined up' data in government. The House of Commons Committee of Public Accounts noted in 2019 that the challenges of using data across government and producing 'joined up' data across different public services include lack of cross-governmental standards, fragmentation across service areas, the absence of common personal identifiers across policy areas and outdated IT systems (House of

Commons Committee of Public Accounts 2019). Public service resource limitations and administrative capacity constraints can also impact on the quality of administrative data and improvement initiatives (House of Commons Committee of Public Accounts 2019, ONS 2021a).

Figure 5: Strengths and weaknesses of administrative data for building up new evidence on missing and invisible groups

Strengths of administrative data

- Large sample sizes
- Supports data inclusivity through:
  - the presence of relevant identifiers (including vulnerability identifiers used for operational policy purposes and ethnicity identifiers)
  - coverage of subpopulations of interest (for example, children and young people in schools, colleges or work-based learning, or children in contact with children's social services)
- Can eliminate sampling variability

Strengths of administrative data linkages

- Brings together information and insights from different datasets
- Supports joined up policy making
- Can improve quality compared to use of individual datasets separately by enabling missingness and invisibility to be better addressed (for example, in the current study, we pool ethnicity information across the component datasets in GUiE, improving identification of Gypsy, Traveller and Roma young people)
- Longitudinal linkages support analysis of interactions with public services over time
- Information about childhood background and circumstances is often not recorded in administrative data systems, but this can be overcome through administrative data-population census linkages

Caveats and limitations

- **Excluded subpopulations:** Administrative data is only created for individuals who interact with public services – key vulnerable subpopulations can be 'missing' (for example, children who go missing for children's social care services or are not in contact with health services, or who do not have a school record). This can be a source of bias and limit the extent to which generalisations can be drawn from statistical findings that draw on administrative data.
- **Missing identifiers.** Administrative data does not include identifiers of interest meaning that subgroups can remain 'invisible'.
- **Quality and error.**
  - Underlying data quality can be limited in important respects: issues can include limited coverage, missing data and inconsistent and variable quality recording practices.
  - Processes of anonymisation (strictly, de-identification) and data linkage (record matching) are associated with error. There is a trade-off between privacy and quality;
  - Error can be compounded in longitudinal linkages.
- **Disproportionality:** linkage error may disproportionately impact on certain groups, including some ethnic minority groups compared to the White group (sometimes characterised as 'hard to link' groups).

Source: draws on analysis in: UKSA (2021ab), Cabinet Office (2022), Harron et al (2017), Blackwell and Rogers 2021, Doidge et al (2021)

Particular concerns have been highlighted in relation to the quality of ethnicity recording in the context of administrative data systems. In its systematic assessment of ethnicity data gaps, the Race Disparity Audit noted that administrative data can suffer high levels of non-recording of ethnicity and overuse of 'other' categories, undermining the ability to identify differences in how people in each ethnic group are treated (Cabinet office 2018).

Raleigh (2023) highlights that the quality of ethnicity coding in administrative data systems is highly variable. Longstanding data quality issues including inconsistencies, errors and bias result from both the use of different ethnicity classifications within different administrative data systems and outmoded ethnicity classifications, as well as from underlying issues such as staff being inadequately trained in relation to coding protocols and processes. ONS (2023e) provides further insights in relation to the scale of inconsistencies between ethnicity recording in health-related administrative datasets and population census data (which is highlighted as the most robust source of ethnicity information). ONS (2023f) examines sources of variation and quality limitations, including variation in local administrative health systems, different data collection modes, inconsistent use of residual categories ('not known', 'not stated'), different staff collecting data differently, quality issues resulting from multiple entries, complex and subjecting coding processes; imperfections in data quality checking and guidance; and financial incentives which focussed on completeness and could have implications for accuracy.

In the context of the Gypsy, Traveller and Roma research, the absence of ethnicity identifiers and the use of inconsistent and inappropriate ethnicity identifiers within different administrative recording systems are key concerns that have repeatedly been raised in the literature and in the advocacy work of civil society organisations. Later sections of this report provide further details.

4.2.2 Missingness, invisibility and bias

Substantial concerns around missingness, invisibility and bias arise in the context of administrative data research. Relevant sub-populations can be missing from administrative data collections and this can impact on the extent to which an administrative data collection is representative of underlying population of interest. For example, administrative records are only created for individuals who interact with public services and individuals

who do not interact with public services are not included in the associated administrative data. As the Children's Commissioner's Office has highlighted, as some of the most disadvantaged, marginalised and at-risk children may not interact with education, health or children's services, these children may not be represented in administrative datasets. These data gaps can be associated with safeguarding concerns as well as lack of knowledge and understanding of the scale of the population of vulnerable children and lack of appropriate public policy response (Children's Commissioner for England 2017, 2019). While the current project makes use of the new GUiE linkage to produce new insights and analysis, there are nevertheless important caveats around the exclusion of Gypsy, Traveller and Roma young people who do not have a record in the 2011 Census or who do not have a record in relevant administrative education systems being excluded from our analysis. Further details of the quality of the GUiE linkage and of important caveats relating to our analysis in terms of missingness, invisibility and bias are provided in sections 5 and 6.

4.2.3 Errors associated with de-identification and record-matching

Error associated with processes of de-identification

Processes of de-identification that remove identifying personal data from administrative datasets are increasingly viewed internationally as critical to protect privacy and confidentiality and in the UK are essential conditions for the release and use of administrative data to researchers. Processes of de-identification (sometimes referred to as 'pseudonymisation' or 'hashing') are increasingly applied to administrative datasets and involve personal data such as names and dates of birth being transformed into a unique string of letters and numbers. However, as this process itself introduces error, ONS (2013) set out a good practice linkage methodology designed to minimise the additional error that is introduced through this process. Options include the use of deterministic methods involving exact matching on multiple fields such as names, addresses and sex; the use of probabilistic matching (similarity score based methods) to maximise linkages, once deterministic method potential is exhausted; the use of clerical and manual reviews which can help to minimize matching error; and transparent quality assurance processes, the publication of precision and recall rates relating to true positive and true negatives. Proposals for improved data linkages and new cross-governmental data linkage networks were set out in ONS (2021a), including the importance of improving

privacy-preserving-record-linkage whilst carrying out linkage 'in the clear' where possible to optimise quality.

Linkage errors

The process of linking records across datasets itself introduces additional error. As ONS note, linkage error can result in missing or incorrect data for analysis; mistaken inclusion or exclusion of records from a dataset; records that are split to look like multiple records; and merging of multiple records to appear as one entity. This affects the interpretation and analysis of data, potentially resulting in incorrect conclusions being drawn and ultimately the misuse of data (ONS 2021a). A basic standard in administrative data quality assurance relates to reporting of the linkage match rate (that is, the proportion of the records that have been successfully linked). However, as ONS (2021a) highlights, assessing linkage quality is about much more than just the match rate. Linkage error includes both false positive errors (that is, pairs of records which are erroneously matched) and false negatives (pairs of records that are not matched but which should have been matched). As linkage strategy impacts on the degree to which false positives and false negatives are 'tolerated' within a data linkage exercise, transparent quantifiable information on precision rates is increasingly viewed as essential. The proportion of records that are matched deterministically and probabilistically should also be clearly communicated; any pre-processing of data identified; and quantitative indicators of quality to support the use of the linked datasets for estimation purposes should be developed and communicated (ONS 2013, 2019, 2020ab, 2021a, n.d; Office for Statistics Regulation n.d.a.b.; UKSA 2023, Harron et al 2017, Blackwell and Rogers 2021; Doidge et al 2021).

Longitudinal linkages and compounding errors

Moreover, these administrative data quality issues can compound in the context of longitudinal linkages (Blackwell and Rogers 2021; c.f. Doidge et al 2021, ONS 2021a). The basics of administrative data collections – for example such as recording practices and standards, policy and operational definitions and classifications, and the use of identifiers - may also change over time while the effects of linkage error can also be amplified over a longer time window. In the context of the GUiE linkage, particular issues may arise in relation to variation on the recording of personal and household details and unique pupil numbers, which are used to match education and children's social care records. This variation in the recording

of names, addresses and other personal information can increase over time, whilst errors associated with personal identifiers in underlying administrative source data, such as unique pupil numbers, can compound in longitudinal data. For example: postcode details will change when children and young people move homes; children and young people may live at more than one addresses; and personal details such as names, gender and ethnic identification can change over time or be inconsistently recorded. If a child moves school, school records will need to be transferred and sometimes errors in recording unique pupil numbers can occur. If a child is adopted, a unique personal identifier may, by intention, be changed. In the context of higher education administrative data linkages, recording of term time and home addresses can be an additional source of error (c.f. DfE 2019 and ONS 2019).

Errors associated with emerging new technologies and methods

The UKSA has highlighted that data linkage methodologies are themselves evolving in the wake of the 'big data' revolution and technological change. The increased application of algorithms and machine learning techniques raise additional issues around quality assurance and automation (ONS 2021a, UKSA 2021ab, 2023).

Disproportionality in linkage error

UKSA (2023) notes that the linkage error that arises in the process of integrating two administrative datasets is likely to be a source of bias as it disproportionately impacts on certain groups resulting in linked datasets that exclude or incorrectly link them. This affects some ethnic minority groups compared to the White group, who are, as a result, sometimes characterised as 'hard to link' groups. Recommendations include setting up workstreams to specifically develop bespoke methods to address the linkage of these groups and to examine bias affecting groups with protected characteristics. This will help to ensure that the quality of linkage is addressed in the context of the Equality Act (2010) and that specific action is undertaken to maximise the inclusivity of statistics based on linked data (UKSA 2023).

4.3 Good practice in assessing the quality and inclusivity of large-scale administrative data linkages

In recognition of the concerns set out above, the explicit assessment of the strengths and limitations of linked administrative datasets is

increasingly viewed as an essential and intrinsic element of administrative data research. This includes undertaking detailed explicit assessments of the quality of underlying administrative datasets and the quality and inclusivity of data linkages (ONS 2021a, UKSA 2023). Good practice in administrative data research requires that empirical findings that draw on linked administrative data sets are based on a clear understanding and communication of quality, bias and error. Statistical findings drawing on administrative and linked administrative data should be appropriately framed and caveated to capture and reflect data limitations and any issues arising in relation to the coverage, reliability, validity and quality of administrative data and linkages (ONS n.d, Diamond 2021a, ONS 2021a, Office for Statistics Regulation n.d.a., UKSA 2021ab, UKSA 2023, Harron et al 2017, Blackwell and Rogers 2021, Doidge et al 2021).

The 'total error' framework

The 'total error' framework has been developed to ensure that the identification of error sources at each stage in the 'data journey' in the creation of a linked dataset - from initial data collection, through pre-processing, linkage and longitudinal linkage - is explicitly built into statistical design and that quality indicators for reporting on the statistical properties of linked data sets are explicitly developed (Blackwell and Rogers 2021). As noted above, a range of quality issues and errors can result from the imperfections of the original underlying separate administrative datasets. These imperfections can relate to coverage, missing data and inconsistent recording practices across different administrative systems. Processes of anonymisation (de-identification) and the process of integrating multiple administrative datasets into a single combined dataset (the record matching process) are both associated with additional error. Longitudinal linkages can introduce further additional errors as well as the possibility that the effects of any errors compound. Additionally, the quality implications of new technologies and automated processes need to be understood (ONS n.d, Office for Statistics Regulation n.d.a., UKSA 2023, Harron et al 2017, UKSA 2021ab, ONS 2021a, Blackwell and Rogers 2021, Doidge et al 2021).

For these reasons, ONS recommends the use of the 'total error' framework when a new administrative data linkage is produced. Underlying principles of error assessment are set out in Doidge et al (2021) and ONS (n.d.) provides an example of the application of the 'total error' framework. This recommends a staged approach to understanding the strengths and limitations of administrative data, involving in the first phase assessment

of the quality of each underlying dataset and in the second phase assessment of the quality of the integrated dataset ONS (n.d.). Quality limitations and error should be explicitly assessed at each stage of the 'data journey' and quality metrics should be produced to capture and communicate the nature and scale of error and how it progresses through the linkage process. This involves the application of a range of different techniques including the use of 'gold standard' (or training) data where true match rates are known; clerical review; comparison of linked and unlinked records; analysis of matching quality; and comparison of linked data to external reference statistics (ONS 2021a).

The explicit assessment of data inclusivity

Explicit assessments of data inclusivity are also increasingly recognised as a critical element of quality assessment in administrative data research. This is especially important in the context of research that aims to tackle data exclusion. While groups with a low population fraction and / or highly disadvantaged groups (sometimes known as vulnerable, seldom heard or hard to reach groups) are becoming increasingly analysable in large scale administrative data linkages, there are good reasons for anticipating that they might be disproportionately impacted by coverage, bias and quality issues including the different manifestations of error that occur at each of the different stages of the development of a new linked administrative dataset.

The specific importance of undertaking assessments of inclusivity in administrative data research was highlighted in the recommendations of the UKSA Inclusive Data Taskforce. The Taskforce raised key concerns around missingness and coverage and recommended a deep dive study to assess the extent to which linked datasets are representative of highly disadvantaged populations and to compare 'what works' in data inclusivity (UKSA 2021ab).

As noted above, issues of inclusivity and disproportionality in data linkage were also explicitly addressed by UK Statistics Authority (2023). This notes that to draw accurate and meaningful conclusions from the analysis of administrative data and to avoid drawing conclusions that are potentially inaccurate, misleading or discriminatory, it is essential that sources of bias and error and their implications are fully understood. This analysis recommends that assessments of inclusivity are undertaken as part of quality and error assessment exercises. Recommendations from this analysis included undertaking clerical matching exercises to facilitate high quality; improving the quality of linkage information on quality, precision

and recall metrics; explicitly assessing inclusivity and any disproportionality in the linkage process; and developing high quality supporting materials so that researchers can fully understand the quality of administrative data (UKSA 2023).

Assessing the 'value added' of the Growing Up in England (GUiE) dataset in this report

Building on these good practice principles, in the context of the current project, we have built an explicit assessment of GUiE quality and inclusivity, and the extent to which the new GUiE dataset can be thought of as being representative of the underlying population of Gypsy, Traveller and Roma children and young people, into our research design. The importance of undertaking an explicit assessment was highlighted in the NSDEC recommendations relating to our project discussed in section 1, which required us to take account of the importance of missingness as a source of potential bias.

The analytical framework that informs our assessment of the strengths and limitations of GUiE for building up new evidence on children and young people from the Gypsy, Traveller and Roma communities in later sections of this report is summarised in Figure 4 and draws on the good practice recommendations set out above. In particular, we build in explicit recognition that quality issues and error, and issues of inclusivity and disproportionality, can occur in each step of the creation of an administrative data-population census linkage such as GUiE. We set out a two-staged approach to assessing quality and inclusivity, involving in the first phase, an assessment of the quality and inclusivity of each of the underlying datasets; and in the second phase, an assessment of the quality and inclusivity of the integrated datasets, with further consideration of the additional issues raised by the longitudinal nature of the data. Additionally, in terms of data inclusivity, we put particular emphasis on the importance of assessing missingness and disproportionality.

Figure 6: Towards a good practice framework for assessing the quality and inclusivity of administrative data linkages

2 stage assessment

For each underlying dataset

- Assessment of coverage/ missingness and missing identifiers/invisibility
- Assessment of underling data quality (e.g. missing data, consistency and equality of recording practices, classifications and identifiers, etc.)
- Assessment of any pre-processing or processes adopted to correct for missing data
- Identification and assessment of any underlying data linkages (any underlying linked data should be assessed using same procedure set out below for linked datasets)

For the integrated (linked) dataset

- Assessment of coverage/ missingness and missing identifiers/invisibility
- Assessment of consistency of categories and classifications
- Assessment of linkage quality and error
    - Overall linkage rates
    - Quantification of false positive and false negative matching (recall and precision rates)
    - Communication of linkage methodology and quality (e.g. proportion of records linked using deterministic methods, proportion linked using probabilistic records, use of clerical review methods, linkage 'in the clear', use of machine learning techniques, duplicates)
    - For longitudinal linkages, assessment of the extent to which errors compound
    - Errors associated with emerging new technologies and methods
- Triangulation of statistics drawing on linked dataset with external reference statistics
- Assessment of any disproportionality in linkage rates (that is, of differential linkage rates for different groups, for example by age, sex, socio-economic disadvantage or ethnic group)
    - Analysis of the residuals from the linkage exercise and comparison of linked and unlinked records (supporting an understanding differences in the matched and unmatched residual cases matched by characteristics such as age, sex, socio-economic disadvantage and ethnicity)

Source: draws on good practice and analysis from a range of sources especially ONS n.d, ONS 2021a, UKSA (2021ab), Diamond 2021a, Cabinet Office (2022), Office for Statistics Regulation n.d.a., UKSA 2023, Harron et al (2017), Blackwell and Rogers 2021, Doidge et al (2021).

5. The Growing Up in England (GUiE) dataset: introduction, overview and general strengths and limitations

This section provides an introduction to the newly linked Growing Up in England (GUiE) dataset and a general assessment of its potential, strengths and limitations. We begin with a discussion of the background to the creation of the dataset and an overview of the underlying data sources that it aims to bring together and integrate. We then provide a summary of the evolution of the GUiE build and of published information on content, design and quality drawing on a series of ONS outputs. Next, we provide a record of some of the key issues relating to content, design and quality that we have followed up during the lifetime of our project and of the key clarifications that we have received. These are important to record as they inform our assessments of the GUiE Wave 1 analysis and our approach to the research exercise we report in section 7 and may also be useful for other researchers. Finally, drawing on the above, we set out our working summary description of GUiE Waves 1–3.

5.1 The GUiE dataset: background, rationale and underlying datasets

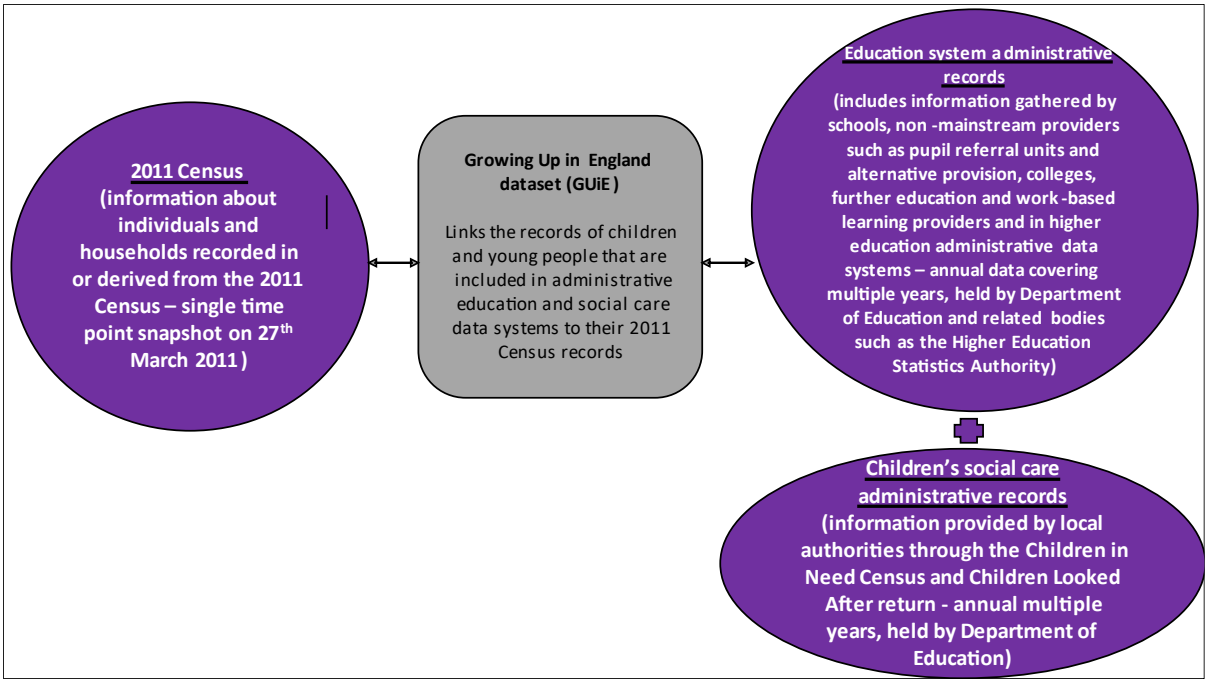
The Growing Up in England (GUiE) dataset is one of the new generation of large-scale longitudinal administrative data population-census linkages discussed in section 2. It brings together administrative education records covering schools, further education and learning and higher education, with administrative records about contact with children's social care services, all held by the Department for Education and related bodies, and information about individual and household characteristics and circumstances from the 2011 Census.

5.1.1 Background and rationale

The GUiE dataset was created as a collaboration between the ONS, Department for Education and ADR UK. The Children's Commissioner for England workstream on child vulnerability has repeatedly highlighted the importance of data and evidence gaps relating to a long list of vulnerable groups and the critical importance of a concerted national statistics response to build up better data and evidence on the needs and outcomes of vulnerable and high-risk groups (Children's Commissioner for England 2017, 2019). The Data for Children Partnership, a strategic partnership

between ADR UK, the Office of the Children's Commissioner England and ONS, was established to address this concern. The GUiE linkage was proposed in this context and progressed as a response to the identified need for a new dataset focusing on vulnerable and highly disadvantaged children.

Figure 7: Growing Up in England (GUiE) dataset – underlying administrative data systems



5.1.2 Underlying datasets

The GUiE linkage (Waves 1-3) as currently designed aspires to draw on and integrate data from the following sources:

- Information about individual educational outcomes recorded for routine operational and policy purposes within the education system, including schools, nonmainstream education, further education, work-based learning and higher education administrative systems;
- Information about children who are recorded for routine operational and policy purposes through children's social care administrative systems.
- Detailed information about individual characteristics and family and household circumstances recorded in the 2011 Census;

Administrative data from the education system

The education component of Wave 1 GUiE is derived from the feasibility All Education Dataset for England (AEDE). The overview information provided in ONS (2019, 2022d, 2023) highlights that the feasibility AEDE is itself a large-scale and complex longitudinal linked education dataset that was pre-processed and linked by Department for Education prior to its secure transfer to ONS in 2017. Specifically, the feasibility AEDE dataset brings together and integrates three types of administrative education records:

1. **Student records sourced from the National Pupil Database (England).** These records cover pupils in schools maintained by local authorities, academies and free schools in England, as well as those in non-mainstream provision such as Pupil Referral Units (PRUs) and Alternative Provision in an academic year. This data is sourced from multiple underlying administrative data collections including the English Schools Census; returns from Pupil Referral Units; and the Alternative Provision Census.
2. **Further education and work-based learning records sourced from Individual Learner records (England).** All providers of further education in England including publicly funded colleges, training organisations, local authorities and employers acting as FE providers are required to return ILR data for learners funded through government funding. Further education colleges are required to return ILR data for all learners (including those *not* funded by government) – consequently, some learners not funded by government are included in ILR data. Higher Education Institutes (HEIs) must also return ILR data for adult learners who receive publicly funded loans. HEIs are also required to return ILR data for learners funded through Advanced Learner Loans (adults who receive loans to cover tuition fees for a range of courses including A levels, general and vocational qualifications, and access to Diplomas of Higher Education.
3. **Higher education records sourced from the Higher Education Statistics Authority HESA (Great Britain).** All government-funded higher education institutes are required to send data to HESA. Additionally, further education bodies that deliver higher education send similar information to HESA. The HESA data excludes students studying outside the UK.

(Source: ONS 2019, 2019b 2022d, 2023)

The feasibility AEDE was originally created by the Department for Education and transferred to ONS as a basis for a feasibility project to investigate the potential of administrative data to provide information on educational qualifications typically collected by the population census and surveys. This research exercise is described in more detail in section 5.2.1 below. Also note that the feasibility AEDE dataset draws heavily on the Young Person's Matched Administrative Dataset (YPMAD). YPMAD is itself a large scale longitudinal linked administrative dataset that is produced by Department for Education from multiple underlying data files including NPD data, ILR data and data from awarding bodies. Further details of YPMAD are provided in section 5.3.3.

Wave 1 GUiE includes information from the feasibility AEDE that was originally recorded for operational and policy purposes within the education, further education and work-based learning systems (the first two categories of administrative education data included in the bullet list above). However, it was not possible to include the higher education data included in the feasibility AEDE in the Wave 1 or Wave 2 GUiE releases, and discussions about the inclusion of the higher education data in Wave 3 GUiE are ongoing at the time of writing.

The Wave 2 and Wave 3 GUiE release includes a range of additional administrative education non-attainment data including additional data relating to exclusions, absences, free school meals, special educational needs and disabilities and further data from the English School Census. This additional data was not included in the feasibility AEDE but rather was separately sourced from the National Pupil Database held by the Department for Education.

Administrative data from the children's social care system

The GUiE Wave 2 and Wave 3 releases also include (or will also include) administrative records from children's social care administrative systems. This information is collected locally within the children's social care services departments of local authorities and is provided to the Department for Education annually through the Children Looked After (CLA) data return and the Children in Need (CIN) Census. The terms 'children looked after' and 'children in need' are official policy categories used within the children's social care system and have specific operational meanings.

The 2011 Census

The 2011 Census for England and Wales provides detailed information on individual characteristics and family and household circumstances for individuals (adults and children) resident in England and Wales on 27th March 2011. Note that this data is a *snapshot* of the characteristics of individuals and households on Census day. Therefore, information derived from the 2011 Census about personal and household characteristics provides a snapshot of the situation in a household where a young person was resident on the day of the Census - not on an annual basis. Additionally, the 2021 Census (which provides a similar snapshot of the England and Wales resident on 21 March 2021 is not currently included in the GUiE linkage).

5.2 Evolution of the GUiE build and summary of quality assessment

The GUiE design and build have been taken forward within ONS, with the linkage undertaken by the ONS data engineering team. As no common personal identifier is included in both the 2011 Census and the feasibility AEDE that could be used to link records across these datasets, it was necessary to produce and refine a GUiE linkage methodology using different combinations of person information such as forename, surname, date of birth, gender and postcodes. The development of the dataset, including the linkage methodology, evolved iteratively over at least four years. Key steps included: the development of a general framework for the de-identification of administrative data; an initial dataset linkage exercise undertaken by ONS in 2019 as a basis for comparing statistics on highest educational qualification derived from population census data, administrative data and social survey sources; the creation of a Proof of Concept (PoC) GUiE dataset in 2020 and initial analysis of the PoC GUiE dataset by an ONS team; and finally the deposit of iterations (or waves) of the GUiE dataset into the ONS Secure Research Service for researcher access. At the time of writing, GUiE Waves 1 and 2 have been deposited into SRS, while Wave 3 is under development, with elements of Wave 3 already being accessible to researchers.

5.2.1 Initial feasibility AEADE-2011 Census linkage

We begin by mentioning two early ONS research exercises which are not strictly speaking part of the GUiE build but which in our view have been important in shaping and influencing the build. The first is an early

experimental linkage of Higher Education Statistics Authority (HESA) and Census 2011 data which was undertaken in the context of foundational research setting out a general framework for the de-identification of administrative data (ONS 2013). This study both advanced administrative education data – population census linkages and set out a general framework for de-identification using an innovative matchkey linkage methodology based on pseudonymised ('hashed') personal data which has influenced subsequent ONS work in this field.

A second key early exercise linked the feasibility AEDE to the 2011 Census as a basis for the ONS highest educational qualification research exercise discussed above. While strictly speaking this exercise was not part of the evolution of the GUiE build, it provides useful background and the discussion of linkage methodology and quality set is informative (ONS 2019 - Haines, Taylor and Tietz). This exercise was undertaken in the context of ONS work to examine the potential for administrative data to replace information on educational qualifications collected by the census. To explore this issue, data on highest education for persons aged 16-24 sourced from three different datasets, the 2011 Census, administrative education records and the Annual Population Survey, was compared. The exercise concluded that the number of people across England with each highest level of qualification was broadly similar in administrative education data to numbers reported in 2011 Census and the Annual Population Survey (Taylor 2019).

ONS (2019) provides further methodological details. Building on the approach to de-identification set out in ONS (2013), analysis of administrative education records for this exercise drew on the feasibility AEDE dataset. The analysis proceeded by creating a version of the feasibility AEDE that was restricted to student records for academic year 2010/11 or the closest available record for that year where a record for 2010/11 was not available (with alternative academic years selected via 'nearest academic year' selection procedure). The rationale for this approach is that 2010/11 is the 'closest' academic year to the point to 2011 Census day (27th March 2011). Hence feasibility AEDE student records for academic year 2010/11 would be expected to produce the highest percentage of successful record matches using deterministic methods (that is, a pattern of exact matches against multiple matchkey fields including forenames, surnames, postcodes, age and gender).

Applying this method, 66% of AEDE student records for academic year 2010/11 were successfully matched to 2011 Census records using deterministic methods, including 93% of records from the national pupil database, 82% of individualised learner record data and 76% of higher

education statistics agency records. Deterministic linkage matches were found to decline as the academic year moved away from 2010/2011 with 79% of NPD, 65% of ILR and 61% of HESA records were deterministically linked to 2011 Census data, regardless of academic year. Having exhausted potential for successful matches using deterministic methods, probabilistic (score based) methods were then applied to remaining unlinked records, with success rates highest for earlier academic years. The application of probabilistic linkage methods increased overall linkage success rates only marginally with 67% of AEDE records for academic year 2010/11 linked to a 2011 Census record using either deterministic or probabilistic matching techniques.

Looking at linkage success rates from the 2011 Census side of the dataset, 84% of 2011 Census records relating to persons resident in England aged 16 to 24 years on 31 March 2011 were linked to the 2010/11 AEDE. Of this total, deterministic matching resulted in 82% of 2011 Census records being linked while probabilistic matching resulted in a further 2% of 2011 Census records being linked (Appendix: ONS 2019 - Haines, Taylor and Tietz).

5.2.2 'Proof of concept' GUiE

The next step in the evolution of GUiE build was the development of a 'proof of concept' (PoC) GUiE dataset. The PoC GUiE linked a *bespoke extract* of the feasibility AEDE (rather than the *full* feasibility AEDE) to the 2011 Census using deterministic methods. The bespoke extract of the feasibility AEDE was restricted to student records sourced from the NPD and further education and work-based learner records sourced from the ILR. That is, higher education (HESA) records were *not* included in the bespoke extract of the feasibility AEDE used as the basis for the PoC GUiE linkage. Additionally, the bespoke extract of the feasibility AEDE used for the PoC linkage was restricted to academic years 2010/11 to 2014/15 (ONS 2020a).

To link the individual records in the bespoke extract of the AEDE to those in the 2011 Census, the AEDE feasibility attributes data containing the source and academic year variables were first linked onto the feasibility AEDE index using the Pupil Reference Number (PRN - an identifier that comes from Department for Education). Second, a process whereby the feasibility AEDE index was "subset" into the 2010/11 academic year with the English School Census identified as the source dataset was applied. Additionally, a de-duplication procedure was then applied to the AEDE subset for linkage, with duplicates (multiple entries of individual records)

removed using the SAS 'nodupkey' routine. This retains only the first instance of a record. A high linkage rate was achieved with 2,035,289 records (90%) of 2,250,655 records linked to a corresponding 2011 Census record (ONS 2020a).

The provision of 2011 Census records in the PoC GUiE dataset includes the individual records of PoC GUiE data subjects (that is, young people whose 2011 Census records have been successfully matched to the feasibility AEDE). In addition, the individual records of other household members - that is, the individual records of the other individuals in the household where the data subject resides - were also included in PoC GUiE. Household identifiers and relationship information, as well as some derived household level variables, were also included in the PoC GUiE dataset. This detailed information about all of the individuals in the household provides strong foundations for research that examines young people's family and parental circumstances and characteristics. The PoC GUiE dataset was used as the basis of a Proof of Concept study undertaken by the ONS Centre for Equalities and Inclusion, with a view to exploring the 'value added' of the linkage in terms of adding to the existing evidence base on children's outcomes, especially for subgroups. This exercise resulted in a set of illustrative findings and breakdowns of children's educational outcomes by children's, parental and household characteristics including child level information on disability and ill-health; mother's country of birth and highest educational qualifications; family type (number of siblings); and household circumstances such as unemployment. These findings were heavily caveated in terms of their preliminary and experimental nature (ONS 2020a).

5.2.3 GUiE research datasets (GUiE Waves 1-3)

At the time of writing, two GUiE research datasets have been made accessible to researchers (Waves 1 and 2). Additionally, Wave 3 is under development (with elements of Wave 3 already released to researchers).

Wave 1 GUiE

As a base for Wave 1 GUiE, a bespoke extract of the feasibility AEDE and Census 2011 were re-linked in a new ONS linkage environment. Like the 'Proof of Concept' GUiE dataset, Wave 1 GUiE was constructed from a bespoke extract of the feasibility AEDE and does not include higher education data. The time coverage of Wave 1 GUiE is longer with all years of data included in the feasibility AEDE included in Wave 1 GUiE (that is,

academic years 2001/2 to 2014/15 for pupil data and academic years 2002/3 to 2014/15 for the further education and work-based learning data).

Several quality improvements were implemented for the Wave 1 GUiE design. The PoC linkage was undertaken using a matchkey methodology that was based on 'hashed' personal data (that is, personal data that was already pseudonymized when the matching exercise was undertaken). However, this method has important limitations, since minor administrative inconsistencies (for example, spelling differences in names) can result in linkage failure. To improve on this methodology, the Wave 1 linkage was undertaken using identifiable data (using a so-called 'in the clear' approach) enabling the use of improved techniques designed to increase matching success rates and reduce the matching error rates for both false positives and false negatives. A total of 91 matchkeys containing different combinations of personal identifiable information were used as a basis for the linkage and a range of techniques (eg Soundex, distance functions, geo-referencing and standardization) were applied to improve matching success rates (ONS 2022d).

As in the context of the PoC linkage, to create the Wave 1 linkage, AEDE attributes data containing the datasource and academic year variables were first linked to the feasibility AEDE index on the Pupil Reference Number. Additionally, multiple entries of individual records were again removed using a de-duplication procedure using a 'nodupkey' routine in SAS which retains only the first instance of a record (ONS 2022d). However, in the context of the Wave 1 linkage, the restriction whereby the index was "subset" into the 2010/11 academic year where the English School Census was identified as the source subset, which was used in the PoC linkage, was not applied (ONS 2023b). Therefore, for the Wave 1 linkage it is *not* necessary to have a 2010/11 record in the feasibility AEDE to be included in the linked dataset.

The Wave 1 linkages used a combination of deterministic (rule based) and probabilistic (score based) methods, with probabilistic methods only utilized after deterministic methods were fully exhausted. Overall, the Wave 1 linkage includes the records of 7 million children and young people. In total, 7,432,115 feasibility AEDE records were successfully linked to the 2011 Census with 98% (7,255,360) linked by deterministic linkage methods and the remaining 2% (176,758) linked by probabilistic linkage methods (ONS 2022d).

A clerical review was undertaken to determine the threshold of acceptability of matching error - where the term 'error' refers to errors introduced by the linkage process rather than any original error relating to

the quality of the underlying administrative data. As defined above, linkage quality is regulated by the percentage of false positive and false negatives that are deemed acceptable. This threshold was a cut-off score of 0.75 (with records below the 0.75 threshold considered to be non-matches and records above the threshold considered to be matches). A manual review of a sample of feasibility AEDE records established a precision rate of 83.9% and a recall rate of 93% and that the quality of the linkage was therefore within the range deemed acceptable (ONS 2022d).

ONS (2022d) does not report precise match rates for the Wave 1 GUiE linkage (that is, the precise rates of successful matching between feasibility AEDE records and those in the 2011 Census). A bespoke analysis for the current project examines match rates by ethnic group. This provides information on match success rates for young people from the Gypsy, Traveller and Roma group and other ethnic groups and includes an overall matching rate, with and without duplicates dropped (see Table 2 in section 6.2.4 for further details). As match rates are likely to be variable by academic year, with rates in 2010/11 likely to be highest as this is the 'closest' academic year to the time of the 2011 Census, reporting of match rates by academic year would also be informative. Therefore, as part of future linkage documentation, we recommend the publication of Wave 1 match rates by academic year, with precise information on both the feasibility AEDE and the 2011 Census residuals (that is, successful match rates expressed as percentages of both 2011 Census records and feasibility AEDE records). A model for this is provided in ONS (2019).

Wave 2 GUiE

Wave 2 GUiE 'joins' Wave 1 GUiE records to several additional datafiles using the 'pupilmatchingrefanonymous' identifier. The additional data made available in Wave 2 includes a wide range of additional non-attainment data derived from administrative education records. This included additional pupil record information on exclusions, absences, Special Education Needs and free school meal status, together with additional pupil record information from the English School Census (these take place in Spring, Summer and Autumn, but only Spring Census information was included Wave 1).

Additionally, Wave 2 GUiE matches Wave 1 GUiE records to administrative data collected locally by children's social care services. This information is provided to central government through the 'Children Looked After' (CLA) return and 'Children in Need' (CIN) census and held by Department for Education. The additional data made available through the

Wave 2 release is restricted to academic years 2010/11 to 2014/15. Our understanding is that Wave 2 release did not include historical information about lifetime care experiences or information about age of first care episode, both of which have been shown in previous studies to be important in understanding the relationship between experiences of care and educational attainment (e.g. Sebba et al 2015). However, additional data for previous years is being made available via Wave 3 (on which, see below).

We have been informed by ONS that the GUiE Wave 1 - Wave 2 linkage is a "join" between Wave 1 and Wave 2. In relation to the CIN/CLA data included in Wave 2, we understand that *all* CIN/CLA records held by DfE within the relevant time period are made available (and therefore that standalone analysis of the variables from the CIN/CLA data should match to analysis undertaken on the standalone datasets held by DfE for these years). However, linked census and educational attainment records are only available for individuals included in Wave 2 where there is a successful match to Wave 1 records. Therefore, analysis involving Wave 2 data in conjunction with education attainment and / or census data from Wave 1 will be restricted by Wave 1 eligibility and age restrictions.

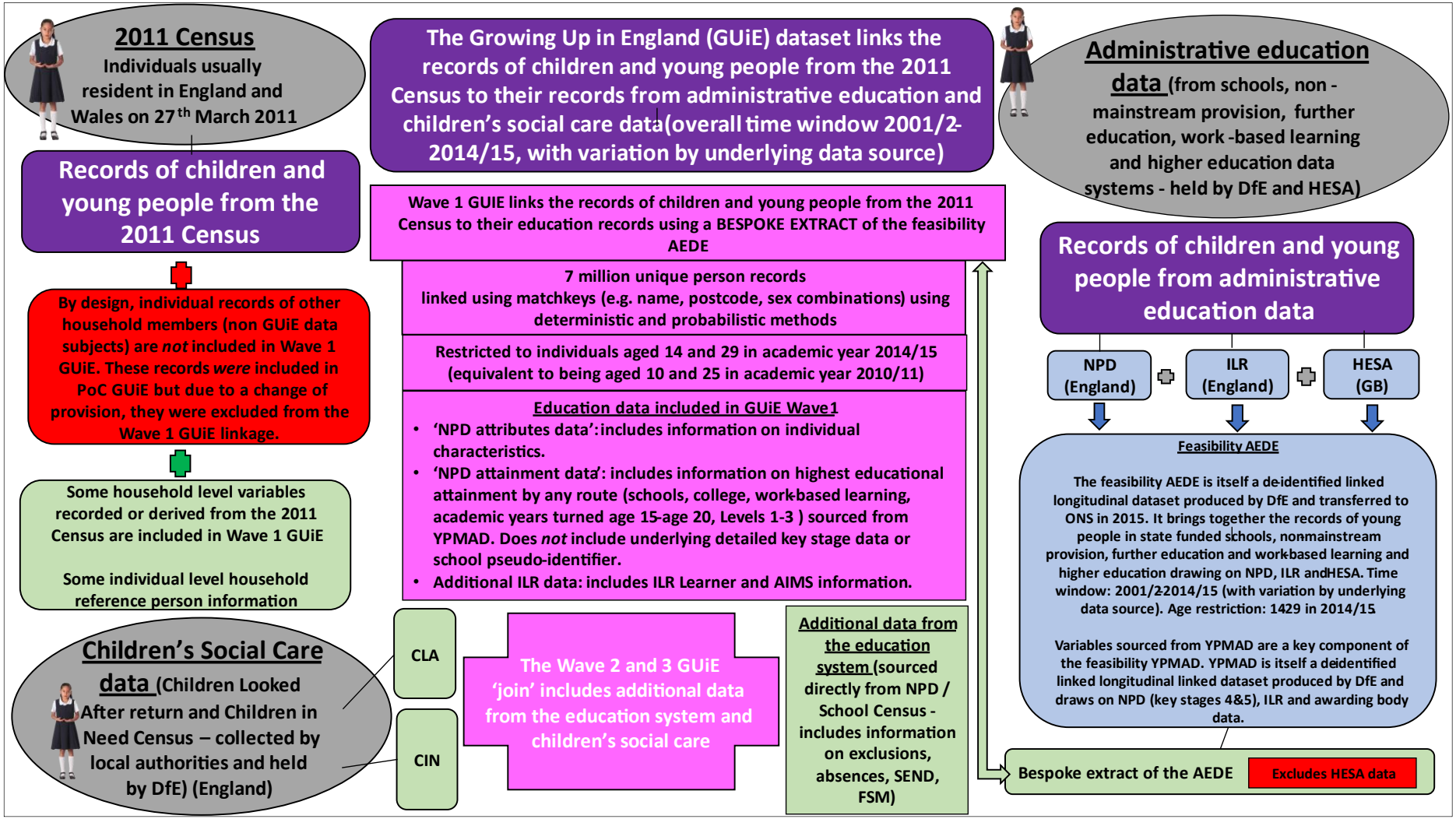
Wave 3 GUiE

Wave 3 GUiE is currently under development and will plug some of the gaps in GUiE Wave 2 (specifically, some terms of English School Census data are missing from Wave 2; and Wave 2 only covers academic years 2010/11-2014/15, whereas Wave 1 covers 2001/02-2014/15). Additionally, in the medium term, it is still intended that, as the final block of Wave 3, a future Wave 3 iteration will include the higher education records from the feasibility AEDE that have not been released in the GUiE research datasets to date. However, this further extension is not due imminently.

At the time of writing, elements of GUiE Wave 3 are already available in the SRS to all accredited researchers. The new release includes additional coverage for English School Census between the 2010/11 – 2014/15 academic years (Wave 3.1) and additional coverage back to academic year 2001/02 for data on Free School Meals, Special Educational Needs, Absences and Exclusions where the underlying data is available (Wave 3.2). Additionally, Wave 3.3 will similarly extend the Wave 2 Children in Need and Children Looked After dataset time window with the provision of additional historical data (Wave 3.3). At the time of writing, the Wave 3.3 data is being processed by the Data Engineering team.

We understand that the Wave 3 "join" adopts a similar methodology to that reported above in the context of Wave 2. Therefore, we understand that standalone analysis of the variables from the CIN/CLA data from Wave 3 should match to analysis undertaken on the standalone datasets held by DfE for the same years. Additionally, analysis involving Wave 3 data in conjunction with education attainment and / or census data from Wave 1 will be restricted by Wave 1 eligibility and age restrictions.

Figure 8: GUiE: conceptualisation v operationalisation (Waves 1-3)



5.3 Record of key issues relating to GUiE content and design that we have followed up during the lifetime of our project

In this section, we provide a record of some of the key issues related to GUiE content, design and quality that we have encountered and followed up with ONS and others over the life-time of our GUiE project. The LSE research team were the first academic researchers to access Wave 1 GUiE. When we first accessed Wave 1 GUiE in June 2021 there was no published dataset metadata or documentation around the dataset, meaning that variable lists could not be shared outside the SRS including with our Advisory Group, and linkage methodologies and the legal gateway for researcher access to GUiE was also complex. At that point in time, there were substantial information gaps relating to Wave 1 GUiE content and linkage methodology. To address issues relating to GUiE design, coverage and quality, the CASE team worked extensively with ONS through a system of meetings and feedback as well as through the newly created ONS GUiE technical and steering groups, broader ADR England Data Linkage Steering Group arrangements and with other stakeholders. Many of the issues we have raised have been subsequently clarified and addressed through the publication of the GUiE metadata and User Guide (ONS n.d.b); the publication of the GUiE methodology documentation and findings in Autumn 2022 (ONScdef); the release of Wave 2 GUiE; the release of the AEDE spine duplicates dataset; and the release of elements of Wave 3 GUiE (with the release of additional elements of Wave 3 being imminent at the time of writing). Researchers are referred to these broader sources of published information for official descriptions of GUiE content and design.

5.3.1 Record of follow up to early issues raised by NSDEC

The NSDEC Committee undertook an ethics review of our project in Spring 2021. As part of the review, it raised concerns relating to planned GUiE coverage, especially the failure of Wave 1 GUiE to include non-attainment education data, as being able to assess spells of missingness and drop out from school as well as exclusions and absences is of particular importance in the context of Gypsy, Traveller and Roma research. We highlighted these issues in our early discussions with the ONS GUiE team. Although we haven't been able to access the GUiE Wave 2 or Wave 3 data as part of our initial GUiE project, exclusions and absences data, together with SEN data, free school meals data, CIN and CLA data and additional Schools Census data, were included within the Wave 2 GUiE release. We also note that while Wave 2 coverage of this data is only for a short time window, for some variables time coverage has already been extended

through the Wave 3 release and for other variables extensions to time coverage are imminent.

Additionally, the NSDEC Committee raised concerns that Gypsy, Traveller and Roma children and young people would be disproportionately excluded from the linkage, as they are more likely to have spells of 'missingness' from school. These concerns are further addressed in section 6. However, we note here that in the context of GUiE Wave 1, individuals who have a record for *any* year in the feasibility AEDE time window and a census record are eligible to be included in the GUiE linkage. This methodology, which builds on the stock design of the feasibility AEDE, maximises inclusion of children who become disengaged from or drop out of formal education, have spells of missingness and / or experience exclusions and persistence absence. This is an important strength and it is critical that the structure of the feasibility AEDE as a cumulative stock is preserved in any future GUiE build.

5.3.2 Record of key issues raised in the context of early data exploration

Our early data exploration in summer 2021 resulted in several queries relating to GUiE content, coverage and quality. These broadly related to the following issues:

- The apparent presence of AEDE spine duplicates (our early data exploration identified duplicates in the AEDE spine).
- Wave 1 linkage methodology (year restriction): we had understood that the linkage required individuals to have a record that matched on one focus year academic year (2010/11), or three focus academic years centred on 2010/11, but early data exploration led us to query this understanding.
- Our inability to identify the individual level 2011 Census records of children and adults who do not themselves have matching records in the feasibility AEDE data. Our understanding had been that Census 2011 information would be made available about *all* individuals living in households with GUiE data subjects (including, for example, the individual level records of the parents of the young people who are GUiE data subjects). However, when we accessed the data, we were only able to identify the individual level records of GUiE Wave 1 data subjects - the individual records of non-data subjects did not appear to be available.
- The observation of cases that are outside of the expected GUiE age range.

- The need to clarify any implications of pre-processing of 2011 Census data for our analysis (especially, a record-swapping exercise was undertaken for Statistical Disclosure Control purposes to protect individual privacy; we requested further information about any impacts on smaller groups such as Gypsy, Traveller and Roma young people).

Below, we provide a record of clarifications received and our current understanding in relation to each of these issues.

Spine duplicates

As noted above, the issue of the AEDE spine duplicates was followed up in detail, with ONS undertaking further quality assessment including clerical review and the development of a regression model that predicts duplicate status. A new AEDE person spine dataset with an additional flag that predicts the duplicate status of each record using a true/false flag using the regression model has also been made available to researchers. Full details are provided in ONS (2022d). However, the underlying reasons for the presence of duplicates have not been fully resolved.

Scope of the 2011 Census data included in Wave 1 GUiE

Wave 1 GUiE includes a wide range of variables sourced from the 2011 Census. However, one key limitation is that only individual level records for data subjects (that is, for those individuals who meet the linkage criteria set out above) are included in GUiE Wave 1. Some 2011 Census household level information is also available in Wave 1 GUiE, including some deprivation characteristics and household level information on disability, and this is a major strength of the dataset. Additionally, some individual level information relating to the household reference person is included. However, the individual records of people living in the same household who are not GUiE data subjects - including parents, grandparents, carers and siblings who are not included in the linkage in their own right – are not made available in Wave 1 GUiE. In contrast, Proof of Concept GUiE included the individual records of all children and adults within households along with a person and household identifier and relationship data.

This change in provision in relation to the individual records of (non-data subject) adults and siblings whose individual level records are not themselves linked to the bespoke feasibility AEDE is a major weakness of Wave 1 and restricts its research potential. Indeed, an important part of the 'value added' and rationale of the GUiE concept is to make available the

rich parental and family information that is recorded in the 2011 Census. It is worth noting though that GUiE Wave 1 *does* include the census person and census household 'id' and the 'relat_XX' variable which were used (in conjunction with other variables) to derive household relationships for the Proof of Concept GUiE analysis. Where siblings are Wave 1 GUiE data subjects, with individual records included in Wave 1 GUiE in their own right, this may provide potential for sibling analysis using Wave 1 GUiE.

Ages that are outside the expected GUiE Wave 1 range

The design of the feasibility AEDE (discussed further below) means that the young people included in the Wave 1 linkage are expected to be aged 14-29 on 31 August 2015 (that is, at the end of academic year 2014/15) or equivalently aged 10 to 25 at the end of the 2010/11 academic year (the year 'closest' to the time of the 2011 Census snapshot). Given this design, we followed up the issue of unexpected ages (that is, cases where census recorded age was out of this range). The clarification provided suggests that the records affected by out-of-range ages were linked using matchkeys that did not include date of birth information. Three of the 91 matchkeys used in the linkage of the 2011 Census to the feasibility AEDE did not include date of birth. Records matched using these matchkeys are matched on variables other than age, such as home postcode. Therefore, there are a small number of records outside of the expected age range (aged 10 to 25 years at the start of the 2011 to 2012 academic year). Some of these records are likely to be incorrect matches between the 2011 Census and the feasibility AEDE, however others will be correct matches. Overall, however, the GUiE match rate meets Office of National Statistics threshold requirements (c.f. ONS 2022d).

Wave 1 linkage year restriction

We also requested further information relating to the year 1 linkage year restriction, as we had understood that the linkage required individuals to have a record that matched on one focus year academic year (2010/11), or three focus academic years centred on 2010/11, but early data exploration led us to query this understanding. Information provided clarified that it was not necessary to have a 2010/11 record either in the feasibility AEDE, or within three focus years centred on 2010/11, to be included in the Wave 1 linked dataset.

Impact of record swapping on subgroup analysis

We also requested information relating to the 2011 Census Statistical Disclosure Control record swapping exercise. This was undertaken to protect against disclosure of personal information and involves the swapping of records of households with unusual characteristics in small areas. Follow up by the ONS GUiE team established that detailed information about the record swapping exercise could not be disclosed to protect against identification of individuals, so we were unable to assess its impact on Gypsy, Traveller and Roma records. However, published information was identified that shows that the record swapping was carried out within a relatively large geographical area (within local authorities) and that households were swapped in and out of small geographical areas (output areas) (Longhurst et al n.d.). As we do not report findings below the local authority level within this report, we believe that there would not be any impact on our analysis.

5.3.3 Record of key issues and clarifications relating to the content and design of the feasibility AEDE and YPMAD

When the team first accessed the Wave 1 GUiE dataset, there were substantial information gaps relating to the feasibility AEDE side of the dataset. Broadly speaking, these included:

- General information gaps relating to the content and quality of the underlying feasibility AEDE dataset; clarifications of the age restrictions imposed on the feasibility AEDE and any follow-on implications for GUiE Wave 1 coverage and eligibility.
- Specific information gaps around the content of the GUiE Wave 1 educational attainment data (which was labelled as NPD data but was neither key stage 4 or key stage 5 data).

Below, we provide a summary of our current understanding of the content and design of the feasibility AEDE side of the dataset, bringing together information that is in the public domain and clarifications that we have received from ONS and others during the life-time of our project.

Clarifications on NPD attainment data

The term 'National Pupil Database' refers to a wide range of individual administrative education datafiles and includes information relating to all

key stages (key stage 1-5) as well as to reception and early years. These datafiles are listed in the Department for Education's National Pupil Database Webtool (Figure 10). Wave 1 GUiE only includes data from a subset of the datafiles that are available through the National Pupil Database. It draws on student data from the English Schools Census (Spring only) and non-mainstream educational provision (the Pupil Referral Unit and Alternative Provision censuses); and data on individuals in further education and work-based learning (held within Individual Learner Records and Learner Aims data). Additionally, Wave 1 GUiE draws directly on data from the Young Persons Matched Administrative Dataset (YPMAD) which is itself a derived dataset that looks across multiple underlying NPD data files. Further details of YPMAD are provided below. The highest educational attainment data included within the Wave 1 GUiE data is sourced from YPMAD.

Note that the education data made available via GUiE Wave 1 does *not* include the standard NPD key stage 1 – key stage 5 data files. This restricts information on qualifications and subjects that is available in the GUiE dataset. Additionally, GUiE does not currently include a school identifier (including a pseudo-identifier). Therefore, school level analysis is *not* possible using Wave 1 GUiE. This excludes analysis of students within school settings, adjustments for school type and adjustments for school clustering effects based on pupil characteristics such as deprivation. These limitations make GUiE less useful for some research and policy purposes. For example, it is not possible to examine key stage 2-4 progression or to control for school composition or other school level effects in the context of multivariate analysis[1].

Also note that additional educational attainment data is included within the GUiE Wave 1 ILR data files.

[1]These limitations also feed through to restrict the research potential of Wave 2 GUiE. For example, controlling for school composition and progression have been identified to be particularly important in the context of assessments of the relationship between educational attainment and time spent in social care (Sebba et al 2015). The limited historical information included in Wave 2 was therefore a concern, as age of first care experience and lifetime care experiences are also important to control for in multivariate assessments of the impact of care episodes on educational attainment. Again see Sebba et al (2015) for the crucial difference this makes to overall narratives relating to the concerning the protective versus the negative / adverse role of care in explaining educational inequalities. The extension of CIN/CLA data and potential for analysing historical information in GUiE Wave 3.3 will be an important advance for these reasons.

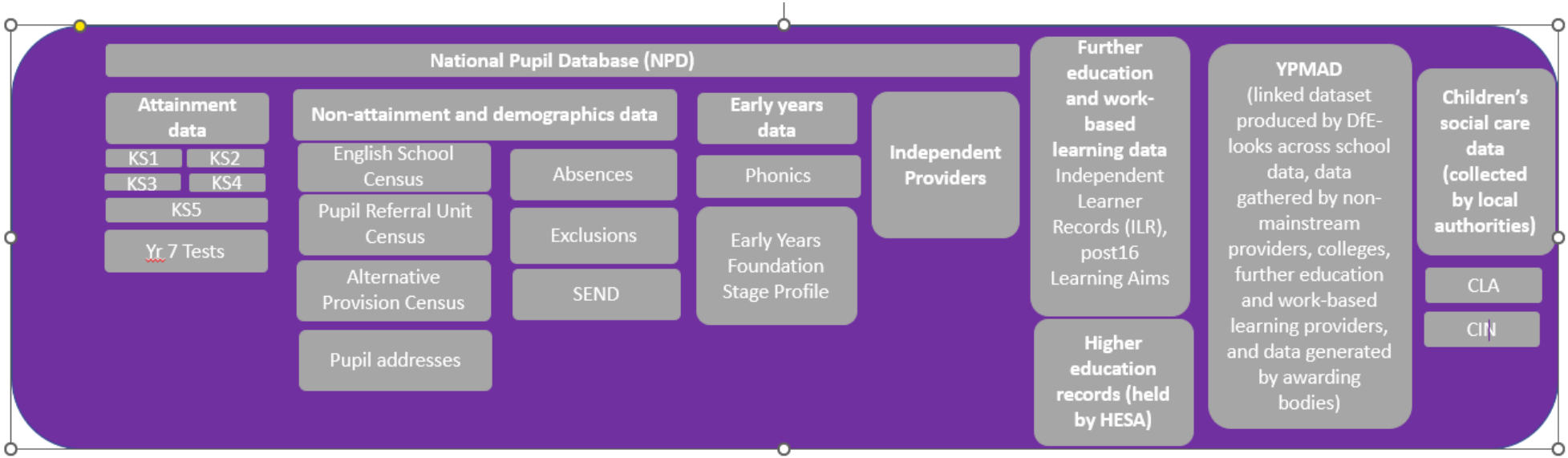
Clarifications relating to YPMAD

Although YPMAD is listed on the NPD data tool, information provided by Department for Education (Connell 2022a) suggests that it is less widely used by education researchers than standard key stage files, such as the pupil attainment data that is made available through the key stage 4 and key stage 5 NPD datafiles. Critically, YPMAD is itself a linked longitudinal dataset produced by DfE which is specifically designed to provide analysis of 16-18 cohorts and supports tracking of individual participation, pathways and attainment from the academic year in which a student is 15 to the academic year in which a young person is 20 (academic ages 15 to 20). Connell (2022a) lists the three major underlying data sources of YPMAD as:

- School census learner and aims data: information on post-16 participation and personal characteristics of pupils in state schools at age 15.
- Awarding organisation data: used for attainment and participation for learners not included in the school census or ILR participation data e.g. independent schools).
- Individualised Learner record (ILR): learners, participation and learning providers in the government funded education and skills sector.

A key strength of YPMAD is that the dataset provides derived information on attainment looking across multiple 16-18 pathways including sixth forms, further education and work-based learning in England and includes both GCSE and A level information and equivalent and vocational qualifications such as BTEC and vocational qualifications and apprenticeships. This comprehensive information on highest educational attainment data derived from YPMAD opens up a series of important research questions that can be pursued using Wave 1 GUiE, looking across further education and work-based learning as well as educational participation and attainment within school sixth form settings. Our understanding is that the use of YPMAD as a base for the feasibility AEDE reflects the original purposes of the creation of the feasibility AEDE, to create a dataset to support an ONS research exercise on the consistency of different sources of information on highest educational qualifications.

Figure 9: Administrative data from the education and children's social care system



Source: uses the categories listed in the National Pupil Database online tool, available here: [Find and explore data in the National Pupil Database - GOV.UK (education.gov.uk)]

YPMAD is used by Department for Education to produce national statistics on Level 2 and Level 3 attainment by age 19 and the dataset incudes binary flags that identify participation, pathways and attainment by age, including highest educational qualification. The attainment variables distinguish between Level 1 qualifications (one to four GCSEs at any grade) or equivalent; Level 2 qualifications (five or more GCSES grades A* to C or equivalent; Level 3 qualifications (two or more A levels or equivalent). The dataset provides information on highest educational qualification by age. For example, achieving Level 2 by age 16 refers to the attainment of learners typically in year 11 and Level 3 attainment by age 19 refers to the attainment of learners in the academic year after they typically complete school.

Clarifications relating to feasibility AEDE linkage methodology

Having an adequate understanding of the design, content, limitations and age restrictions of the feasibility AEDE is critical for having an adequate overall understanding of the GUiE Wave 1 build. For example, the feasibility AEDE time window, age restrictions, excluded sub-populations as well as the content of the feasibility AEDE education attainment data all feed through into overall Wave 1 GUiE content and design and determine and limit the research potential of GUiE Waves 1-3 in critical ways.

Unfortunately, detailed documentation with detailed descriptions of the design, content and quality assessment of the feasibility AEDE dataset have *not* been identified and it is not possible for researchers to access the feasibility AEDE – the starting point of the GUiE build – directly themselves. However, overview information about the feasibility AEDE dataset is provided in several ONS publications (ONS 2019, 2020a, n.d.b, 2022d, 2023). This information is summarised in Figure 10. In terms of time windows, note that while the pupil data covers the period 2001/2-2014/15, the further and higher education data cover the period 2002/3-2014/15. The geographical coverage of the data is also different, with pupil and further education / work-based learning data covering England while the higher education data covers Great Britain.

We believe that the feasibility AEDE population can be best conceptualised as individuals who are recorded as attending government funded schools in England between 2001/2 and 2014/15, or participating in further education or work based learning in England between 2002/3 and 2014/15, or higher education institutions in Great Britain between 2002/3 and 2014/15 and are recorded in the relevant underlying data sources AND who were aged between 14 and 29 years on 31 August 2015 (that is, at

the end of academic year 2014/15, where age refers to age as recorded in the feasibility AEDE). However, note that there is some ambiguity in relation to excluded sub-populations (including in relation to students in private schools and home educated young people) - further details are discussed below.

It is particularly critical to understand the age restriction that was imposed when the feasibility AEDE was created, as this determines the current scope and research potential of the GUiE Wave 1 dataset (and has follow on implications for analysing GUiE Wave 2 and Wave 3 data in conjunction with Wave 1 census and education attainment data). Specifically, the feasibility AEDE was restricted to children and young people aged between 14 and 29 years on 31 August 2015 (that is, at the end of academic year 2014/15 - where age refers to age as recorded in feasibility AEDE records). Additionally, the YPMAD data file is restricted to ages 14-20 and analysis that uses GUiE Wave 1 educational attainment data sourced from YPMAD is further restricted by this age range.

The feasibility AEDE linkage was undertaken by the Department for Education *prior* to transfer to the ONS in 2017. While detailed linkage information has not been identified, the available overview documentation suggests a high matching success rates averaging 95% with some variation between different data sources and cohorts. One underlying means of identifying young people in education data is the unique pupil number (UPN) which is a code that is allocated to all pupils in the state-maintained school system and identifies all pupils in administrative data records. Additionally, anonymised personal identifiers were produced by DfE prior to transfer of the dataset to ONS, so that individuals can be tracked longitudinally and across the different component parts of the feasibility AEDE dataset (ONS 2022d).

Clarifications relating to sub-populations excluded from the feasibility AEDE linkage

At the time of writing, there remains some ambiguity in relation to the subpopulations that are excluded from the feasibility AEDE. The information on feasibility AEDE coverage included in ONS (2019) suggests that the dataset provides information on the socio-demographic characteristics and educational qualifications for individuals who attended government-funded schools, further education and higher education in England. In assessing sub-populations that are excluded from the feasibility AEDE linkage, this document further notes that those who have never interacted with government-funded education will not be included in the feasibility AEDE, for example, those who have only ever attended independent schools or

institutions or those who have been home-educated outside of local authority provision. It reports that around 86,000 individuals aged 16-18 attended independent schools in England in 2011, representing around 4% of all individuals of that age[2]. It further notes that while data on the number of home-educated children who receive no funding from government are not collected centrally, figures from a Freedom of Information response from 86 out of 152 councils suggest that in 2011, around 15,000 children were home educated, representing around 0.2% of all individuals aged 5 to 16 years. Finally, ONS (2019) notes that for older learners, data for school activity is not held in a way that enables consistent linkage with the other datasets; older learners are recorded in HESA and ILR data but cannot be linked back to their school record.

However, while it is currently agreed that information on private school and home educated students would not enter into the feasibility AEDE through pupil data (the School Census), the most recent available information identifies that it is possible that some information on students in private schools / home educated young people may enter into the linked dataset through attainment data in YPMAD sourced from awarding bodies. Information provided by DfE states that the YPMAD dataset primarily captures information for *state funded providers* but that some additional attainment information from students who attend private schools may also be included in the data from awarding bodies (Connell 2022a, ONS 2022d). By analogy, it seems reasonable to conclude that attainment information on home educated students may be present in YPMAD attainment data. Additionally, as noted in ONS 2022d and above (section 5.1.2), further education colleges are required to return ILR data for all leaners (including those *not* funded by government) – consequently, some learners not funded by government are included in ILR data.

ONS (2019) highlights the understanding of the feasibility AEDE as a *stock* dataset (that is, a cumulative stock estimate of all persons in the AEDE age cohort). This means that once individuals are included in the dataset they remain in the dataset even if they have died or emigrated (ONS 2019). This stock understanding of AEDE has important implications in the context of the current project and is a key strength. The implication is that the records of individuals that are initially included in AEDE *remain* in the stock data even if they subsequently drop out of school, are temporarily disengaged from school, are excluded or persistently absent, or elect for home education. This design goes some way to addressing some

[2] Other estimates suggest that across age groups, 7% of pupils attended private schools in England in 2015. See Green et al (2017) Figure 1b, using data from the School Census (state schools) and School Level Annual School Census (independent schools).

of the concerns raised by the NSDEC Committee as part of our project ethics review (for details, see section 1 ‘ethics and engagement’).

Figure 10: Summary of available information relating to the design and content of the feasibility All Education Dataset for England (AEDE)

Overall description: the feasibility AEDE is a large-scale longitudinal education dataset that was created by Department for education. It includes socio-demographic and attainment data and brings together administrative records from three main sources:

1. **Student records sourced from the National Pupil Database (England).** These records cover pupils in schools maintained by local authorities, academies and free schools in England, as well as those in non-mainstream provision such as Pupil Referral Units (PRUs) and Alternative Provision in an academic year. This data is sourced from multiple underlying administrative data collections including the English Schools Census; returns from Pupil Referral Units; and the Alternative Provision Census. Additionally, the feasibility AEDE draws on the Young Person's Matched Administrative Dataset (YPMAD), which is itself a large scale longitudinal matched administrative dataset produced by Department for Education from multiple underlying data files.
2. **Further education and work-based learning records sourced from Individual Learner records (England).** All providers of further education in England including publicly funded colleges, training organisations, local authorities and employers acting as FE providers are required to return ILR data for learners funded through government funding. Further education colleges are required to return ILR data for all leaners (including those *not* funded by government) – consequently, some learners not funded by government are included in ILR data. Higher Education Institutes (HEIs) must also return ILR data for adult learners who receive publicly funded loans. HEIs are also required to return ILR data for learners funded through Advanced Learner Lonas (adults who receive loans to cover tuition fees for a range of courses including A levels, general and vocational qualifications, and access to Diplomas of Higher Education.
3. **Higher education records sourced from the Higher Education Statistics Authority HESA (Great Britain).** All government-funded higher education institutes are required to send data to HESA. Additionally, further education bodies that deliver higher education send similar information to HESA. The HESA data excludes students studying outside the UK.

Time window: the feasibility AEDE time window is generally specified as 2001/2 to 2014/15. However, the within this time overall window, time periods covered vary by component underlying dataset, with the pupil data covering the period 2001/2-2014/15 and the further and higher education data covering the period 2002/3-2014/15.

Age coverage: the feasibility AEDE was restricted to children and young people aged between 14 and 29 years on 31 August 2015 (that is, at the end of academic year 2014/15 - where age refers to age as recorded in feasibility AEDE records).

Population: the feasibility AEDE population can be broadly conceptualised as individuals have a record from government funded school or non-mainstream / alternative provision in England between 2001/2 and 2014/15; or further education institutions (or government funded work-based learning or apprenticeships) in England between 2002/3 and 2014/15; or government funded higher education institutions in Great Britain between 2002/3 and 2014/15.

Excluded subpopulations:

- Individuals within the relevant age range who do not have a pupil, further education or higher education record in the underlying data sources / within the relevant time windows.
- The pupil data covers pupils in government funded education; pupils who have not interacted with government-funded education within the relevant time window, for example, those who only attended independent schools, or who were missing from school or who are electively home-educated within the relevant time window, would not be included. However, it is possible there is some coverage of private school students and home educated data through the attainment data derived from awarding bodies.
- Linked data: for older learners, data for their school activity is not held in a way that enables consistent linkage with the other datasets; older learners are recorded in HESA and ILR data but cannot be linked back to their school record.

Stock dataset: the feasibility AEDE is a stock dataset – the records of individuals who have died, migrated or dropped out of school remain in the feasibility AEDE.

Note: the information in Figure 9 is sourced from: ONS 2019, ONS 2020a, ONS 2022cde, ONS 2023.

5.4 Our working summary description of GUiE Wave 1-3

Building on the information summarised above, in this section we set out our working summary of the basic building blocks of GUiE.

Feasibility AEDE

The starting point of the GUiE build is the feasibility AEDE. Our working definition of the feasibility AEDE population is that individuals are included in the bespoke extract of the feasibility AEDE if:

- They have a record from government funded school or non-mainstream / alternative provision in England between 2001/2 and 2014/15; or further education institutions (or government funded work-based learning or apprenticeships) in England between 2002/3 and 2014/15; or government funded higher education institutions in Great Britain between 2002/3 and 2014/15.
- They were aged between 14 and 29 years on 31 August 2015 at the end of academic year 2014/15 (where age refers to age as recorded in feasibility AEDE records).

The feasibility AEDE draws on the following underlying datasets:

- **NPD**: Administrative records from the English Schools Census, PRU census and the Alternative Provision census held by Department for Education.
- **YPMAD** (data from a matched administrative data set derived by Department for Education which provides summary information on highest educational attainment in an academic year looking across attainment in key stage 4, key stage 5 and alternative pathways)
- **Independent Learner Records** (ILR dataset), including records from further education colleges and other publicly funded providers of further education, work-based learning and apprenticeships.
- **Higher Education records** held by the Higher Education Statistics Authority (HESA).

Bespoke extract of the feasibility AEDE used in the GUiE Wave 1 linkage

As noted above, Wave 1 GUiE does not include *all* of the data that is included in the feasibility AEDE (listed above) but rather links 2011 Census data to a *bespoke extract* of the feasibility AEDE. The bespoke extract of the feasibility AEDE used in the GUiE Wave 1 linkage includes the first three datasets listed above but excludes the higher education (HESA) dataset.

GUiE Wave 1

GUiE Wave 1 links a bespoke extract of the feasibility AEDE to the 2011 Census. Individuals are included in the Wave 1 GUiE linkage if:

1. They satisfy the criteria for inclusion in the bespoke extract of the feasibility AEDE set out above.
2. They have a record in the 2011 Census;
3. Their record in the bespoke extract of the feasibility AEDE has been successfully matched to their record in the 2011 Census as part of a record matching exercise undertaken by ONS.

Again as noted above, the feasibility AEDE age restriction means that AEDE records are available for students who were aged between 14 and 29 years on 31 August 2015, the end of academic year 2014/15 (where age refers to age as recorded in feasibility AEDE records). Note that this is equivalent to students being aged between 10 and 25 four years earlier on 31st August 2011 at the end of academic year 2010/11 (the academic year during which the 2011 Census took place). Also note that in practice, some GUiE observations fall outside of the expected age range. See section 5.3.2 for further details / explanation.

It should also be noted that at the time of writing, there remains some ambiguity in relation to whether young people in private schools / who are home educated are excluded from the linkage. Although the pupil data in the bespoke extract of the feasibility AEDE is restricted to government funded schools and non-mainstream provision records, it is possible that some data relating to individuals in private schools or who are home educated feeds through into the linkage via the attainment data from awarding bodies included in YPMAD. See section 5.3.3 for further details / explanation.

The GUiE Wave 1 release includes the following datafiles:

- **NPD attributes data:** sociodemographic information (including ethnicity) sourced from the English Schools Census (Spring only) and the Alternative Provision and Pupil Referral Unit censuses.
- **NPD attainment data**: attainment data sourced from YPMAD, looking across attainment in key stage 4, key stage 5 and alternative pathways (including by drawing on awarding bodies data). Note that the NPD attainment data file is restricted to 14 to 20 year olds.
- **Additional ILR data**: ILR Aims data, ILR Learner data and additional ILR datafiles.
- **2011 Census data** (2011 Census individual records for data subjects; selected derived household level variables; some household reference person information; household identifier; relationship information).
- **AEDE data and person spines.**

The datafiles require merging before analysis using the linked Wave 1 GUiE dataset can proceed. Pseudonymized person identifiers are included in datafiles to enable users to track individuals. The AEDE residuals and the 2011 Census residuals (i.e. non-matched residual records from each side of the linkage) were *not* made available as part of Wave 1 GUiE meaning that patterns and disproportionality in unmatched cases cannot be directly assessed by researchers.

GUiE Waves 2 and 3

As noted in section 5.2.3, GUiE Waves 2 and Wave 3 provide additional educational data (information on exclusions and absences; additional information derived from the English school census; and information about SEND) and information on interactions with children's social care services (recorded in the CIN and CLA datasets). Wave 2 provides data for 2010/11-2014/15. Wave 3 plugs some gaps in the Wave 3 2010/11-2014 provision and provides a longer historical run of both the additional education and children's social care data (going back to 2001/2 where data availability allows). We have been informed by ONS that the GUiE Wave 1 - Wave 2/3 linkage is a "join" between Wave 1 and Waves 2 and 3.

In relation to the CIN/CLA data included in Waves 2 and 3, we understand that *all* CIN/CLA records held by DfE within the relevant time period are made available (and therefore that standalone analysis of the CIN/CLA data should match to analysis undertaken on the standalone datasets held by DfE. Similarly, the English School Census data included

within Wave 2 and Wave 3 should match DfE records when analysed on a standalone basis. However, linked 2011 Census and educational attainment records are only available for individuals included in Waves 2 and 3 where there is a successful match to Wave 1 records. Therefore, any analysis using GUiE Wave 2 and / or Wave 3 data in conjunction with GUiE Wave 1 education attainment and / or 2011 Census data will be restricted by the same eligibility and age restrictions that apply to Wave 1 analysis.

6 Further assessment of the strengths and limitations of the Growing Up in England (GUiE) Wave 1 dataset for research on the lives, circumstances and needs of young people from the Gypsy, Traveller and Roma communities

This section assesses provides further details of the strengths and limitations of the Wave 1 Growing Up in England (GUiE) dataset for undertaking research on the lives, outcomes and needs of Gypsy, Traveller and Roma children and young people. We begin by highlighting the rich fine-grained ethnicity information that is available in Wave 1 GUiE as a major strength and breakthrough in terms of supporting the identification of the records of Gypsy, Traveller and Roma young people at scale. We then address a series of caveats and limitations relating to missingness, invisibility and inclusivity and caveat the extent to which the dataset can be viewed as being representative of the underlying population of Gypsy, Traveller and Roma children and young people. Next, we provide detailed statistical assessments of the extent to which the records of Gypsy, Traveller and Roma young people are successfully matched into the Wave 1 dataset on an equal basis and the extent to which the linkage with the 2011 Census as associated with deprivation-related compositional effects. Finally, we draw attention to the importance of interpreting statistical findings on Gypsy, Traveller and Roma young people produced using the GUiE dataset in the light of these caveats and limitations.

6.1 Addressing missingness: identifying the records of Gypsy, Traveller and Roma young people at scale using Wave 1 GUiE

The Wave 1 Growing Up in England (GUiE) dataset combines the ethnicity information that is available in two of the best existing data sources we have for identifying the records of children and young people from the Gypsy, Traveller and Roma communities: data from the administrative education system, and data from the 2011 Census. This is the major strength of the dataset from the point of view of Gypsy, Traveller and Roma research and is a major breakthrough in terms of supporting the identification of much larger analytical samples than has previously been possible in research on Gypsy, Traveller and Roma children and young people.

As we highlighted in section 3, the inclusion of relevant ethnicity identifiers within education administrative systems is an area of (relative) good practice – particularly when compared with the complete historical

absence of Gypsy, Traveller and Roma ethnicity identifiers within the NHS data dictionary. Irish Traveller and Gypsy/Roma categories have been included in the English School Census from the early noughties and fine-grained ethnicity information is made available through the ethnicity 'minor' variable in the NPD dataset. Further education administrative systems include a Gypsy/Irish Traveller identifier which is made available in the ILR dataset.

Additionally, again as highlighted in section 3, the introduction of the Gypsy or Irish Traveller ethnicity identifier in the 2011 Census was an important step forward in the evolution of inclusive infrastructure. The introduction of the new ethnicity identifier was moreover coupled with pro-active measures to maximise participation including measures to engage with local authorities and civil society organisations and to obtain as comprehensive a list as possible of authorised and unauthorised sites; training of enumerators; hand-delivery of questionnaires to sites; and follow up of non-response (Burchardt et al 2018). These were important advances and have provided a basis for the identification of Gypsy and Irish Traveller group in statistical findings drawn from the 2011 Census over the last decade.

The research exercise we report in section 7 demonstrates that the rich and fine-grained ethnicity information included in the Wave 1 GUiE linkage supports the identification of the records of Gypsy, Traveller and Roma children and young people at scale. The research exercise exploits the unprecedented large analysis samples that are available in the Wave 1 GUiE dataset and we establish that using pooled ethnicity information derived from the different component parts of the integrated GUiE dataset improves Gypsy, Traveller and Roma identification when compared to using ethnicity information from each of these underlying datasets separately. Moreover, the longitudinal nature of the education data in GUiE opens-up a substantially longer time window for Gypsy, Traveller or Roma identity to be revealed. Exploiting these different dimensions of the GUiE Wave 1 data enables us to boost analysis sample size substantially, overcoming the challenges of small sample size typically encountered in Gypsy, Traveller and Roma research.

Note that a further key strength of the GUiE Wave 1 education data that supports the identification and analysis of the records of Gypsy, Traveller and Roma records at scale relates to its coverage of alternative pathways including further education, work-based learning systems and apprenticeships, as well as schools. Our analysis of educational pathways in section 7, as well as the feedback from the direct engagement workshops we present in the section that follows, highlights that information on college

and vocational pathways is particularly valuable, informative and relevant in the context of research on Gypsy, Traveller and Roma young people. Previous research also shows that 'low attainers' within overall cohorts of young people are more likely in 16-18 education to be in college or work-based learning, and less likely to be in schools, than there higher attaining peers (Lupton et al 2021). This broad education data is important both for substantive analysis but also for analytical sample size, as the number of Gypsy, Traveller and Roma students identified within A level cohorts is typically very low (see for example the frequency counts in the A level column in Table 1 and the related discussion in section 3.3.1).

6.2 Continued concerns around missingness, invisibility and inclusivity

Nevertheless, while the GUiE Wave 1 linkage represents a major break-through in terms of Gypsy, Traveller and Roma research potential and identifying and analysing the records of Gypsy, Traveller and Roma children and young people at scale, important caveats and limitations relating to missingness, invisibility and inclusivity remain. For the reasons set out below, these restrict the extent to which the GUiE Wave 1 dataset can and should be conceptualised as being representative of the underlying population of Gypsy, Traveller and Roma children and young people.

6.1.2 Limitations of Gypsy, Traveller and Roma recording in the 2011 Census and administrative education systems

While administrative education data and data from the 2011 Census are two of the best sources of evidence we have on Gypsy, Traveller and Roma young people, it is nevertheless important to acknowledge that ethnicity recording in both of these underlying data sources is limited in important respects. These limitations feed into the GUiE Wave 1 linkage, resulting in ongoing concerns around both 'missingness' and 'invisibility'.

In relation to the 2011 Census, despite the pro-active measures taken at the time of the 2011 Census, it is acknowledged that not all sites or Gypsy, Traveller or Roma people participated in the exercise, resulting in gaps in inclusion and a substantial undercount of Gypsy, Traveller and Roma people. Further details on the overall 2011 Census person response rate and item-response and imputation in relation to the 2011 Census ethnicity question are provided in the Appendix. Additionally, an important caveat to the discussion above is that a pre-coded Roma option was *not* included in the 2011 Census questionnaire. While a small number of

respondents entered Roma or Roma combinations within an open response field (Burchardt et al 2018), the Roma population is essentially invisible within the 2011 Census records and in evidence and official statistics that have been derived from the 2011 Census. This limitation was addressed in the 2021 Census, which includes a new 18+1 ethnicity classification standard with a separate Roma tick-box.

Additionally, the ethnicity classifications used in administrative education pupil data were not harmonised with the 2011 Census ethnicity standard. The attainment data within the feasibility AEDE dataset includes the National Pupil Database ethnicity classifications Gypsy/Roma and Traveller of Irish Heritage – which means there is at least some basis for identifying the Roma subpopulation. However, the pupil data ethnicity classifications are inconsistent with the 2011 Census identifier. Moreover, as noted earlier in this report, the use of aggregated Gypsy / Roma categories has been critiqued as inappropriate as Gypsy and Roma people are heterogeneous groups with distinct circumstances and different needs. For example, Roma people are more likely to be recent migrants, to have English as a second language and are highly unlikely to be nomadic (see for example, the critique in Mulcahy et al 2017, the Traveller Movement 2022). More generally, the Race Disparity Unit (2020) identified that aggregated ethnic groups can mask substantial differences between constituent detailed groups while the Commission on Race and Ethnic Disparities (2021) recommended a move away from the use of aggregated ethnic categories.

These complexities are further exacerbated by use of different ethnicity classifications within different parts of the education system. Specifically, further education administrative systems use a Gypsy or Irish Traveller identifier, which is consistent with and has the same limitations as the 2011 Census, and is inconsistent with the NPD pupil data. The inconsistency of Gypsy, Traveller and Roma classifications across different datasets has been highlighted as problematic in the Cabinet Office Race Disparity Audit and follow up analysis (Cabinet Office 2018, 2022) and the UK Statistics Authority Inclusive Data Taskforce (UKSA 2021ab). Similarly, the Commission on Race and Ethnic Disparities (2021) also recommended the consistent use of harmonised ethnicity classifications across datasets.

6.1.3 Non-disclosure of Gypsy, Traveller and Roma identity

As discussed in sections 2 and 4 and in the feedback from our engagement workshops discussed in section 8, there are a multiple reasons why Gypsy, Traveller and Roma people might not disclose their ethnic

identity in administrative data collections. Ongoing day-to-day experiences of discrimination and hostile treatment, as well as historical persecution and distrust of government and the purposes of official data collection, and previous experiences of data misuse, mean that Gypsy, Traveller and Roma people can be reluctant to disclose their identity in their interactions with government and public services and in data collections. In the context of schools, concerns around bullying and harassment can also mean that children and families are reluctant to disclose their identity. Non-disclosure of identity due to fear of adverse consequences is well documented in previous research and has been highlighted by civil society organisations and in several of the high-level inquiries and reviews discussed earlier sections of this report (e.g. Equality and Human Rights Commission, 2016; Women and Equalities Committee 2019, Friends, Families and Travellers, 2022a; Traveller Movement 2022). If identity is not revealed in data collections, then Gypsy, Traveller and Roma people cannot be separately identified as a group in research analysis, even when they are included in the dataset.

6.1.3 Other coverage restrictions

The NSDEC committee that reviewed our project raised concerns that administrative education data may fail to adequately represent the underlying population of Gypsy, Traveller and Roma young people due to 'missingness' from school and the impact of exclusions and absences. The evidence review in section 3 identified that rates of non-participation in and drop-out from formal education, as well spells of missingness from school and rates of exclusions, persistent absence, are substantially higher amongst Gypsy, Traveller and Roma children and young people than amongst their peers. These higher rates of non-participation and drop-out occur at an increasing rate with age, particularly from the secondary school stage, while rates of participation in 16+ education and higher education further are both substantially lower. Rates of mobility and school moves are also higher than amongst the general population. Civil society organisation advocacy, qualitative research, official statistics and analysis by the Women and Equalities Committee (2019) and the House of Commons Education Committee (2022) all identify relatively high rates of elective home education amongst children and young people from the communities.

These factors are all likely to impact on the inclusivity of the GUiE data linkage to a certain extent. At the very least, Gypsy, Traveller and Roma people young people who do not have a 2011 Census record, or who do not

have a feasibility AEDE education record, are not included in the linkage. As noted above, notwithstanding the pro-active engagement measures put into place, Gypsy, Traveller and Roma participation in the 2011 Census is known to be substantially lower than that for other groups. Additionally, recent migrant children and young people who have a school record and are included in the feasibility AEDE but who were not resident at the time of the 2011 Census would also be excluded from the linkage.

Gypsy, Traveller and Roma children and young people are also likely to be excluded from the feasibility AEDE at a greater rate than other children and young people. Administrative education data captures information about individuals who interact with education services and children and young people who do not interact with education services will not have a record in the feasibility AEDE. From the point of view of Gypsy, Traveller and Roma research, it is particularly important to understand whether home education students are included in the GUiE linkage. As we noted in section 5, there remains some uncertainty in relation to the subpopulations that are entirely excluded from the GUiE Wave 1 linkage including in relation to students who have always been home educated or only attended independent schools. While Gypsy, Traveller or Roma children and young people who are home educated or have never registered at a school would not be included in the English School Census, the question of whether their attainment data may enter the GUiE linkage through the YPMAD dataset remains has not yet been fully resolved (c.f. section 5.3.3 for further details). Additionally, as we discuss below, difficulties with the unique pupil reference numbers used to track individuals across different datasets also arise and can be exacerbated by factors such as mobility and moves between schools, and children from the communities are likely to be more affected. These difficulties may result in children and young people not being included in the feasibility AEDE data (which was pre-processed and matched by Department for Education, prior to transfer to the Office for National Statistics).

On a positive note, two of the Wave 1 GUiE design features identified in section 5 may help to mitigate against 'missingness'. Specifically, the feasibility AEDE is a stock dataset – that is, individuals remain in the feasibility AEDE dataset if they fall within the relevant age range and have an eligible education record, even if for example they leave a school, move between schools, leave the country, drop out of school, are excluded or persistently absent. Additionally, the Wave 1 GUiE linkage inclusion criteria do *not* require children to have a record in any particular year (such as 2010/11) or in any particular time window (such as 2009/10-2011/12).

Figure 11: Potential of GUiE Wave 1 for challenging data exclusion and building up new data and evidence on children and young people from the Gypsy, Traveller and Roma

Strengths

Underlying data sources

- 2011 Census included a Gypsy/Irish Traveller identifier for the first time
- Administrative education records include Gypsy/Roma and Irish Traveller identifiers
- Administrative further education records include Gypsy/Irish Traveller identifiers

Integrated dataset
Wave 1 GUiE

- Supports identification of the records of Gypsy, Traveller and Roma children and young people at scale, overcoming the challenge of small sample sizes
- Supports pooling of ethnicity information across underlying data sources, improving identification and addressing 'invisibility'
- Enables the educational attainment of children and young people from the communities to be followed longitudinally
- Covers alternative pathways including further education, work-based learning and apprenticeships as well as schools
- Supports analysis of child multidimensional deprivation (including identification of caravan or mobile structure housing status)

Waves 2 / 3 GUiE

- Supports additional analysis of variables such as exclusions, absences and SEND which are critically important for understanding the educational participation and progression of Gypsy, Traveller and Roma young people
- Provides further information on deprivation characteristics such as Free School Meal status
- Supports analysis of interactions with the children's social care system and of needs, vulnerabilities and adversities recorded within the Children Looked After return and the Children in Need Census

Caveats

- Both the 2011 Census and administrative education data are known to under-count the Gypsy, Traveller and Roma populations
- A Roma tick-box was not included in the 2011 Census
- Gypsy, Traveller and Roma identity may not be disclosed for a range of reasons including fear of discrimination and hostile treatment.
- Gypsy, Traveller and Roma have relatively high rates of nonparticipation in the formal education system and this increases with age. This includes non-registration and missingness from school; spells in and out of school; higher rates of drop out from school, with transitions from primary to secondary school, year 11 exams and transitions into 16-18 education all identified as being key trigger points; with some young people and parents opting for elective home education. Gypsy, Traveller and Roma children and young people also experience higher rates of exclusion and persistent absences.

Two positive design features that mitigate against 'missingness'

- The AEDE is a stock dataset: individuals stay in the AEDE dataset if they fall within the relevant age range and have an eligible education record, even if for example they leave a school, move between schools, leave the country or drop out of school.
- Wave 1 GUiE linkage criteria do not require children to have a record in any particular year (such as 2010/11) or in any particular time window (such as 2009/10-2011/12) with children and young people whose 2011 Census records match to any year of AEDE data (2001/2-2014/15) included in the linkage. This mitigates against the disproportional exclusion of Gypsy, Traveller and Roma children, who are more likely to be mobile or to drop out or disengage from the formal education system, particularly as they age, from the AEDE linkage. It also means that children who have a record but who experience periods of disengagement from school, spells in and out of school, and regular exclusions or persistence absence, should nevertheless have a AEDE record.

Exclusions from coverage

- See Figure 9 for a full list of subpopulations excluded from the GUiE Wave 1 linkage.
- Gypsy, Traveller and Roma children and young people who do not have a 2011 Census record are excluded from the linkage.
- Gypsy, Traveller and Roma children and young people who do not have a record in the feasibility AEDE are excluded from the GUiE Wave 1 linkage. Note that Gypsy, Traveller and Roma children and young people who have not interacted with government-funded education within the relevant time window are not included in the English School Census – this includes those who have not registered for school or who are electively home-educated within the relevant time window. However, there remains some uncertainty as to whether there is some coverage of these groups within the YPMAD attainment data. See section 5.3.3. for further details.

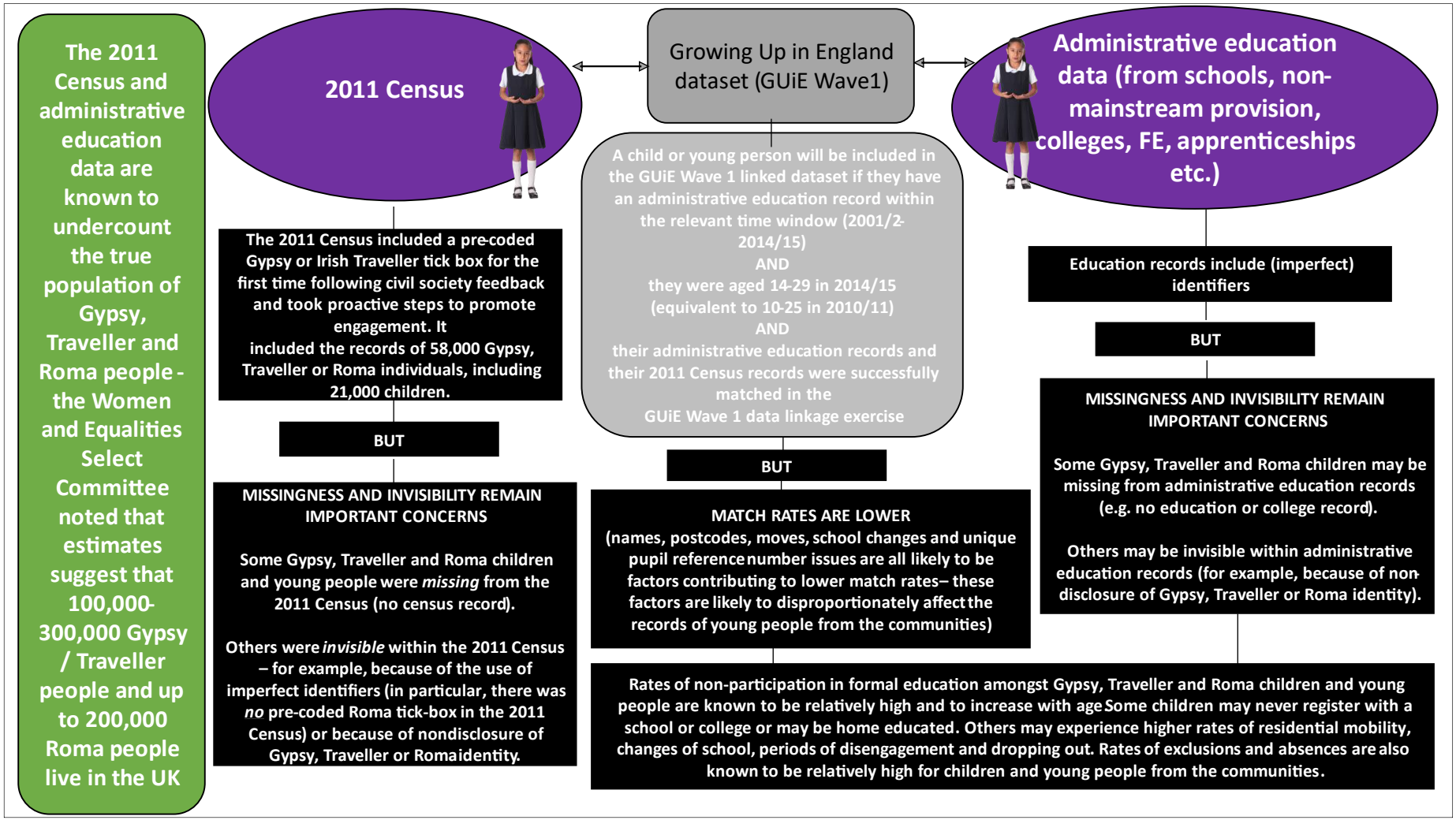
Lower record matching success rates

- The records of Gypsy, Traveller and Roma children and young people are not successful matched into the GUiE Wave 1 linkage on an equal basis.

GUiE Wave 1 composition effects linked to deprivation

- There is some evidence that the linkage is associated with deprivation-related compositional effects.

Figure 12: Strengths and limitations of the GUiE Wave 1 linkage in challenging data exclusion and building up new inclusive evidence on children and young people from the communities



Note: Gypsy, Traveller and Roma population estimates are cited in Women and Equalities Committee (2019), drawing on estimates produced by the Council of Europe and University of Salford.

6.3 Assessing disproportionality in Gypsy, Traveller and Roma record matching success rates

As we noted in section 1, the UK Statistics Authority has highlighted that the linkage error that arises in the process of integrating two administrative datasets is likely to be a source of bias as linkage error can disproportionately impact certain groups, with linked datasets more likely to exclude or incorrectly link them. This affects some ethnic minority groups compared to the White group, who are, as a result, sometimes characterised as 'hard to link' groups. Recommendations include setting up a series of work packages to develop bespoke methods for the linkage of 'hard to link' groups and analysing bias affecting groups with protected characteristics (UKSA 2023).

There are also reasons to anticipate that Gypsy, Traveller and Roma children and young people might be disproportionately excluded from the GUiE record linkage exercise. As noted in section 5.2, the matching exercises exploits common information that is recorded across datasets including names and postcodes. Common reasons for reduced record matching rates or failure include issues with the recording of names (for example, inconsistent or different spellings of names, use of shortened names, changes in names) and issues with the recording of postcodes (for example, inconsistent postcodes or changes in postcodes). These issues and inconsistencies can be higher amongst ethnic minorities and might be anticipated to be disproportionately high among individuals from the communities as they may be exacerbated by factors such as mobility, moves between schools and cultural factors.

To investigate any disproportionality in the GUiE record matching exercise, we originally planned to undertake analysis of the residual (unmatched) Gypsy, Traveller and Roma records looking at both the 2011 Census residuals (2011 Census records not linked to a feasibility AEDE dataset record) and the feasibility AEDE residuals (feasibility AEDE dataset records not linked to a 2011 Census record). We aimed to establish any variation in linkage match rates between Gypsy, Traveller and Roma children and young people compared to other groups and to provide feedback on the characteristics of the unmatched Gypsy, Traveller and Roma 2011 Census cases (the 2011 Census residuals) compared to the characteristics of the matched Gypsy, Traveller and Roma 2011 Census cases.

Unfortunately, we were unable to access the GUiE residuals data or 100% Census data within the SRS for legal reasons. Instead, ONS undertook bespoke analysis for us on our behalf based on a specification

that we provided. The results of this exercise are set out in Table 2, which provides detailed information on GUiE linkage rates by ethnic group. These results are subject to three important limitations and caveats.

First, the residuals exercise is more limited than we originally anticipated in that it covers the feasibility AEDE residuals only. That is, the exercise provides information on the proportion of feasibility AEDE records that have been successfully matched to a 2011 Census record as part of the GUiE linkage. Unfortunately, ONS were not able to provide information on the 2011 Census residuals. Information on the 2011 Census residuals was previously provided as part of the quality assessment for previous census-feasibility AEDE linkage exercises (see section 5.1) and this information is essential for understanding quality, missingness and bias. Therefore, we recommend that information on the GUiE 2011 Census residuals is provided in the future.

Second, it is important to understand that in reporting feasibility AEDE linkage success rates, it is not possible to distinguish between linkage failure that results from "true 2011 Census missingness" (that is, feasibility AEDE cases that are not linked because a young person doesn't have a 2011 Census record – for example, because their family didn't complete a 2011 Census form or only moved to the country and acquired an education record subsequent to the 2011 Census) and record match failures - that is, cases where a record failed to match but should have matched – for example, matching failures that result from inconsistencies in recorded postcodes or difficulties with names. The records of young people from the Gypsy, Traveller and Roma communities are less likely to be successfully matched because of both of these reasons.

Third, additional limitations and caveats result from the ways in which the data have been deduplicated as a basis for the residuals exercise which involves an element of pseudo-randomness. Further details of this issue are provided in the notes under Table 2 (notes 3 and 4).

Subject to these caveats, the results of the residuals exercise raise concerns about disproportionality and substantially lower match rates for young people from the Gypsy, Traveller and Roma communities. Match rates are as high as 73% for the White group. When only the feasibility AEDE part of the GUiE dataset is used as a basis for ethnicity identification, matches are as low as 40% for Gypsy/Roma and 39% for Irish Traveller records. When the ethnicity data pooling methodology we set out in Section 7 is used, combining ethnicity information from both parts of the dataset, Gypsy, Traveller or Roma match rates rise to 51%. The overall GUiE linkage match rate suggested by this exercise is 70% (with or without dropping 'true duplicates according to the AEDE Person Spine duplication flag').

These results limit the extent to which the GUiE dataset can and should be thought of as representative of the underlying true population of Gypsy, Traveller and Roma young people; and the extent to which generalisations and inferences can be drawn from the experimental statistical results we present in section 7. They also suggest the need for a follow-up clerical exercise that examines the extent to which linkage error might be reduced for Gypsy, Traveller and Roma young people though the development of bespoke linkage methodologies that aim to retain as many Gypsy, Traveller and Roma cases within linked population census-administrative datasets as is possible. Looking forward, recommendations on bespoke clerical exercises of this type that aim to maximise the extent to which population census-administrative data linkages are representative of 'seldom heard' and 'hard to link' groups are included in the concluding section of this report.

Figure 13: The GUiE linkage: record matching exercise

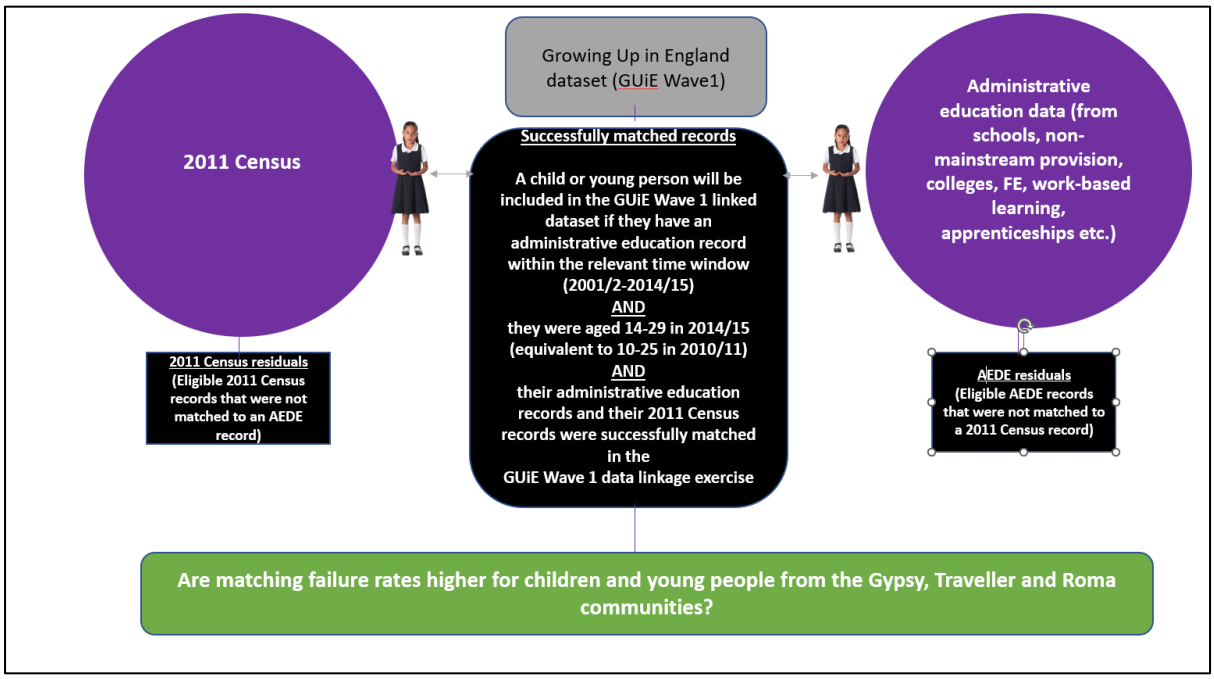


Table 2: Results of residuals analysis (differential matching rate by ethnic group)

		Without dropping duplicates		With duplicates dropped (that is, merge in the flags from the spine, and drop case if case is flagged up as a duplicate)		% matched (match rate)	
Ethnicity (from feasibility AEDE records – except for combined GTR variable specified below)	All feasibility AEDE subset for linkage (number)	Feasibility AEDE – Census linked cohort (number)	Unlinked feasibility AEDE (number)	Feasibility AEDE – Census linked cohort (number)	Unlinked feasibility AEDE (number)	**Without dropping duplicates**	**With duplicates dropped**
White	**6168375**	**4510895**	**1657480**	**4510892**	**1652803**	**73.13**	**73.18**
White British	5874162	4313570	1560592	4313567	1556261	73.43	73.49
White Irish	27508	19933	7575	19933	7544	72.46	72.54
Feasibility AEDE any year - Gypsy/Roma	14181	5628	8553	5628	8521	39.69	39.78
Feasibility AEDE any year - Irish Traveller	3804	1472	2332	1472	2323	38.6961094	38.78788
Feasibility AEDE any year - combined Gypsy/Roma/ Irish Traveller	17839	7033	10806	7033	10766	39.42	39.51
Feasibility AEDE any year Gypsy/Roma/ Irish Traveller OR Census 2011 Gypsy/ Irish Traveller(ie using combined ethnicity information from any AEDE year AND from 2011 Census)	21879	11073	10806	11073	10766	50.61	50.70
White other	266705	177392	89313	177392	88998	66.51	66.59
Mixed	261134	199183	61951	199183	61602	76.28	76.38
Mixed White/Asian	53473	42122	11351	42122	11313	78.77	78.83
Mixed White/Black African	27425	20848	6577	20848	6527	76.02	76.16
Mixed White/Black Caribbean	88750	66442	22308	66442	22191	74.86	74.96

Mixed other	91486	69771	21715	69771	21571	76.26	76.38
Asian	604316	396184	208132	396181	207098	65.56	65.67
Bangladeshi	93977	53412	40565	53410	40316	56.84	56.99
Indian	182935	136753	46182	136753	46036	74.75	74.81
Pakistani	232448	136713	95735	136712	95333	58.81	58.92
Asian other	94956	69306	25650	69306	25413	72.99	73.17
Black	344861	247898	96963	247895	95633	71.88	72.16
Black African	198372	143648	54724	143645	53837	72.41	72.74
Black Caribbean	106703	76623	30080	76623	29776	71.81	72.01
Black other	39786	27627	12159	27627	12020	69.44	69.68
Chinese	31159	22270	8889	22270	8585	71.47	72.18
Other	96347	67614	28733	67614	28350	70.18	70.46
Unknown	1913110	1141980	771130	1141969	766420	59.69	59.84
TOTAL (overall – all ethnicities)	9554122	6672776	2881346	6672756	2868310	69.84	69.94
Extra							
Invalid and Missing	23	11	12	11	12	47.83	47.83
Not Obtained and Refused	116958	79708	37250	79708	37041	68.15	68.27

Source: ONS.

Notes

1. Analysis undertaken by ONS to team specification.
2: These data are not official statistics.
3.The estimates are subject to a level of pseudorandomness as a result of the methodology used to prepare data for linkage, meaning that estimates may vary to a small degree. ONS advise caution when using the data.
4. An important limitation of the data stems from how the data are deduplicated. An individual may have multiple recordings of ethnicity in a given academic year. ONS took the approach to deduplicate the feasibility AEDE attributes data on adr_id and academic year prior to merging the data to the 2011 Census such that there was only one record per individual in a given academic year. From this it was determined the mode ethnicity (or most recent ethnicity in multimodal cases). This approach may lead to slight variation to the resultant dataset due to the pseudorandom nature of data partitioning when this deduplication method is implemented. Future work could involve partitioning data by id and ranking all recordings of ethnicity across all available years to determine the mode ethnicity for an individual prior to deduplication. The approach to multimodal records is under development but could involve creating a multimodal response category within the dataset.

6.4 Assessing Wave 1 GUiE compositional effects

It is also important to note that there also appear to be some important compositional differences between the deprivation characteristics of the sample of young people identified as Gypsy, Traveller and Roma at ages 14 and 15 in GUiE and the deprivation characteristics of a sample of young people identified as Gypsy, Traveller and Roma at ages 14 and 15 in the 100% Census. The deprivation characteristics of the sample of young people identified as Gypsy, Traveller and Roma at ages 14 and 15 in GUiE are set out in Table 9 in the following section of this report (section 7). For the purposes of comparison, the deprivation characteristics of young people identified as Gypsy, Traveller and Roma at ages 14 and 15 in standalone analysis of the 100% Census are presented below in Table 3.

Comparing Table 9 and Table 3, the differences are particularly apparent in relation to the percentage that are living in a caravan or mobile structure. Amongst the GUiE sample at age 14 and 15, this is 10%. Amongst the 100% Census sample, this is 23% at age 14 and 25% at age 15. Overall rates of Gypsy, Traveller and Roma childhood deprivation at ages 14 and 15 are also lower when examining the GUiE Gypsy, Traveller and Roma sample than is the case based on the analysis of 100% Census, with 14% of the GUiE 14 and 15 year old Gypsy, Traveller and Roma young people are classified as 'not deprived' in any dimension, compared to 9% of 14 and 15 year olds identified as Gypsy, Traveller and Roma in the 100% Census. Rates of education deprivation, while still extremely high in the GUiE sample at 51% compared to 18% of other young people, are nevertheless 10 percentage points lower in GUiE than in the 100% Census sample (and in our previous study).

There are several potential explanations of this difference. First, the data pooling methodology we apply to identify the GUiE sample results in the identification a much larger Roma population – and Roma people in the UK are known to be considerably less likely to be mobile than Gypsies or Travellers. We also observe a substantially higher proportion of 14 and 15 year olds identified as Gypsy, Traveller and Roma with a non-UK country of birth in GUiE compared to the 100% Census (23% compared to 12% and 13% at ages 14 and 15 respectively in 100% Census). This is again not surprising, as it is well known that Roma people are more likely to be recent immigrants to the UK.

Additionally, age composition may also be playing a role. For example, our previous study reported that one third of Gypsy, Traveller and Roma young people aged 18 or under lived in a caravan or other mobile structure, compared to a figure of 25% amongst 100% Census sample and 10% in

the GUiE sample. Our analysis of 100% Census shows that rates of living in a caravan or mobile structure are consistently lower than 30% amongst young people aged 9-25, suggesting that rates may be higher amongst the youngest children.

However, as the GUiE sample only includes young people whose records have been successfully matched across administrative education data and the 2011 Census, it is also a possibility that the exclusion of particularly deprived subpopulations from the GUiE linkage is also affecting the deprivation characteristics of the GUiE sample. For example, we might anticipate that the young people from the communities who do not have a matched record and may be missing from the formal school system and / or were not recorded in the 2011 census are on average likely to be more mobile and more likely to be disadvantaged than those in the GUiE sample. The possibility highlights the concerns raised by the NSDEC Committee at the outset of our project, that some of the most disadvantaged Gypsy, Traveller and Roma young people may be systematically excluded from and under-represented in the GUiE linkage.

Table 3: Deprivation characteristics of Gypsy, Traveller and Roma young people at ages 14 and 15 (100% Census, percentages)

	Gypsy, Traveller and Roma		Other young people	
Age	14	15	14	15
Sex				
Male	47.8	52.7	51.2	51.5
Female	52.2	47.3	48.8	48.5
Family type				
Child in couple family	42.0	45.5	68.0	67.5
Other	5.8	7.5	3.6	4.0
Child in single parent family	52.2	47.0	28.5	28.5
Count of families in the households				
two or more families	2.5	5.1	1.9	2.0
0 families	1.9	1.9	0.4	0.5
one family	95.6	93.0	97.6	97.5
Country of birth				
Born outside the UK	12.3	13.0	8.3	8.5
Born within the UK	87.7	87.0	91.7	91.5
Household tenure				
Private rent	20.9	20.6	13.2	12.6
Rent free	2.2	2.3	0.8	0.8
Social rent	46.8	49.3	21.7	21.4
Owned (incl.shared ownership)	30.1	27.8	64.3	65.1

Accommodation type				
House/Flat	74.6	77.1	99.9	99.9
Caravan/mobile or temp structure	25.4	22.9	0.1	0.1
Household overcrowded (bedroom standard)				
overcrowded	39.2	37.9	12.5	12.1
not overcrowded	60.8	62.1	87.5	87.9
Number of dependent children in the family				
Three or more	51.9	49.7	30.4	27.6
One or two	48.1	50.3	69.6	72.4
Number of dependent children in the household				
Three or more	53.4	52.4	31.1	28.3
One or two	46.6	47.6	68.9	71.7
Bad/fair general health or limiting long-standing illness/disability				
no	94.2	95.7	97.8	97.7
yes	5.8	4.3	2.2	2.3
Household education deprivation (ONS)				
no	39.0	42.7	81.0	80.7
yes	61.0	57.3	19.0	19.3
Household employment deprivation (ONS)				
no	63.7	58.4	83.5	83.0
yes	36.3	41.6	16.5	17.0
Household health and disability deprivation (ONS)				
No	49.5	49.0	73.5	72.8
yes	50.5	51.0	26.5	27.2
Household housing deprivation (ONS)				
No	54.2	56.3	84.4	84.9
Yes	45.8	43.7	15.6	15.1
Deprivation in:educ,empl,health,housing (ONS)				
deprived in 4 dimensions	7.9	7.3	0.6	0.7
deprived in 3 dimensions	22.2	23.3	5.3	5.5
deprived in 2 dimensions	34.6	34.2	14.4	14.5
Not deprived in any dimension	8.9	9.1	49.3	48.8
deprived in 1 dimension	26.4	26.2	30.4	30.6
Deprived in any:educ,empl,health,housing(ONS)				
deprived in at least 1 dimension	91.1	90.9	50.7	51.2
not deprived	8.9	9.1	49.3	48.8

Source: ONS. Note: data provided by ONS based on Obolenskaya specification.

Table 4: Gypsy, Traveller and Roma young people living in a caravan or a mobile or temporary structure (100% Census)

Age	9 to 13	14	15	16	17	18	19	20 to 25
House/Flat	75.1	74.6	77.1	74.8	78.0	76.0	82.1	81.2

Caravan/mobile or temp structure	24.9	25.4	22.9	25.2	22.0	24.0	17.9	18.8

Source: ONS. Note: Table provided by ONS based on Obolenskaya specification.

6.5 Implications: caveating the extent to which GUiE is representative of the underlying population of Gypsy, Traveller and Roma young people

The analysis in this section highlights the substantial role that the Growing Up in England dataset, as one of a new generation of large-scale administrative data - population census linkages, can play in challenging data exclusion and building up new evidence on groups that in the past have been systematically missing from, or invisible within, social statistics. However, while highlighting the 'value added' of GUiE and its research potential, we have also noted multiple limitations and weaknesses relating to imperfect identification, missingness and disproportionality. These limitations restrict the extent to which the GUiE Wave 1 dataset can be thought of as being representative of the underlying population of Gypsy, Traveller and Roma young people and the extent to which generalisations and inferences should be drawn from our findings. Therefore, it is critically important that the experimental results that we report in section 7 are interpreted in conjunction with the analysis of limitations and caveats provided in this section.

7 Experimental statistical findings: results of an initial research exercise using Wave 1 GUiE to build up new evidence on the lives, circumstances and needs of Gypsy, Traveller and Roma young people

This section sets out findings from an initial research exercise that uses the Growing Up in England (GUiE) dataset to build up new experimental evidence on the lives, circumstances and needs of young people from the Gypsy, Traveller and Roma communities. In the research exercise, we use the Wave 1 GUiE dataset to identify the records of substantially larger numbers of Gypsy, Traveller and Roma young people than has been possible in most previous research. Our methods for maximising the identification of the records of Gypsy, Traveller and Roma young people within the GUiE Wave 1 data exploit the large-scale and longitudinal nature of the dataset as well as the potential for pooling ethnicity information sourced from the 2011 Census and administrative education data systems. Applying these methods, we use the education side of the GUiE Wave 1 dataset to follow the educational journeys of three academic year cohorts of young people over four years, from the academic year they turned 15 to the academic year that they turned 19. We report on the educational pathways of the Gypsy, Traveller and Roma young people within the cohorts as they transition through 14-16 and 16-18 education and training through any route (schools, further education, vocational education, work-based learning or apprenticeships) and on their highest educational attainment (Level 1, Level 2 or Level 3 qualifications) by ages 17 and 19, comparing these to the pathways and attainment of their peers. Additionally, drawing on the rich information about individual characteristics and household circumstances recorded in the 2011 Census, we use the GUiE Wave 1 data to examine the multidimensional child disadvantage characteristics of the Gypsy, Traveller and Roma children identified in GUiE Wave 1 (observed at ages 14 and 15) and to assess the strength of the associations between different indicators of multidimensional child disadvantage and highest educational attainment by age 19 attained by any route for Gypsy, Traveller and Roma and other young people. More detailed results from the experimental research exercise (including full unweighted frequency information and methods) are included in detailed notes produced for the purposes of Statistical Disclosure Control (Obolenskaya 2022ab).

The section is organised as follows. We begin with an introductory discussion on aims and scope. We then acknowledge the main limitations of the research exercise. Next, we provide further methodological details,

including further details of the analysis samples used in the research exercise; our innovative methodology for identifying the records of Gypsy, Traveller and Roma young people in Wave 1 GUiE by pooling ethnicity information from across the different component parts of the linked dataset; and our selection of academic year cohorts for the longitudinal analysis. Finally, we set out our experimental statistical findings on the educational journeys of Gypsy, Traveller and Roma young people, reporting on their main educational pathways and highest educational attainment and the relationship between educational outcomes by age 19 and multidimensional child disadvantage.

7.1 Introduction to the research exercise

We have designed this experimental research exercise based on our assessment of the strengths and limitations of the GUiE Wave 1 dataset, as set out in earlier sections of this report. There are three key elements of the research design.

First, we build on our analysis of the discussion of the strengths and limitations of the ethnicity identifiers available in different parts of the GUiE dataset set and their basis for Gypsy, Traveller and Roma identification (c.f. section 6.1). We acknowledge that the Gypsy, Traveller and Roma identifiers that are available in Wave 1 GUiE fall short of the new good practice 18+1 ethnicity classification standard that was introduced in the 2021 Census. However, the research exercise makes optimum use of the (imperfect) ethnicity information that is available from the 2011 Census and administrative education data systems and explores how Gypsy, Traveller and Roma identification can be maximised by pooling the ethnicity information that is available in the different component parts of the GUiE Wave 1 dataset.

Second, the research design is shaped by the nature and scope of the education data that is available through the feasibility AEDE side of the GUiE Wave 1 dataset. As we discussed in section 5.3.4, the primary purpose of underlying dataset from which the educational attainment data in GUiE Wave 1 is derived (YPMAD) is to produce national statistics on Level 2 and Level 3 attainment by age 19. Drawing on this data, the GUiE Wave 1 dataset includes binary flags that identify main pathways and attainment by age, including highest educational qualifications by Level. The attainment variables distinguish between Level 1 qualifications (one to four GCSEs at any grade) or equivalent; Level 2 qualifications (five or more GCSES grades A* to C or equivalent; Level 3 qualifications (two or more A levels or equivalent). For example, ’achieving Level 2 by age 16’ refers to

the attainment of learners typically in year 11; and 'achieving Level 3 by age 19' refers to the attainment of learners in the academic year *after* they typically complete school. The longitudinal data made available in the feasibility AEDE side of the dataset enables us to follow the Gypsy, Traveller and Roma young people within our focus academic year cohorts over a four-year period and to compare their pathways and attainment to those of other young people.

Third, while official education statistics are one of the key available data sources on the lives of Gypsy, Traveller and Roma young people, they do not include information about family background and household circumstances. As discussed in section 5, this is a general limitation of administrative data and the key strength of the GUiE linkage it that it makes it possible to combine the information gathered within administrative education data systems with the rich information on individual and household circumstances recorded in the 2011 Census. This has particular importance and 'value added' in the context of Gypsy, Traveller and Roma research. The 2011 Census data includes detailed information on living in a caravan or other temporary or mobile structure, as well as broader information on household employment, household education deprivation, family structure, and ill-health and disability. This information is highly informative for understanding the lives, circumstances and needs of children and young people from the Gypsy, Traveller and Roma communities and is rarely available from other data sources.

Two previous outputs producing experimental statistical findings using the GUiE dataset have been produced by ONS (ONS 2022ef). In terms of the broader literature in this field, we note that in addition to the specialist literature on Gypsy, Traveller and Roma inequalities reviewed in section 3, there is a vast body of research on educational disadvantage. A recent study (Lupton et al 2021) stands out as having particular substantive and methodological relevance for the current research exercise. This study does not separately identify the progression, attainment or support needs of young people from the Gypsy, Traveller and Roma communities. However, it is particularly important in highlighting the need for universal measures to support educationally disadvantaged young people, with a particular emphasis on post-16 alternative further education and vocational pathways. Methodologically, the Lupton et al study is also particularly relevant to the current project as it analysed data on the five cohorts of young people who completed GCSEs between 2013 and 2017 by linking the National Pupil Database to Individual Learner Records. Specifically, the study constructed a dataset that enabled the educational journeys of young people to be followed from the end of KS4 in Year 11, when they sit their

GCSE exams, up to the end of the post-16 phase, looking at progress and attainment both up to age 18 and age 19 (recognising that some young people, particularly 'lower attainers', take three years over this phase). Like our own study, the quantitative analysis for the study was supplemented by a qualitative research exercise and detailed analysis of post-16 opportunities and requirements.

7.2 Caveats and limitations

The analysis in this section is intended to be read in conjunction with our assessments of the strengths and limitations of GUiE undertaken in sections 5 and 6. We do not repeat the findings of those sections here but rather underline that the experimental statistical findings set out in this section require careful interpretation and caveating in the light of the limitations of the GUiE dataset identified in sections 5 and 6 as well as additional issues identified in our literature review and direct engagement exercise. In particular, we acknowledge that the discussion in section 6 highlights a series of concerns around imperfect ethnicity identification, missingness, disproportionality and compositional effects. These limitations restrict the extent to which the GUiE Wave 1 dataset can be thought of as being representative of the underlying population of Gypsy, Traveller and Roma young people and the extent to which generalisations and inferences should be drawn from our findings.

In relation to data cleaning, our initial step in using the Wave 1 GUiE dataset involved undertaking the Wave 1 GUiE merge, which links the different parts of the GUiE dataset together. In undertaking this merge, duplicate records were identified in the feasibility AEDE spine data. Some of the queries that we raised in relation to our early data exploration and data cleaning exercise were discussed in section 5 of this report; and one of our key early concerns related to this concern. A follow up clerical review exercise was undertaken by ONS and the feasibility AEDE Person Spine dataset was resupplied with a 'record duplication flag' variable made available in the SRS. This provides a basis for identifying and dropping duplicate cases. However, as the analysis samples and academic year cohorts for the current study had already been identified when this new dataset was deposited – and as in any case we did not want to entirely drop duplicate cases – we have not made use of the flag dataset. Rather, our methodology has been to retain the first record or the record with the richest information where duplicates were identified.

Issues of conflicting ethnicity information within the GUiE dataset have been flagged in other early GUiE research exercises (see in particular ONS

2022c). In the context of this previous research exercise, the approach adopted was to make use of the most recent recorded ethnicity information. Alternatives such as modal ethnicity information are also being explored by ONS. The approach we have adopted in this study is rather different and takes account of expert advice and guidance from our Advisory Group, the findings of our literature review and the findings from our direct engagement workshops relating to the non-disclosure of Gypsy, Traveller and Roma identity in official data collections and administrative data systems. Drawing on these insights and findings, our identification of Gypsy, Traveller and Roma children and young people relies on *any* mention of Gypsy, Traveller and Roma identity in *any* of the underlying GUiE data (using 2011 Census as a starting point for identification and integrating it with information on ethnicity from *any* year of the feasibility AEDE longitudinal records). Further details of our data pooling approach to Gypsy, Traveller and Roma sample and academic year cohort identification are provided below.

As noted in section 5, information derived from the 2011 Census about personal and household characteristics only provides a snapshot of the situation on census day – and does *not* provide information on an annual basis. However, below we establish that through careful selection of the age and years of the young people in cohorts, this data can be understood as providing a snapshot of childhood multidimensional disadvantage at a given age, which can then be related to subsequent educational attainment at age 19. While missing non-data subject Census 2011 records (discussed in section 5.3.2) have limited the analysis we can undertake, our approach has been to make the best use of the data that is available.

As a basic early cross-check, we triangulated GUiE-derived highest educational attainment by age 19 with DfE published figures on overall educational attainment by age 19. As one key use of YPMAD data (which provides the basis of the highest educational qualification information in GUiE) is to produce national statistics on highest educational attainment by age 19, we would anticipate that information on overall highest educational attainment at age 19 derived from GUiE would provide a reasonable match with published data. Reassuringly, our analysis indicated a reasonable degree of alignment between the GUiE highest educational attainment data and published national statistics. This is shown by comparing the overall Level 3 attainment at age 19 results based on published data for 2018/19 (Table 1, this report) with the overall Level 3 at age 19 results derived from GUiE (Table 6, this report), both of which suggest around 40% of young people do not attain Level 3 qualifications by age 19. These results also

align with the findings on 'low attainers' from a previous study using linked NPD and ILR records (Lupton et al 2021).

7.3 Further methodological details

We now provide further methodological details of the research exercise including our methodology for maximising the identification of the records of Gypsy, Traveller and Roma young people; our analysis sample sizes; and our selection of academic year cohorts.

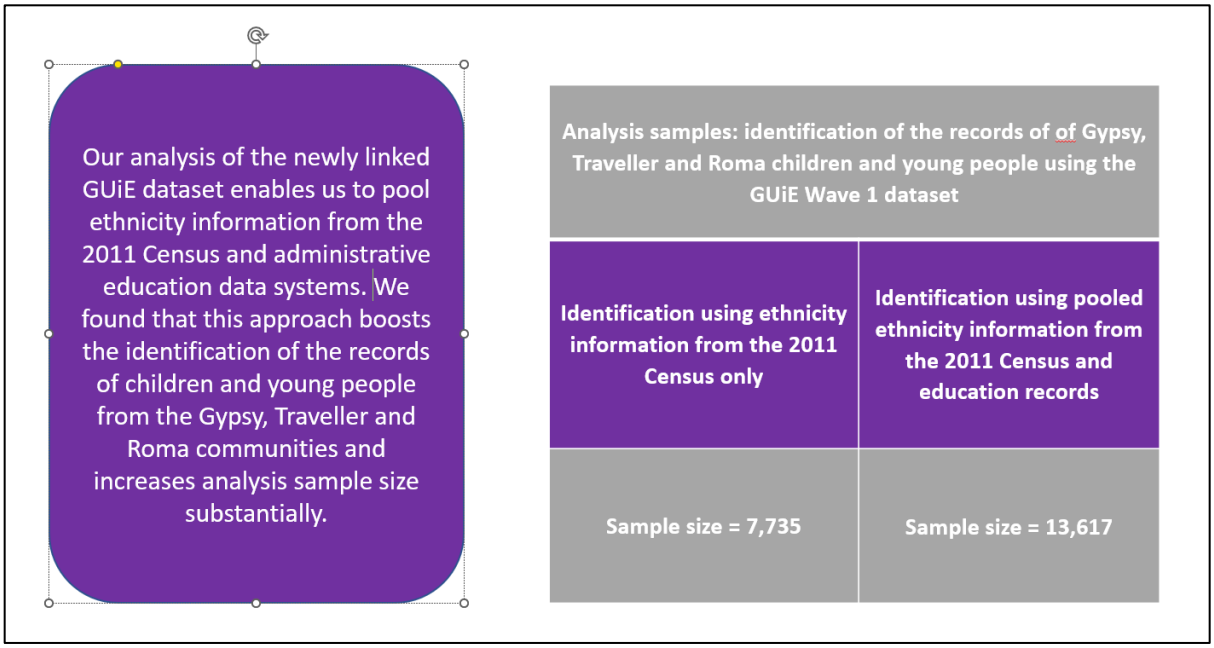
7.3.1 Methodology for pooling ethnicity information across the different component parts of the GUiE Wave 1 dataset

In the research exercise, we use the Wave 1 GUiE dataset to identify the records of substantially larger numbers of Gypsy, Traveller and Roma young people than has been possible in most previous research. This in part reflects the large-scale nature of the dataset. Additionally, our methodology for maximising the identification of the records of Gypsy, Traveller and Roma young people and identifying large analysis samples involves pooling the ethnicity information that is available from the 2011 Census and administrative education data systems.

As discussed in detail in section 6.1, the ethnicity information available in both the 2011 Census and administrative education data systems are limited in important respects. The 2011 Census records Gypsy or Irish Traveller status under the White ethnic group - but Roma identity was not included as a pre-coded option. In relation to the feasibility AEDE side of the dataset, the NPD ethnicity minor variable records ethnicity using Gypsy / Roma and Traveller of Irish Heritage categories, while the ILR learner data includes a Gypsy or Irish Traveller identifier.

The research exercise demonstrates that by pooling ethnicity information sourced from the 2011 Census and administrative education records, it is possible to increase the identification of Gypsy, Traveller and Roma children and young people substantially Additionally, the longitudinal nature of the feasibility AEDE data opens up a substantially longer time window for disclosing and observing Gypsy, Traveller or Roma identity. Gypsy, Traveller and Roma analysis sample size is boosted substantially from 7,735 young people based on 2011 Census ethnicity information alone to 13,617 when ethnicity information sourced from across the different component parts of the GUiE Wave 1 dataset is pooled (Figure 14).

Figure 14: Maximising the identification of the records of Gypsy, Traveller and Roma young people in Wave 1 GUiE by pooling ethnicity information from the 2011 Census and administrative education data systems



Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. Further details of the analysis are provided in Obolenskaya 2022a.

7.3.2 Selection of academic year cohorts for the longitudinal analysis

Our selection of analytical cohorts for the longitudinal analysis takes account of the focus of the feasibility AEDE on academic (rather than calendar) years as well as the design and limitations of GUiE Wave 1 described in section 5.

Cohort selection for the analysis of main educational pathways and educational attainment

For the analysis of main educational pathways and educational attainment reported in section 7.4 below, we focus on three academic year cohorts: specifically, young people who turned 15 in academic years 2008/9, 2009/10 and 2010/11. We follow the young people in these cohorts over four years to academic years 2012/13, 2013/14 and 2014/15 respectively - the academic years in which the young people in each of the three cohorts turned 19. Information across these three academic year

cohorts is combined as a basis for reporting the main findings on pathways and attainment in section 7.4. In total, we identify the records of 2,966 Gypsy, Traveller and Roma young people across these three academic year cohorts[3].

The selection of these three academic year cohorts for the analysis in section 7.4 takes account of GUiE age restrictions. As highlighted in section 5, the feasibility AEDE side of the dataset is restricted to young people aged 14-29 in 2014/15 (based on age recorded in the feasibility AEDE - or equivalently, young people aged 10-25 in 2010/11). Additionally, the upper age limit of the education attainment data we are using is age 20. Moreover, in line with national statistics on highest educational attainment by age 19, we have chosen to observe highest educational qualification in the academic year students turn 19 rather than the academic year they turn 18 as a typical student is studying for 16-18 qualifications (key stage 5) in the academic year they turn 18. As the last academic year for observing highest educational qualifications by age 19 using the GUiE dataset is 2014/15, the final cohort of young people we can follow for this analysis is the cohort that turned 15 in 2010/11.

Cohort selection for the multidimensional child disadvantage analysis

For elements of the multidimensional child disadvantage analysis reported in section 7.5.2 below, we further restrict the analysis to a single cohort of young people – specifically, the cohort of young people who turned age 15 in the academic year 2010/11 and age 19 in academic year 2014/15. This single academic year analytical cohort includes 1,088 Gypsy, Traveller and Roma young people. We have selected this academic year cohort as a focus for the analysis in section 7.5.3 because we can observe *both* the household deprivation characteristics before the young people in the cohort turn age 16 *and* their highest educational attainment by age 19 using the GUiE Wave 1 dataset. Specifically, the household deprivation information we use in the analysis was recorded on 27th March 2011 (2011 Census day). This fell during the academic year 2010/11, when the young people in this academic year cohort were aged 14 or 15. Additionally, for the young people in this cohort, we can observe educational attainment by

[3] Note that below, we sometimes refer to the young people in these three analytical cohorts in terms of their ages on 31st August of a given academic year. For example, young people in the cohorts may be referred to as being aged 15 at the end of academic years 2008/9, 2009/10 and 2010/11 on 31st August 2009, 2010 and 2011 respectively. Alternatively, they may be referred to as being aged 18 at the beginning of academic years 2012/13, 2013/14 and 2014/15 on 31st August 2012, 2013 and 2014 respectively.

age 19 in academic year 2014/15, the final year of the GUiE Wave 1 education data.

The selection of this academic year cohort also reflects the limitations imposed by the change of provision in 2011 Census records that occurred between the development of Proof of Concept and GUiE Wave 1. Our aim is to examine the relationship between childhood deprivation (that is, household deprivation before the age of 16) and highest educational attainment at age 19. However, as noted in section 5, due to the change in provision between Proof of Concept and Wave 1 GUiE, the 2011 Census individual records for non-data-subjects are not included in the Wave 1 GUiE dataset. One implication is that we have not been able to construct household deprivation variables ourselves using the individual level records of all the individuals in the household as we did in the context of our previous study (Vizard et al 2018, Burchardt et al 2018). Rather, we have been reliant on derived variables on household deprivation which *have* been made available in the GUiE dataset. While these derived deprivation variables are rich in many respects, they have some disadvantages for our analysis. In particular, the ONS derived measure of household education deprivation counts households as being educationally deprived if no one in the household has Level 2 qualifications and no one is a student between ages 16 and 18. As many of our data subjects are aged 16-18 and are in education, and others will have siblings in education within this age range, this restriction has important implications for our analysis. To make best use of the derived variable that is available, it has been necessary to observe household educational deprivation status before young people turn 16. Note that this is a second-best solution and one of our key recommendations is that 2011 Census records for *all* household members are included in future GUiE releases and / or future population census-administrative data linkages.

Figure 15: Analysis sample size: identification of the records of Gypsy, Traveller and Roma young people within our academic year cohorts

Analysis of educational pathway and attainment (3 analytical cohorts)	Multidimensional child disadvantage analysis – (single cohort)
Young people who turned 15 in academic years 2008/9, 2009/10 and 2010/11	Young people who turned 15 in academic year 2010/11
Total number of Gypsy, Traveller or Roma young people identified across three analytical cohorts = 2,966	Total number of Gypsy, Traveller or Roma young people =1,088

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. Further details of the analysis are provided in Obolenskaya 2022a.

7.4 Experimental findings: the educational pathways and attainment of children and young people from the Gypsy, Traveller and Roma communities

We now use the GUiE Wave 1 dataset to follow the educational journeys of the young people within our three focus academic year cohorts (young people who turned 15 in academic years 2008/9, 2009/10 and 2010/11) over four years to the academic year they turned age 19 (academic years 2012/13, 2013/14 and 2014/15, respectively). We report on the 14-16 and 16-18 educational pathways of the Gypsy, Traveller and Roma young people in these cohorts through further education and vocational and work-based learning as well as schools and assess their educational attainment, drawing comparisons with their peers. Specifically, we report on main educational pathways (the main qualifications young people were working towards in each year) as the young people in the cohorts turn 16 and, for those who transition into post-16 education, as they turn 18. Additionally, we assess the highest level of educational attainment (Level 1, Level 2 or Level 3 qualifications) attained via any route by ages 17 and 19.

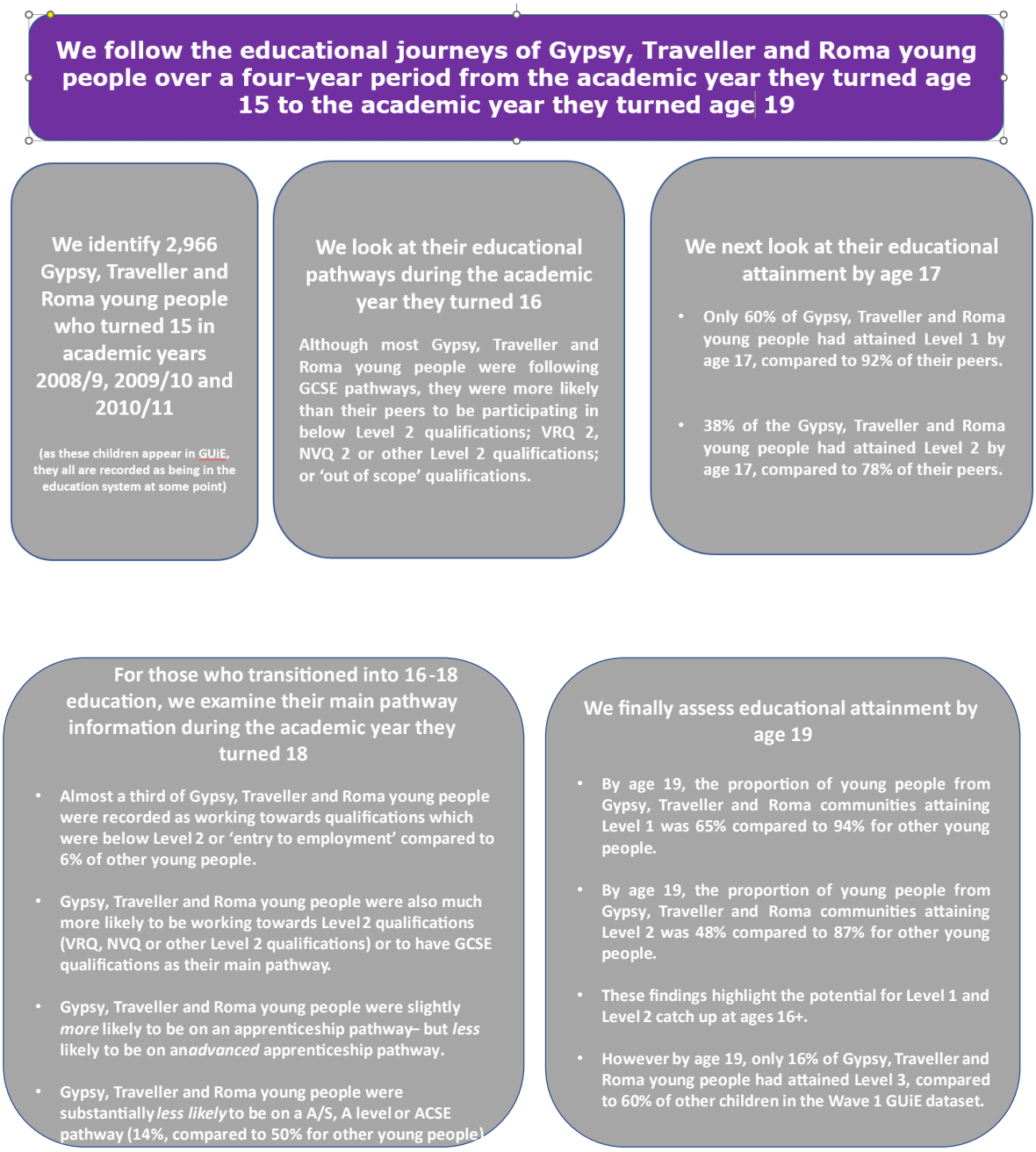
Findings on the main educational pathways of the young people in the cohorts during the academic year they turned 16 are set out in Table 5. The results show that while a majority of Gypsy, Traveller and Roma young people were following GCSE (or IGCSE) pathways, young people from the communities were substantially more likely than other young people to be participating in below Level 2 qualifications; in National Vocational Qualifications (NVQs) 2, Vocationally Related Qualifications (VRQ) 2 or other Level 2 qualifications; or participating in 'out of scope' qualifications.

Findings on the educational attainment of the young people in the cohorts by age 17 are set out in Table 6. The results show that only 60% of the Gypsy, Traveller and Roma young people in the cohorts had attained Level 1 by age 17, compared to 92% of other children in the same cohorts. Additionally, 38% of Gypsy, Traveller and Roma young people in the cohorts had attained Level 2 by age 17, compared to 78% of other young people.

For those who transitioned into 16-18 education, we also assess main pathway information in the academic year the young people turned age 18 (Table 7). Almost third of Gypsy, Traveller and Roma young people in the cohorts were recorded as working towards qualifications which were below Level 2 or 'entry to employment' compared to 6% of other young people. Gypsy, Traveller and Roma young people in the cohorts were also much more likely to be working towards GCSEs (or IGCSEs) or other Level 2 qualifications (NVQ 2, VRQ 2 or other Level 2 qualifications) as their main pathway. Gypsy, Traveller and Roma young people in the cohorts were slightly *more* likely to be on an apprenticeship pathway – but *less* likely to be on an *advanced* apprenticeship pathway. Gypsy, Traveller and Roma young people in the cohort were also substantially *less likely* to be on a A/S, A level or AVCE pathway (14%, compared to 50% for other young people).

Findings on educational attainment by age 19 are set out in Table 8. By age 19, the proportion of young people from Gypsy, Traveller and Roma communities who had attained Level 1 qualifications stood at 65% and the proportion who had attained Level 2 qualifications was 48%. Although figures remained substantially lower than for other children (at 94% and 87% respectively), these findings nevertheless point towards the potential for Level 1 and Level 2 catch up at ages 16-19. However, by age 19, only 16% of the Gypsy, Traveller and Roma in the cohorts had attained Level 3 qualifications, compared to 60% of other young people.

Figure 16: Educational journeys of Gypsy, Traveller and Roma young people



Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. Further details of the analysis (including unweighted frequencies) are provided in Obolenskaya 2022ab.

Table 5: Main educational pathways of young people in the academic year they turned 16 (cohorts of young people who turned 15 in academic years in 2008/9, 2009/10 and 2010/11 - pooled data)

	Gypsy, Traveller and Roma	Other ethnic groups	All	Gypsy, Traveller and Roma	Other ethnic groups	All
	Unweighted frequency			Percentage		
Participating but not in scope qualifications recorded	120	11,654	11,774	4.0	0.8	0.8
Below L2 / E2E	276	18,152	18,428	9.3	1.3	1.3
VRQ 2/NVQ 2/Other level 2	74	8,139	8,213	2.5	0.6	0.6
GCSE/IGCSE	2,418	1,262,900	1,265,318	81.5	90.7	90.6
Apprenticeships (including advanced)	24	9,378	9,402	0.8	0.7	0.7
VRQ Level 3/Other Level 3	18	38,282	38,300	0.6	2.7	2.7
AS/A Level/AVCE or higher	36	44,370	44,406	1.2	3.2	3.2
Total	2,966	1,392,875	1,395,841	100	100	100

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The Information on main educational pathway is taken from the feasibility AEDE. Identification of Gypsy, Traveller or Roma ethnicity uses combined information from the 2011 Census and the feasibility AEDE. Further details are provided in Obolenskaya 2022a.

Table 6: Main educational pathways of young people in the academic year they turned 18 (cohorts of young people who turned 15 in academic years in 2008/9, 2009/10 and 2010/11 - pooled data)

	Gypsy, Traveller, Roma	Other ethnic groups	All	Gypsy, Traveller, Roma	Other ethnic groups	All
	Unweighted frequency			Percentage		
Below L2 / Entry to employment	625	76,010	76,635	30.56	6.03	6.07
VRQ level 2/NVQ 2/Other L2	494	119,289	119,783	24.16	9.46	9.49
GCSE/IGCSE	42	15,233	15,275	2.05	1.21	1.21
Apprenticeship	160	79,489	79,649	7.82	6.31	6.31
NVQ3/Other Level 3	172	89,840	90,012	8.41	7.13	7.13
VRQ Level 3	215	208,899	209,114	10.51	16.57	16.56
AS/A level/AVCE	289	633,282	633,571	14.13	50.25	50.19
Advanced Apprenticeships	48	38,307	38,355	2.35	3.04	3.04
Total	2,045	1,260,349	1,262,394	100	100	100

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The Information on main educational pathway is taken from the feasibility AEDE. Identification of Gypsy, Traveller or Roma ethnicity uses combined information from the 2011 Census and the feasibility AEDE. Further details are provided in Obolenskaya 2022b.

Table 7: Highest educational attainment by age 17 (cohorts of young people who turned 15 in academic years 2008/9, 2009/10 and 2010/11 - pooled data)

	Unweighted frequency			Percentage		
	Gypsy, Traveller and Roma	All other	All	Gypsy, Traveller and Roma	All other	All
Level 1						
No	1,187	107,976	109,163	40.02	7.75	7.82
Yes	1,779	1,284,872	1,286,651	59.98	92.25	92.18
Total	2,966	1,392,848	1,395,814	100	100	100
Level 2						
No	1,837	305,790	307,627	61.94	21.95	22.04
Yes	1,129	1,087,058	1,088,187	38.06	78.05	77.96
Total	2,966	1,392,848	1,395,814	100	100	100
Level 3						
No	2,856	1,051,155	1,054,011	96.29	75.47	75.51
Yes	110	341,693	341,803	3.71	24.53	24.49
Total	2,966	1,392,848	1,395,814	100	100	100

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The Information on highest educational qualification is taken from the feasibility AEDE. Identification of Gypsy, Traveller or Roma ethnicity uses combined information from the 2011 Census and the feasibility AEDE. Further details are provided in Obolenskaya 2022a.

Table 8: Highest educational attainment by age 19 (cohorts of young people who turned 15 in academic years in 2008/9, 2009/10 and 2010/11 - pooled data)

	Unweighted frequency			Percentage		
	Gypsy, Traveller and Roma	All other	Total	Gypsy, Traveller and Roma	All other	Total
	Level 1					
no	1,037	81,779	82,816	34.97	5.87	5.93
yes	1,928	1,311,075	1,313,003	65.03	94.13	94.07
Total	2,965	1,392,854	1,395,819	100	100	100
	Level 2					
no	1,549	185,157	186,706	52.24	13.29	13.38
yes	1,416	1,207,697	1,209,113	47.76	86.71	86.62
Total	2,965	1,392,854	1,395,819	100	100	100
	Level 3					
no	2,484	556,539	559,023	83.78	39.96	40.05
yes	481	836,315	836,796	16.22	60.04	59.95
Total	2,965	1,392,854	1,395,819	100	100	100

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The Information on highest educational qualification is taken from the feasibility AEDE. Identification of Gypsy, Traveller or Roma ethnicity uses combined information from the 2011 Census and the feasibility AEDE. Further details are provided in Obolenskaya 2022a.

7.5 Experimental findings: the relationship between highest educational attainment by age 19 and childhood multidimensional disadvantage

We now turn to our analysis of the relationship between highest educational attainment by age 19 and multidimensional child disadvantage. Here, we exploit the rich information about individual characteristics and household circumstances recorded in the 2011 Census. Building on the framework for assessing multidimensional child disadvantage we developed in our previous study (Vizard et al 2018, Burchardt et al 2017), we have produced new experimental statistics on the relationships between the multidimensional child disadvantage characteristics of the young people we identify as Gypsy, Traveller or Roma observed at ages 14 or 15 (assessed using information recorded in the 2011 Census) and their highest educational attainment observed by age 19 (assed using information from administrative education data systems). The findings provide insights into the lives, circumstances and needs of children and young people from the communities and highlight the scale of the challenges they and their families face through adolescence and as they transition into young adulthood. In particular, they put the spotlight on inter-generational cycles of educational disadvantage and the importance of whole family support.

7.5.1 Assessing the multidimensional child disadvantage characteristics of the Gypsy, Traveller and Roma young people in Wave 1 GuiE

As noted above, due to the limitations of the GUiE data, for the multidimensional child disadvantage analysis we use household deprivation variables that have been derived from 2011 Census records by ONS (rather than using our own bespoke multidimensional child disadvantage indicators derived directly from 2011 Census individual level information). The ONS derived household deprivation categories are based on the following definitions:

- **Household employment deprivation:** Where any member of a household, who is not a full-time student, is either unemployed or long-term sick.
- **Household education deprivation:** No person in the household has at least Level 2 education and no person aged 16 to 18 is a full-time student.

- **Household health and disability deprivation:** Any person in the household has general health that is 'bad' or 'very bad' or has a long-term health problem.
- **Household housing deprivation:** The household's accommodation is either overcrowded, with an occupancy rating -1 or less, or is in a shared dwelling, or has no central heating.

Source: UKDS (n.d.).

Table 9 sets out our findings on the multidimensional child disadvantage characteristics of Gypsy, Traveller and Roma young people within Wave 1 GUiE who were aged 14 or 15 at the time of the 2011 Census (27th March 2011). The results reveal a systematic pattern of multidimensional disadvantage for Gypsy, Traveller and Roma young people across each of the indicators considered. At ages 14 or 15, young people who identify as Gypsy, Traveller or Roma are substantially more likely to be living in a single parent family; to be living in a family with three or more children; to live in privately rented or socially rented accommodation; to be living in over-crowded accommodation; to be living in a caravan or other mobile or temporary structure; or to experience ill health or a disability themselves.

The results in Table 9 also show that at ages 14 and 15, young people identified in GUiE Wave 1 as Gypsy, Traveller or Roma are substantially more likely to be living in a household classified as employment deprived, educationally deprived, housing deprived or health and disability deprived than their peers. Take childhood experiences of household educational deprivation; this is where the young person in question had no educational resources within the household to draw on when they were a child in the sense that no person aged 16 or above in the household had an educational qualification at Level 2 or above and no one in the household is a full-time student aged 16-18. As many as 51% of young people identified as Gypsy, Traveller or Roma were resident in households that are assessed as educationally deprived at ages 14 or 15 using this definition, compared to 18% of other young people. Additionally, 36% were resident in households that were assessed as employment deprived compared to 16.5% of other similar young people; 47.7% were resident in households assessed as health and disability deprived compared to 26.7% of other similar young people; and 38% were resident in households that were assessed as housing deprived, compared to 14.2% of other similar young people.

Overall, a staggering 86.5% of young people identified as Gypsy, Traveller or Roma who were aged 14 or 15 at the time of the 2011 Census were deprived in at least one deprivation dimension. Looking at disadvantage across multiple dimensions simultaneously, 19.8% were

deprived in three dimensions, compared to 5.1% of other similar young people; and young people from the communities were more than seven times as likely to be deprived in all four of these deprivation dimensions.

Table 9: Prevalence of multidimensional child disadvantage (children and young people in Wave 1 GUiE who were age 14 or 15 at the time of 2011 Census)

	Gypsy, Traveller and Roma young people aged 14 or 15	Other young people 14 or 15
Sex		
Male	46.60	50.20
Female	53.40	49.80
Family type		
Child in couple family	49.30	70.10
Child in single parent family	46.20	27.60
Other	4.50	2.40
Number of families in the household		
One	96.50	98.00
two or more families	3.50	2.00
Country of birth		
Born within the UK	77.20	92.00
Born outside the UK	22.80	8.00
Household tenure		
Owned(incl.shared ownership)	26.20	66.10
Private rent	45.90	21.00
Social rent	26.00	12.10
Rent free	1.90	0.80
Type of accommodation		
House/Flat	89.50	99.90
Caravan/mobile or temp structure	10.50	0.10
Household overcrowded (bedroom standard)		
overcrowded	31.20	11.10
not overcrowded	68.80	88.90
Number of dependent children in the family		
No dependent children	N/A	N/A
One or two	52.20	71.20
Three or more	47.80	28.80

Poor/fair general health or limiting long-standing illness (own)		
No	87.80	93.80
yes	12.20	6.20
Household education deprivation (ONS)		
No	48.80	81.80
yes	51.20	18.20
Household employment deprivation (ONS)		
No	63.80	83.50
Yes	36.20	16.50
Household health and disability deprivation (ONS)		
No	52.30	73.30
Yes	47.70	26.70
Household housing deprivation (ONS)		
No	62.00	85.80
Yes	38.00	14.20
Deprivation in multiple dimensions simultaneously		
Not deprived in any dimension	13.50	50.10
deprived in 1 dimension	28.80	30.30
deprived in 2 dimensions	33.40	14.00
deprived in 3 dimensions	19.80	5.10
deprived in 4 dimensions	4.50	0.60
deprived in at least 1 dimension	86.50	49.90

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The individual and household characteristics in this table are derived from information recorded in the 2011 Census. Identification of Gypsy, Traveller or Roma ethnicity uses combined information from the 2011 Census and the feasibility AEDE. Further details (included unweighted frequencies) are provided in Obolenskaya 2022a.

7.5.2 Assessing the relationship between highest educational attainment by age 19 and multidimensional child disadvantage

Highest educational attainment by age 19 by characteristics recorded in the 2011 Census

Table 10 sets out our findings on the relationship between educational attainment by age 19 and different individual and household characteristics

recorded in the 2011 Census for the GUiE cohorts that turned 15 in academic years 2008/9, 2009/10 and 2010/11. The results show that amongst Gypsy, Traveller and Roma young people in the cohorts, females were more likely to gain Level 2 or Level 3 qualifications by any route by age 19 than males. Rates were 53% for females compared to 43% for males in relation to Level 2 qualifications, and 18% for females compared to 15% for males in relation to Level 3 qualifications. Similar patterns of differentiation are also apparent amongst non-Gypsy, Traveller and Roma young people in the cohorts, with rates of 90% for females compared to 84% for males in relation to Level 2 qualifications, and 65% for females compared to 55% for males in relation to Level 3 qualifications.

Table 10 reveals several other factors that are associated with lower attainment amongst young people in the cohorts. This includes living in a single parent family, living in a family with three or more children, experiencing a longstanding illness or disability, living in rented accommodation (both privately rented and socially rented) compared to living in owner occupied accommodation and living in a caravan or mobile or temporary structure rather than in a house or flat. Particularly low rates of attainment are observed amongst Gypsy, Traveller and Roma young people who were resident in single parent households or who live in a caravan or mobile or temporary structure. For example, only 43% of Gypsy, Traveller and Roma young people who lived in a single parent family attained Level 2 qualifications by age 19 and only 12% attained Level 3 qualifications. Amongst all young people in the cohorts, living in a caravan or other mobile or temporary structure is associated with substantially lower attainment. Amongst Gypsy, Traveller and Roma young people, only 43% of those living in a caravan or other mobile or temporary structure had attained Level 2 and 13% Level 3 by age 19. Rates amongst Gypsy, Traveller and Roma young people living in socially rented households were even lower (with 39% attaining Level 2 and 11% Level 3 by age 19).

Findings by country of birth are more complex to interpret, with lower rates of Level 2 attainment amongst Gypsy, Traveller and Roma young people who were born outside the UK but with observed gaps in Level 3 attainment being substantially narrower. Additionally, amongst non-Gypsy, Traveller and Roma young people in the cohorts, *higher* rates of attainment at age 19 are observed for those born *outside* of the UK for both Level 2 and Level 3 qualifications.

Table 10: The relationship between educational outcomes at age 19 and characteristics recorded in the 2011 Census (cohorts of young people who turned 15 in academic years 2008/9, 2009/10 and 2010/11 – pooled data)

	Level 2 and Level 3 qualifications attained by any route by age 19 as recorded in Department for Education administrative records (%)			
	% achieved Level 2		% achieved Level 3	
Individual and household characteristics recorded in the 2011 Census	Young people from the Gypsy, Traveller and Roma communities	Other young people	Young people from the Gypsy, Traveller and Roma communities	Other young people
Sex				
Male	42.69	83.89	14.55	55.4
Female	52.66	89.55	17.84	64.72
Family type				
In couple family	54.07	89.86	21.36	65.37
In single parent family	43.19	80.78	11.49	49.1
Country of birth				
Born within the UK	50.2	86.57	16.51	59.54
Born outside the UK	39.34	88.3	15.36	66.01
Household tenure				
Owned(incl.shared ownership)	63.84	91.99	27.92	69.11
Private rent	43.01	73.29	11.72	37.05
Social rent	38.78	80.51	10.8	48.26
Type of accommodation				
House/Flat	48.36	86.76	16.52	60.04
Caravan/mobile or temp structure	42.58	75.15	12.89	42.84
Number of dependent children in the family				
One or two (incl. none)	51.7	87.6	18.37	61.24
Three or more	43.89	85.77	14.01	58.52
Poor/fair general health or limiting long-standing illness (own)				
no	49.28	88.29	16.82	61.69
Yes	37.64	62.79	12.36	35.13

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The individual and household characteristics in this table are derived from information recorded in the 2011 Census. Identification of Gypsy, Traveller or Roma ethnicity uses combined information from the 2011 Census and the feasibility AEDE. Further details (included unweighted frequencies) are provided in Obolenskaya 2022a.

Highest educational attainment by age 19 by multidimensional child disadvantage

To examine the relationship between highest educational attainment by age 19 and multidimensional child disadvantage, we focus on the single analytical cohort of young people who turned 15 during the academic year 2010/11 – that is, the academic year of the 2011 Census. Table 11 sets out our findings on the relationship between educational outcomes at age 19 and multidimensional child disadvantage observed at ages 14 or 15 using the four ONS derived variables discussed above (household employment deprivation, household education deprivation, household health and disability deprivation and household housing deprivation).

The overall patterns of attainment at age 19 for this cohort are very similar to the findings for the three cohorts combined. Overall, as many as 40% of young people did not progress to attain Level 3 qualifications by any route by age 19. The percentage of Gypsy, Traveller and Roma young people *not* securing Level 3 education by any route by age 19 was more than double that of other young people (87%). Risks were also higher in relation to Level 2 education (with 53% not securing full level 2 qualifications compared to 13% of other children) and Level 1 education (37% compared to 6%).

The results show that educational attainment at age 19 is substantially differentiated by household level deprivation factors for both the Gypsy, Traveller and Roma cohort and the non-Gypsy, Traveller and Roma cohort. However, as attainment at age 19 is substantially lower for Gypsy, Traveller and Roma young people than for other young people, the lowest rates of attainment are observed for Gypsy, Traveller and Roma young people who lived in deprived households as children aged 14 or 15. Additionally, the risks of educational disadvantage increase with multiple deprivation (i.e. where more than one deprivation factor is recorded). In total, 43% of Gypsy, Traveller and Roma young people with 3 or 4 household deprivation factors at age 14 or 15 did not go on to attain Level 1 qualifications by age 19, 62% did not go on to secure Level 2 qualifications and 90% did not go on to secure Level 3 qualifications.

While rates of low educational attainment at age 19 are substantially lower amongst non-Gypsy, Traveller and Roma young people in the cohort, the scale of low attainment at age 19 amongst other young people who experienced multidimensional deprivation at age 14 or 15 is also strikingly high. Amongst the non-Gypsy, Traveller and Roma young people in the cohort who recorded multidimensional household deprivation (household deprivation in 3 or 4 domains) at ages 14 or 15, 17% had not secured Level

1 qualifications by age 19, 30% had not secured Level 2 qualifications by age 19 and 65% had not secured Level 3 qualifications by age 19.

Inter-generational cycles of educational disadvantage

Concerns about the multidimensional transmission of educational disadvantage were expressed in witness testimony to the Women and Equalities Committee inquiry in 2019. Table 11 directly addresses this concern as GUiE enables us to observe educational deprivation within a young person's household (at ages 14 or 15) and to relate this to the young person's own educational deprivation status (observed at age 19). The research identifies that 89% of Gypsy, Traveller and Roma young people who experienced household educational deprivation as children aged 14 or 15 at the time of the 2011 Census had not secured Level 3 qualifications four years later by age 19. In addition, 59% failed to attain a Level 2 qualification and 42% failed to go on to attain a Level 1 qualification by age 19.

Rates of inter-generational transmission of educational disadvantage amongst non-Gypsy, Traveller and Roma young people in the cohort are somewhat lower; but are also striking. Amongst the non-Gypsy, Traveller and Roma young people in the cohort, 59% of those who experienced household educational deprivation as a child aged 14 or 15 did not attain Level 3 qualifications by any route by age 19, with 24% and 12% failing to attain Level 2 and Level 1 qualifications respectively.

Table 11: The relationship between educational outcomes by age 19 and multidimensional child disadvantage at ages 14 or 15 (cohort of young people who turned 15 in academic year 2010/11)

Childhood deprivation characteristics (household level deprivation, recorded when young people in the cohort were aged 14 or 15 in the 2011 Census)	Highest educational qualifications at age 19 (recorded in administrative education data for academic year 2014/15)					
	% who had *not* attained Level 1		% who had *not* attained Level 2		% who had *not* attained Level 3	
	Gypsy, Traveller, Roma	Other	Gypsy, Traveller, Roma	Other	Gypsy, Traveller, Roma	Other
All	37.04	6.06	52.57	13.08	86.86	39.88
Male	39.55	7.49	56.39	15.93	89.05	44.7
Female	34.96	4.61	49.41	10.19	85.04	35.01
Household education deprivation	42.03	12.02	59.31	23.62	88.87	59.1
Household employment deprivation	41.75	11.93	60.19	22.84	90.05	56.16
Household health and disability deprivation	38.68	11.31	54.71	20.4	89.58	50.08
Household housing deprivation	42.25	10.13	55.35	20.02	89.84	51.49
Not deprived in any dimension	23.13	2.33	36.88	6.97	75	29.51
Deprived in at least 1 dimension	39.43	9.78	55.23	19.21	89	50.43
Number of deprivation dimensions = 1	33.54	7.3	48.78	15.44	86.89	45.08
Number of deprivation dimensions = 2	42.35	12.34	56.47	23.32	90.59	56.31
Number of deprivation dimensions = 3-4	43.2	16.84	62	29.52	89.6	64.86

Source: All statistical results remain Crown Copyright. Note: Experimental statistics. The analysis was undertaken in the Secure Research Service (SRS) by Obolenskaya using Wave 1 GUIE (merged cases only). The full citation for this dataset is: Office for National Statistics, released 14 June 2021, ONS SRS Metadata Catalogue, dataset, Growing Up in England Wave 1, https://doi.org/10.57906/ct4b-k289. The individual and household characteristics in this table are derived from information recorded in the 2011 Census. Identification of Gypsy, Traveller or Roma ethnicity uses combined information from the 2011 Census and the feasibility AEDE. Further details (included unweighted frequencies) are provided in Obolenskaya 2022a.

7.6 Implications: the importance of tailored support for Gypsy, Traveller and Roma young people and their families at all stages of education

As noted above, the experimental statistical findings set out in this section require careful interpretation and caveating in the light of the limitations of the GUiE dataset identified in sections 5 and 6 as well as additional issues identified in our literature review and direct engagement exercise. In particular, we acknowledge that the discussion in section 6 highlights a series of concerns around imperfect ethnicity identification, missingness, disproportionality and compositional effects. These limitations restrict the extent to which the GUiE Wave 1 dataset can be thought of as being representative of the underlying population of Gypsy, Traveller and Roma young people and the extent to which generalisations and inferences should be drawn from our findings.

Nevertheless, while recognising these limitations and the importance of careful interpretation, our findings demonstrate the important and innovative role that administrative-data population census linkages can play in challenging data exclusion and building up new evidence and insights on groups that have previously been largely missing from national statistics. We have demonstrated that it is possible able to identify much larger analysis samples of Gypsy, Traveller and Roma young people than has been possible in most previous studies using the GUiE Wave 1 dataset. This advance results from both the scale of the dataset as well as the ability to pool ethnicity information from the 2011 Census and longitudinal education data. Additionally, the analysis shows that it is possible to follow the educational journeys of Gypsy, Traveller and Roma young people using the feasibility AEDE data and to assess strengths of the relationships between multidimensional child disadvantage and highest educational attainment by age 19 using rich information on individual and household circumstances recorded in the 2011 Census. These are important steps forward and highlight the key role that administrative data–population census linkages can play within broader efforts to develop inclusive data infrastructure and evidence for the 2020s.

Overall, the findings provide important new insights into the educational journeys of children and young people from the communities and the scale of the challenges that they and their families face as they transition from adolescence into early adulthood. In terms of policy implications, the research exercise underlines the critical importance of policies, educational pathways and support packages that are specifically tailored to meet the needs of children and young people from the Gypsy, Traveller and Roma

communities and their families across all educational pathways and routes and at all stages of education. These findings, together with the qualitative findings from the direct engagement exercises reported in section 8, inform the policy recommendations that we include in the concluding section of this report.

Additionally, the research exercise lays the foundations for a more detailed and richer future cohort study focussing on the lives, circumstances and needs of Gypsy, Traveller and Roma young people. Looking forward, we plan to up-scale this study and to undertake a larger programme of research, incorporating into our analysis additional data on exclusions, absences, SEND, drop-out rates and interactions with children's social care services released in Waves 2 and 3 GUiE, together with additional data made available in complementary large-scale administrative data linkages.

8 Qualitative findings from two direct engagement workshops with children and young people from the Gypsy, Traveller and Roma communities and their families addressing the issues raised by our project

In this section, we provide a record of the feedback from two direct engagement workshops with young people 19 from the Gypsy, Traveller and Roma communities and their families. The direct engagement workshops were co-organised with and facilitated by the national charity Friends, Families and Travellers (gypsy-traveller.org) and were held in July 2022. Our aim in undertaking the direct engagement workshops was to hear directly from children and young people from the Gypsy, Traveller and communities about the issues raised by our project. Feedback was invited on experiences of the education system (including non-mainstream provision and home education as well schools, colleges) and in relation to administrative data collection and use (including experiences, views and perspectives relating to ethnicity monitoring and the identification of Gypsy, Traveller and Roma people within administrative data collections).

Embedding a direct engagement exercise within an administrative data research project provides recognition of the substantive importance of direct experiences, perspectives and views for understanding, interpreting and framing statistical findings and for drawing overall project conclusions and recommendations. Our motivation in undertaking the direct engagement exercise has been to move away a passive 'data subject' approach to research design and to recognise the substantive role of voice and agency within in administrative data research. The direct engagement exercise also builds on ADR UK public engagement initiatives and has enabled us to address specific recommendations of the National Statistician's Data Ethics Advisory Committee (NSDEC) relating to our project as well as broader NSDEC ethical principles, including those relating to engagement, transparency and accessibility (c.f. section 1.3).

This section begins with some background relating to the broader public engagement exercises being undertaken by ADR UK and partners around administrative data and the public good and the GUiE data linkage. We then set out the direct engagement workshop methods (including the processes of recruitment and for obtaining informed consent and the numbers and composition of participants). Next, we provide a record of the feedback from the first part of the workshop, which focused on experiences of the education system (schools, colleges and training) and home education. We then move on to the second part of the workshop, and provide a record of

experiences and view-points in relation to official data collection and statistics, including issues around the disclosure of Gypsy, Traveller and Roma identity. We finally reflect on how the findings from the engagement workshops have fed into the overall project recommendations set out in the conclusion to this report (section 9).

8.1 Background: broader engagement exercises in the context of administrative data research and the development of the GUiE dataset

Administrative Data Research UK (ADR UK) and the Office for Statistics Regulation have taken forward an engagement exercise on public attitudes to data use. An initial analysis of public attitudes towards sharing and linking of administrative data found that the public are broadly supportive of the use of administrative data for research purposes provided the research is in the public interest, that data subjects are protected from identification (and re-identification) and a framework of trust is in place (ADR UK 2020). In 2021, ADR UK launched a Public Engagement Strategy which defines public engagement as a range of different activities that are designed to promote an ongoing dialogue with the public about administrative data research, driven by active listening and responding (ADR UK 2021). Different researcher understandings of the public good were examined in Office for Statistics Regulation (2021). Findings from a public dialogue on what the public perceive as 'public good' in the use of data for research and ethics underlined the importance of public involvement; research and statistics that addresses real-world needs; clear and accessible communication; minimisation of harm; and universal application of best practices in relation to data security, confidentiality and privacy, to maximise trust (Office for Statistics Regulation and ADR UK 2022) (c.f. Gordon and Humpherson 2022).

In terms of public engagement exercises that specifically relate to the Growing Up in England (GUiE) dataset, there were two initial stakeholder meetings held in 2018 and 2019 respectively (see Waind 2019). Additionally, the ADR England Children & Young People Representative Panel is made up of experts from civil society organisations which represent children and young people and their interests in relation to issues raised by administrative data research. This provides an additional opportunity for community representatives to raise ethical or other concerns or to address how risk of harms can be mitigated, with a view to increasing public acceptability and establishing that administrative research is in the public interest. The panel has considered the issues raised by the GUiE linkage, with key messages and feedback set out in ADR UK (2022), and includes a

civil society organisation representing the Gypsy, Traveller and Roma communities (Waind 2021, Kashef 2023).

Additionally, as noted in the introduction to this report, in undertaking the current project, and building on specific recommendations of NSDEC we have taken several additional steps. First, the civil society organisation Friends, Families and Travellers has been integrally involved in the project since its inception and has participated in our Advisory Group. Second, working with Friends, Families and Travellers, we embedded a direct engagement exercise into our project design. Third, we have produced accessible outputs that are specifically designed for onward sharing with children and young people from the communities and their families.

8.2 The two direct engagement workshops: organization, methods and informed consent

8.2.1 Recruitment and participation

Young people aged 14-29 and their broader family groups (for example, parents, siblings, aunts, uncles, grandparents and carers) were invited to attend and participate in the engagement workshops. The extension of invitations to participate to broader family groups as well as young people themselves reflects advice from Friends, Families and Travellers in relation to good practice. Recruitment was undertaken through Friends, Families and Travellers outreach teams through established networks and using face-to-face methods. The aim was to achieve a balance (1) from all three communities: Gypsy, Traveller and Roma; (2) by gender; (3) across the age range most relevant to the GUIE Wave 1 dataset (14-29). In total, 23 participants attended the workshops (nine attendees at the first workshop and 14 at the second). Of this total, the vast majority of participants were direct invitees, with two additional participants attended as broader family member chaperones.

Direct invitees included individuals from across the 14-29 age spectrum and included children and young people as well as working age adults and parents, some of whom were parents or carers and attended with their own children. Participants included Romany Gypsies, Traveller and Roma young people and family groups, ensuring diversity of experiences and perspectives. At the first workshop, participant backgrounds were six Romany Gypsies and three Travellers. At the second workshop, participant backgrounds were 11 Travellers, one Romany Gypsy and two Roma. All participants in the workshops received vouchers as compensation for their time (including young people and their broader family groups and siblings).

8.2.2 Informed consent

Paper copies of project information forms were provided to participants by the outreach teams as part of the recruitment process. Young person and adult friendly versions of the project information form were developed and as literacy is a major concern the out-reach teams provided verbal explanations of both the project information sheet and the informed consent process. Before the workshops began, participants were reminded of what would happen at the workshops and we paused to provide time for questions and for individuals to provide informed consent. For under 18s, consent from both parents/guardians and the young people themselves was obtained. We provided options for both written and verbal consent, although in practice informed consent for all participants was provided as a written signature.

8.2.3 Facilitation, introductory discussion and warm up

The workshops were facilitated by experienced Friends, Families and Travellers staff. The workshops were audio-recorded and detailed notes were taken on the day. Although we planned options for smaller breakout groups and visual storytelling exercises, in practice, oral discussions with family groups as a plenary were the most productive and preferred method of discussion on the day.

The workshops began with an introductory discussion by Friends, Families and Travellers and the LSE research team about the aims of the day and the GUiE project. The starting point of the discussion was the importance of inclusive data and evidence to understand the needs of children and young people from the Gypsy, Traveller and Roma communities. By way of introduction, we highlighted the importance of tackling data exclusion and ensuring that data and evidence captures and reflects the lives, experiences and outcomes of everyone in society. We also explained that researchers from the London School of Economics and national charity Friends, Families and Travellers (FFT) are working together to tackle data exclusion, and develop better data to better support children and young people from the Gypsy, Traveller and Roma communities. In addition, we highlighted the research as contributing to these broader efforts using the newly linked Growing Up in England (GUiE) dataset and explained that GUiE includes the records of around 7 million children but is 'de-identified' - that is, it is not possible to identify specific individual people from the dataset.

As part of the introductory discussion, the FFT team also explained to participants how the day would work and highlighted that participation was

entirely voluntary and that participants could leave at any time without any adverse consequences. Participants were reminded about the information included in the project information forms. The session was then paused to enable the FFT team to obtain informed consent from all participants.

8.2.4 Organisation of the day

Following on from the introductory discussion, there were two main parts of the day, before and after lunch. During the first part of the day, we invited participants to talk to us about their experiences of the education system – what had gone well and what hadn't gone well - and of home education where relevant. During the second part of the day, we focussed on data collection and data use, including the use of data that doesn't use real names (de-identified data) but that reports on Gypsy, Traveller and Roma identity. We asked about whether participants felt that it was important for public services like schools and hospitals to record Gypsy, Traveller and Roma. Finally, we asked whether participants felt comfortable revealing their identities at school and in other environments such as hospitals and at work, the issues this raised for them and what their experiences of this had been, good and bad.

8.2.5 Feedback

In the write up of the findings from the engagement workshops below, we use the term 'child/young person' or 'young person' to refer to feedback from direct invitees to the workshops (that is, young people from the Gypsy, Traveller and Roma communities aged 14-29). As noted above, this includes individuals from across the 14-29 age spectrum including a mix of school age children, adolescents, young people, working age adults and parents (some of whom attended the event with their own children). Feedback from younger siblings (who might be aged under 14) who participated in the workshops is described in the same way (as feedback from a child or young person). Additionally, below, we refer to feedback from 'parents/carers/member of family group' – this includes feedback from directly invited participants who are parents or carers and attended the workshops with children; and others who attended as family chaperones (for example, grandparents, aunts, uncles and carers). Finally, we use the term 'family group' where the feedback is part of a collective discussion amongst a family group.

Findings are written based on detailed notes from the day and audio recordings. We capture the views, experiences and perspectives expressed in the discussions as closely as possible. As the numbers at the workshops

were limited, disaggregated feedback by Gypsy, Traveller or Roma young people and their families has not been possible. We acknowledge that the extension of invitations to broader family members raise some issues of power relating to the ability and willingness of children and young people to speak freely.

8.3 Feedback on experiences of the education system and home education

Participants at the workshops told us about their experiences of the education system – what had gone well and what hadn't gone well – and about home education where relevant. Participants were invited to share their educational journeys – whether they are in school or college, whether they are home schooled, whether they have had education interrupted for any reason, and what their experiences of school, college or home schooling had been.

8.3.1 Worse experiences of education linked to identity

We learnt that most participants – young people as well as parents, carers and siblings - felt that they have worse educational experiences because of their identity.

- **Lack of respect, hostile treatment and bullying.** Several participants reported being treated with a lack of respect at school and some participants reported instances of hostile treatment, stereotyping and bullying. While there were some references to the social side of school and to friends as being positives, several participants shared experiences of themselves or younger siblings or broader family members being upset, picked on, called names or being bullied. Some parents and carers told us about raising incidents with teachers but not having an adequate response. Young people reported teachers not doing anything about the poor behaviour of other young people when they or their siblings or broader family members were picked on or bullied. One family told us about an incident which they felt escalated because of inaction by teachers and there were references to situations escalating to physical fights or altercations. Participants told us that young people from the communities get the blame with assumptions being made about what had happened and that they then are singled out for punishments such as exclusions or isolation. There were references to not being able to rely on teachers to sort a situation out and references to young people from the communities having to stick up for themselves and young people feeling they had to take matters into their own hands.

One mother told us that one of the teachers at the school said that the Traveller young people were like wolves sticking together. One participant referred to being bullied when they lived on a site while they were at primary school but that when they moved to a house it was easier to blend in.

- **Lack of support for young people with special educational needs**. There were several reports of a lack of support for children and young people living with neurodiverse conditions, autism, ADHD, dyslexia and other special educational needs and disabilities. Some families felt that this lack of support and intervention was linked to their Gypsy, Traveller or Roma identity. A member of a family group referred to a lack of support and interventions. She referred to more than one of her children needing specialist support but that this was not being provided. She felt that this was because of who they are and that other people got the support and interventions they need. There was a reference to one good support assistant who helped a son to start to read; however, she had to leave and there were also difficulties when she was threatened with disciplinary action when she stood up for him after a fight. One mother told us that her son found it difficult to sit still in a classroom all day but that the curriculum had to be stuck to. She explained that her son did much better learning outside and her own experiences with using chalk outside with her son had been positive. The SENCO had helped and there were positive experiences of learning outside and cooking with one learning assistant. However, the support was stopped and she felt that she hadn't been given a valid reason why it couldn't continue. She felt unsupported by the headteacher and also frustrated that the school has to do what's in the government curriculum. The mother referred to her own experiences of cooking outside and making pegs outside with her son as her own Mum used to do. Another participant told us about a Forest School with outside learning and making tree swings and marshmallows.
- **Experiences of punitive sanctions**. Several participants were struggling with repeated punitive sanctions by school including long and repeated periods of exclusions as well as isolation and fines. There was a perception that these were common events, rather than one offs. Generally, the feeling was that these punitive actions were unfair and used disproportionately against them and that this was linked to identity. There were reports of children and young people from the Gypsy, Traveller and Roma communities experiencing exclusions and isolation when situations involving lack of respect, being picked on or bullying had escalated. In some instances, exclusions and isolation impacted on children with special educational needs and disabilities. Children and young people reported concerns about having nothing to do in isolation rooms and that this was putting them off school. Parents

and carers expressed similar concerns and told us that their circumstances were not adequately recognised by schools. There was a reference to fines being imposed when children were absent from school to attend cultural events such as the Appleby Horse Fair and being followed up by an attendance officer when children attended a family funeral. One mother felt there was a lack of understanding by the school when her children were late, as she had to take seven children to school, and dropped the older children later.

- **Repeated and multigenerational adverse experiences**. These adverse educational experiences were in some instances reported across multiple generations of the same families, with parents and carers reporting their own experiences of lack of respect, hostile treatment and bullying as well as segregation and punishments when they were at school. Adverse experiences were reported by both mothers and fathers. One Dad recalled his brothers and sisters being bullied at primary school as well as having fights and getting kicked out and then not going to a 'proper' secondary school. Some participants reported these adverse experiences as part of day-to-day life, rather than as one-off events. Some participants commented that this is just the way things are, and that things wouldn't change. However, others highlighted some positive experiences and specific recommendations about what would make things better.

Quotes

Lack of respect, hostile treatment and bullying

"How do you want the Traveller kids to be good if they don't get any respect at school?" [child/young person]

"I got in a fight because I got called a p*key ... I got the isolation and all that and he didn't ..." [child/young person]

"She got excluded from school because some mean girls were bullying her little sister … we told the teachers and they didn't do anything ..." [parent/carer/member of family group]

"I love my little girl's playschool … if another child says something they'll pull out the mum and dad and say look this is what they've done and we're not going to put up with it … whereas at my kids' school they won't bother … they say it's tit for tat … it's a different standard … lack of doing anything …" [parent/carer/member of family group]

Punitive sanctions, exclusion and isolation

[He] [g]ets an hour a day for one day a week of schooling - in school support - everything he does is 'wrong' but he's always punished and kicked out for things even when other children are doing it or worse" [parent/carer/member of family group/family member]

"They excluded her for a week …. she had to stay home … they didn't give her anything to do … they had to sit in isolation … you just have to sit there for a week …. the school is aware that she has ADHD and everything …. so there were two weeks of nothing …" [parent/carer/member of family group]

Parents and carers recollections of their own experiences at school

"I used to always have fights so I'd get kicked out all the time … I never got put into a proper school - the secondary school was like 6 people in a class - at primary school I was bullied all the time … my brothers and sisters got bullied there as well …" [parent/carer/member of family group]

"I went from year 7 all the way through to year 11 … every single lesson and every single teacher I got bullied after bullied after bullied for coming from where I come from … It got a little bit better at year 11 … its only because we basically just didn't talk to the other children it was just the Travellers … I wouldn't mix with the gorjas … we were stuck into one classroom .. that was us …" [parent/carer/member of family group]

"[T]he whole five years I was there was - well - it was living hell, cause everyone knew … everyday it was p*key, caravan on wheels … house on wheels … I couldn't bear school … hated it, I only went because me Mum made me … just because me Mum told me to go" [parent/carer/member of family group]

Quotes

Concerns about lack of support for children and young people with special educational needs and disabilities

"He can't sit in a classroom … he wants to be active … … he can't cope with noises and being still … [there's] noise hollering and you have to sit there from 8.30 to 3 …"

"I do work with him at home … I go and sit the garden and do work with him in the garden … you know like chalk … I've got a wall in the garden … I put his maths numbers … and he'll copy it on the wall …"

"At school he used to have one of them learning assistants … she used to get chalk [and] take him on the playground .. and put the spellings that he needed on the floor … he was loving it …"

"But in September changed it … he doesn't have that any more … now it's going to go back to square one … he's not going to learn because he's stuck in the classroom again … but the other teacher she realised if you take him in the playground or the big field they've got he's going to learn, he's going to do it…"

"He loves to cook .. make brownies … so she [the learning assistant would] take him in the cooking room just her and him ... she's like you need to weigh out 100 … butter… he would read 'sugar' and repeat [it].. that's how he would learn .. [but it was] cut off in September [when] he was making progress … I think it's ridiculous."

"I know he's not going to secondary school … he ain't going to sit down in the classroom and do it … and that's what worries me … if he goes to secondary he won't be able to cope … different classrooms … he can't even sit in a classroom for an hour and then go to another classroom and for an hour … it's not going to work".

[parent/carer/member of family group]

8.3.2 Limited options and opportunities

Participants told us that the focus and requirements of schools and colleges often does not recognise and support the needs of children and young people from the Gypsy, Traveller and Roma communities. Many participants felt that their options and opportunities were extremely limited and that these were worse because of their identity. Many participants referred to a lack of valued opportunities and options. Subject content and subject requirements were viewed as barriers to educational participation and progression.

- **Limited subject relevance.** Several participants expressed the view that at secondary school there were too many subjects that were irrelevant to the needs of children and young people from the communities – although English, maths and science were viewed as important. Subjects that were mentioned as being irrelevant to their needs and futures included geography and history - some felt these subjects were a waste of time ("what do I need history for, what do I need geography for"? [child/young person]). One young person commented that half the stuff they learnt they didn't need – they wanted more about the outside world. Another referred to hours of boredom. The importance of more practical options focused on the outside world were noted. Subjects of particular interest that were mentioned included hair and beauty and plumbing and electrician courses.
- **Practical options at an earlier age.** There was a concern from some parents that practical subjects are available too late and that by then the children and young people have fallen too far behind. A specific issue was raised about funding not always being available for 14 year olds to attend college to complete technical qualifications alongside general qualifications such as English and maths at key stage 4.
- **Formal subject requirements as barriers to 16-18 educational participation.** Where young people remained in the school system in key stage 4 (the year that GCSEs are taken), concerns about progression into key stage 5 were raised as a specific concern, with formal subject requirements including GCSE maths and English being identified as a particular barrier after year 11. GCSE requirements for apprenticeships and courses at ages 16-18 were highlighted as an important concern. Some parents felt that formal requirements for post-16 training and apprenticeships are now higher and more insurmountable than they were in the past, based on their own experiences of training and college. One mother was concerned that her children, currently in year 11, wouldn't meet these requirements at the end of the year. She didn't know what they would do as her child was unlikely to attain the necessary formal requirements. She felt it was

harder now than in the past and recollected her own (positive) experience of having not learnt at school but as being able to take an NVQ level 2 in childcare after leaving school and working in a nursery. Other participants highlighted positive examples of support with gaining the necessary qualifications to progress, such as support at college with maths and English skills while taking a bricklaying or plumbing course.

- **Limited opportunities for work experience were raised as a specific concern.** One participant referred to lack of support with work experience. One young person felt that there were extremely limited opportunities for her – that she had been rejected from most opportunities and that the restricted opportunities open to her were not gender appropriate. There was a careers day at school, the only person who was interested in her was a tree surgeon. She was accepted for who she was and didn't need qualifications but it's for boys so she couldn't do it. Another young person reported not having work experience and having to stay at school instead and a mother highlighted that one difficulty is that girls would not be allowed to go unaccompanied for example to do work experience at Tesco with lots of strangers.
- **Barriers to progression linked to identity**. Several participants reported their identity being a barrier to their progression and the need to conceal identity to get on at school, college and in relation to work experience. There were references to participants hiding their identity in order to get on and some participants questioned why they should have to do this. One participant referred to wanting to become an air hostess, but that she would have to make out that she is not a Traveller to get that job. Another referred to needing to hide identity in the context of apprenticeships. One participant referred to their experience of being at college on a flower arranging course and arranging work experience. She had been doing really well but when her Traveller identity was discovered, she didn't progress.
- **Importance of role models.** One participant referred to one of the children in his broader family finishing at primary school and a broader family debate about whether she should go to secondary school. She is going to go. One issue had been that the girl in question was very scared but the participant was able to help out, recollecting that when he was her age he had no one - no one went to secondary school and he was the first one going to secondary school let alone university. It was important for him that he felt like he had paved the way for her.

Quotes

Limited subject relevance and the importance of practical and vocational subjects and life skills

"We go to school but we don't like it .. there's nothing for us .. they need to teach us differently … what they teach us … I don't really need it …" [child/young person]

"They waste my day - I could go in there for three hours … I could do English, maths and science … and I would be fine – but I do six hours of like boringness - what do I need history for? what do I do geography for? ... I don't need them … it goes through one ear … and out the other …" [child/young person]

"… [h]alf the stuff they learn they don't need ..." [parent/carer/member of family group]

"He's dropped history, dropped geography, dropped languages - he just does English, maths and science and what he wants to do … he was just playing up in lessons … didn't want to go to school" [parent/carer/member of family group]

"Schools don't learn them anything about the outside world … like if they want to do hair and beauty they don't teach anything like that .. if they want to be a plumber or gas engineer they don't teach them anything like that … they don't teach them how to cook unless you … do that lesson .. I think that they need to learn them what they are going to need in life …" [parent/carer/member of family group]

"If you're interested in something you're gonna learn it, if you're not interested you're not gonna learn it…. you might not be the best reader or writer … but you could be a good hairdresser… but I don't think children get the opportunity anymore" [parent/carer/member of family group]

Identity as a barrier to progression

"I applied for college while I was in the school … one day a week … I got to the point where I couldn't sit there all day and do the same thing … I got myself work experience once a week for two years .. [but] as soon as the school told them I was a Traveller … no, they don't need me .. can't have me" [parent/carer/member of family group]

Quotes (continued)

Formal subject requirements as barriers to 16-18 educational participation and progression [Family group]

"… but that's when they're older and you have to get certain GCSEs to get into that course, but if they're not learning anything before that it's not going to make a difference … I think that's why children part and don't like school because they're not interested in what they're learning … " [parent/carer/member of family group]

"… [you need] maths, English, science" … [child/young person] … "and some of them you need like a C in your GCSE …" [parent/carer/member of family group]

"[W]hen I went to school, I didn't learn at school, I didn't get no GCSEs, nothing .. I worked in a nursery with my nan ... and then went to college and done my nvq level 2 … I learnt more at college, doing something I wanted to do … I didn't have no GCSEs, so I literally just left school and was working and then did my nvq 2 level after in college … they're asking a lot more of children these days, a lot more …" [parent/carer/member of family group]

Positive experiences of combining core qualifications and more practical subjects

"My mate did a bricky course and left in year 10 – he went to a college and got a college course - they let him do his GCSEs and helped him with the maths and English and science in the course … he got straight in …. he did his maths, his English … then he did his course … [facilitator: is that a good approach?] .. of course it is .. if I want to go on a course … if I don't have it I can go into the college … I can learn it, what I need to know .. I can do it alongside it … if I start to progress I am actually doing something" [child/young person]

Limited work experience opportunities [Family group]

"[T]hey just had work experience …[she] wanted to do hair and beauty .. there was a training place down the road .. no sorry you need to have this, you need to have that... When you say you're a Traveller or any thing like that you ain't going to get a job anyway ... there's a lot of like stereotyping of Travellers" [parent/carer/member of family group]

"I don't get it … we're only in year 10, we haven't even done our GCSEs … it's meant to help us .. we just get rejected by it all the time … there's just no point" [child/young person]

"She didn't even have work experience … so she stayed in school … they didn't find anything … and obviously … being Travellers [she might not be allowed to go to] certain places …. They have to be with someone … [she wouldn't be allowed] to work in Tescos for work experience with loads of strangers" [parent/carer/member of family group]

8.3.4 Home education and support with education at home

Several young people we talked to had left the formal education system or were not participating in the formal education system and they and their parents/carers told us that they were being home schooled. Some parents told us about their experiences with home education during the pandemic.

- **Home education.** Some participants had attended primary school but were home schooled at secondary level. Reasons for and experiences of home schooling were mixed. One participant told us that being home schooled he felt free and able to get the information from the internet. He felt more comfortable at home and less restricted. He preferred to be home schooled as he didn't like school. Issues raised in the discussion such as bullying and exclusions and isolation also made him feel that he didn't want to go school. If you were excluded or put in isolation – you might as well go home as sit and look at four walls. One family said that their girls had gone to primary but were now being home schooled. They were home schooling because they left primary and didn't want to go to secondary. They have to pay for a private tutor. They felt that half of school was pointless and not practical. The girls didn't need that for what they would be doing. One participant suggested that resources have been cut and that there aren't resources to support home education any more – councils have had their budgets cut.
- **Support with education at home**. One participant mentioned that their Dad did know how to read but that he could barely write and had no way of getting help. One mother referred to her child who is six bringing homework home and her (the mother) not understanding it.
- **Pandemic experiences.** Some parents and carers raised the difficulties of the pandemic and lockdowns, including difficulties with online learning and concerns about supporting children and computer equipment. One parent referred to experiences of online schooling: it was structured every hour but when you are at home with brothers and sisters this didn't work and they ended up scrapping it. Work was sent that she (the parent) couldn't understand. She felt like she was being laughed at - her children wouldn't be able to do the work because their mum wouldn't be able to help. She complained to the school and said that she would not do the home-schooling. One family referred to being given an ipad without a charger. No one reported being supported through pandemic catch up individual tutoring, including when prompted.

Quotes

Home schooling

"I'm home schooled - I prefer it home schooled I never really liked school .. it's not very free … sticking to that same schedule, its boring ... [I can] mix it up … do something different every day - get information off the internet which is … better … I feel more comfortable … if I go to school [there is a] certain curriculum … I ain't got no restrictions … I can read a book and then I can start maths … it's got a bit more freedom than sitting there in school" [child/young person]

"And that's why I don't go to school ... [I] don't mix with other people, just my family … keep myself to myself and that's that …" [child/young person]

"I can't keep fighting it everyday… I might as well do my own thing" [child/young person]

Concern with support with education at home

"My youngest who is six brings homework home and I don't understand it" [parent/carer/member of family group]

8.3.5 Relationships, trust and cultural awareness

- **Poor relationships with teachers**. While there were some references to good teachers, there were also references to poor relationships with teachers (including headteachers) and other staff at school. In some instances, participants referred to degrading treatment by teachers as well as other children. One participant referred to a headteacher not wanting to talk to Travellers. One participant referred to multiple meetings with schools that never ended well.
- **Cultural awareness, stereotyping and assumptions.**
  - Some young people and their families reported some poor experiences of measures that schools had taken to build up cultural awareness and referred to activities that weren't helpful, that disclosed their identity and made life worse for them. There were references to having had to stand up on stage in assembly by both young people and their parents and carers and that this resulted in hostile treatment from other children and young people. Some parents and carers reported similar experiences when they had been at school. Some participants also highlighted the use of stereotyping materials at school – for example, imagery that focussed on old fashioned ways of living, traditional gender roles and use of old photos rather than contemporary living arrangements and accommodation.

- There were several references to LGBTQ and other ethnicities being celebrated at school (for example, Black history month), but to a lack of focus and awareness of Gypsy, Traveller and Roma culture and history.
- As noted above (under 'school experiences'), participants referred to assumptions being made and them being singled out in the context of bullying and punitive sanctions and some parents and carers told us that important cultural events and their circumstances and ways of life were not adequately recognised by schools.
- One participant told us that when she joined school her school didn't think she could read or write - she felt that they stereotyped her and made assumptions about her and that she was judged before they knew her.
- TV programmes were also identified as stereotyping.

- **Importance of pro-active measures to build up cultural awareness and better relationships and trust between schools and families.** Some parents and carers identified good practice and positive experiences. For example, one participant reported good practice in an early years setting, where Traveller history month was celebrated in a positive way, and felt that there was much the school attended by another child could learn on how to teach young children not to victimise others. One participant thought that building up better relations and trust with parents is key to the progression of children from the Gypsy, Traveller and Roma communities to the end of year 11. He had experience of a model of schools pro-actively working with parents to build up trust and how this could work and felt that this is a key recommendation.

Quotes

Cultural awareness, stereotypes and assumptions

"They wanted us to do an assembly and stand up in front of everyone … I said it no … I'm not doing it … it's just gonna make it worse" [child/young person]

"They done that ... what they said with assembly … with my little sister and me and the other Travellers when I was at school... The whole time we was there we were bullied even more … the way they made it was like the men do what they want - of course they do - but they made it look like us women … are slaves ... they would say we were dirty because the school would put in old pictures from about 100 years ago … make us look like we're that … [but] that was 100 years ago, back in that day .. times have changed … they made it ten times worse for me being at school … that's what they did to us, they made us stand on the stage ... tell everyone about it, like the whole school …" [parent/carer/member of family group]

"You know Traveller history month … my kids' school don't do it … not one thing about it … but her playschool she's two years old and they do Traveller history month in a playschool … they could educate other little children … not to victimise other children for the way we are … if they can do it, why can't the school do it?" [parent/carer/member of family group]

8.4 Feedback on data collection and data use

During the second part of the workshop, we asked participants about whether they would feel comfortable revealing their identity in data collections undertaken by schools, colleges and other public services like hospitals. This raised lots of questions about disclosing their identity in school for fear of biased treatment, not only in education and colleges but also in healthcare, and employment and many other contexts. Some felt recording of Gypsy, Traveller and Roma identity is essential and that it is important to have this option. However, many participants raised important concerns about revealing identity because of their previous experiences of discrimination and hostile treatment, and those of their families and communities, in schools and colleges and in other contexts like hospitals and GP surgeries, at work and when making purchases. For some, there were issues of historical persecution and of personal security and safety. For these reasons, some preferred not to reveal their identity in data collections. Concerns were also raised about children and young people having their identity recorded and revealed by schools and colleges without their consent.

8.4.1 Experiences of disclosing Gypsy, Traveller and Roma identity

Participants shared their experiences of data use (good and bad) and whether they would feel comfortable revealing they are from the Gypsy, Traveller or Roma communities in their interactions with public services such as schools and hospitals.

- The discussions included many references to feelings of the needs and circumstances of Gypsy, Traveller and Roma people being ignored and not understood in the context of interactions with public services such as schools and hospitals.
- Some participants expressed the view that it is important for public services to record Gypsy, Traveller and Roma identity. They felt they had the resources, but didn't do it or didn't use this information. Information was collected for other groups but not for Gypsy, Traveller and Roma people.
- Several participants raised concerns about how exactly identity is recorded and the limited options that are often available. Sometimes this meant that participants had to tick the 'Other' option. The importance of being able to record Gypsy, Traveller and Roma identity individually and separately (rather than Gypsy, Traveller and Roma status being regarded as one single combined identity) was mentioned. Not having a separate Roma tickbox was also raised as a specific concern. One participant told us about a recent experience at a hospital of not having a separate Roma tickbox and having to record nationality instead. Another participant highlighted that people from the communities have different skin colours but that Gypsy and Traveller identity is often reported under the White tickbox. Another highlighted that the category Gypsy, Traveller and Roma is used to refer to a range of diverse backgrounds and groups such as showmen.
- Some participants told us that they did not feel that identity was relevant in the context of hospital services as all people simply require the appropriate medical treatment, while others told us that it is important for public services to understand the needs of Gypsy, Traveller and Roma people. For example, some hospital understood need for female doctors and preferences women and girls might have for female doctors. Experiences were mixed, with some hospitals being willing to do this but others not.
- Participants also referred to difficulties of collecting data, because for example of the issue of bricks and mortar versus people living on the roadside and the nomadic way of life, and people not wanting to come out as Gypsy, Traveller or Roma because of fears of discrimination.

8.4.2 Reasons why young people and their families might be reluctant to reveal their identity and backgrounds in data collections

Participants also told us that there are many different reasons why Gypsy, Traveller and Roma children and young people and their families might be reluctant to reveal their identity and background in their interactions with public services such as schools and hospitals. Many participants felt that revealing they are from a Gypsy, Traveller or Roma background – or having this discovered or disclosed by someone else - is a major concern and feared discrimination and adverse consequences.

- **School context.** When asked about whether they are comfortable revealing their identity at school or college, participants referred to the multiple adverse experiences and concerns relating to the disclosure of Gypsy, Traveller or Roma identity highlighted above - including the potential for exposure to hostile treatment, stereotyping and bullying as well as barriers to progression, discrimination and other adverse treatment. As highlighted above, adverse experiences and concerns were also specifically raised in relation to progression into apprenticeships, work experience and college.
- **Other contexts.** Many participants referred to hostile treatment and discrimination in other broader environments including in the context of health, employment, social services, taxis, delivery and insurance services.
    - In relation to health, one participant reported difficulties in registering with a GP resulting from paperwork and informational requirements, including requests for passports and addresses. Another participant referred to her variable experiences using hospital maternity services. Experiences of the birth of two children had been good, including one great midwife, however, when having one child the situation had deteriorated and the participant linked this to her identity and the way maternity staff reacted to her broader family. Differences in experiences at different hospitals relating to preferences for female health professionals were another key concern. At one hospital staff just listened and respected requests for female health professionals in a straight-forward way. However, at another hospital a request for a female health professional had resulted in an escalating altercation with the situation becoming very difficult. Others mentioned broader family members fearing that hospitals might not give them the right treatment.
    - Other participants mentioned social services being discriminatory towards Travellers as well as highlighting experiences of

discrimination in the context of delivery and insurance services. For example, if someone gets a quote for a car and their address is a caravan site the insurance quote comes back sky high.

  - The difficulties to getting employment and the need to conceal identity at work was a recurring theme and many participants referred to job opportunities being limited because of who you are. One participant highlighted that members of his broader families and communities tend to do what they know, like scrap metal, handy man work, farm work, labour, gardening, and that there are no real other options for them but that they make their own businesses in these types of occupations.

- **Historical persecution and personal security**. Some participants referred to their own fears or the fears of their families revealing their identity because of the broader historical contexts of adverse treatment, discrimination and persecution in the past. One participant mentioned the holocaust and fears that something like that could happen again so that some from the communities might feel that it would be better not to have documents saying who you are. Another referred to the use of personal data such as ethnicity and nationality on passports. There were also references to ethnicity information not being routinely collected or used for social policy purposes in some European countries. Therefore, ethnicity might be viewed as a private matter and those not born in the UK might have particular concerns for their personal security and around the misuse of ethnicity data.

Quotes

Adverse experiences of revealing identity and reasons why identity might not be revealed – non-school context

"I've had the same doctor since I was four or five … but he knows what we are and doesn't discriminate … he knows my children ... knows my mum … and my family … but if I went to the hospital I would literally say white British because I don't want them to discriminate me … because of what I am" [parent/carer/member of family group]

"I don't like male doctors … and they tried to get a male doctor in … no, I said I want a female … but they didn't respect that ... I had to scream at them … it should have been 'alright she's from the Traveller community … we should go and get a female, what she's asking' … they ignored what I was saying to the point I was angry, shouting …." [parent/carer/member of family group]

"We moved down this way about three years ago … we couldn't get a doctors … they was asking for passports, proof of address … we had just moved down here … we didn't have that … we said like Travellers site - they was like no, literally" [family group]

"Some people in my family won't go to hospital because they fear they will get discriminated against" [young person]

"I make out that I'm a gorja to keep my job ... I've done that for many years … you've got to go out and work" [young person]

"I was [about] 12 before I finally got to mark myself on papers .. going to hospitals they might not give you the right treatment … or down the line there will be an extremist government … you don't know … what will happen to us ... there's that fear of the holocaust repeating itself" [young person]

"It's a strange experience for us because in [country] … no one is collecting ethnicity data, it's considered to be private .. for protection .. sometimes we think that it's better not to reveal it for security reasons .. we also have the … [country] background … [with] passports … [but] … I think I would tick it … I could understand it can be used for statistics … it's good to have statistics … to advocate" [young person]

Theme 3: whether participants feel comfortable revealing their identity in data collections

When we asked participants whether they are comfortable revealing their identity at school and in other environments and contexts, important concerns around revealing their identity were repeatedly expressed by most participants.

- For all of the reasons above, participants said that they or that families might sometimes not disclose their Gypsy, Traveller or Roma identity on documents and forms. Some participants told us that they would or might prefer to (or sometimes prefer to) tick the White British box. In some instances this was because this might be easier and not a big deal. In multiple instances, there were references to not wanting to reveal identity, or concealing identity, because of fears of discrimination and adverse consequences. One participant suggested that whether or not they revealed their identity would depend on context and her perceptions of hostility and consequences.
- An important distinction was made between self-identification and identification by others. Some participants reported that sometimes identity is revealed or disclosed by schools without their knowledge and without checking with the families or individuals concerned, and that this could make life worse for them or have adverse consequences for them. Participants emphasised choice and that schools and others should not tick boxes in official documentation or reveal Gypsy, Traveller or Roma identity without consent.

Quotes

Perspectives on whether participants would reveal their identity in data collections by schools and other public services

"We always tick .. but then there's not an option … it says 'other' … you have to write it down…[t]hey should [ask] ... we're a culture … we're a race … you have to be born into it" [young person/parent/carer/member of family group/family member]

"At my doctor's surgery they do have Roma Gypsy plus Irish Traveller … they have it at her playschool .. they don't have it at the school … but they do ask you" [parent/carer/member of family group]

"It's not a really big deal but it just would be nice if it was on ... other races have it .. why can't we?" [child/young person]

"I don't care … if I've got a broken arm I just want it fixed, I don't care about what I am" [child/young person]

"Recently I had to go to the hospital .. on the form I had to include my personal information … there was a box - Gypsy or Traveller … but … we identify ourselves as Roma … there was no box for Roma .. we don't think we are Gypsies and for sure not Travellers … there should be additional information you can write by yourself" [young person]

"In education there's a box to tick … I didn't tick it but my school ticked it on my behalf .. which I don't think is right really .. it's someone's culture or ethnicity - they might not want to come out for certain reasons .. public services are presuming people's ethnicity –at the end of the day it's how you define yourself" [young person]

"Gypsies and Roma and Travellers are too scared to say that they are gypsies or whatever … for myself I always mark myself as English or White British …" [young person]

"I try to not to tell people" [young person]

"[There's an important] difference between bricks and mortar and the roadside – it can be hard to collect data for people [who have] a nomadic lifestyle … [and] some people don't always come out as Gypsy Travellers … so it's really hard to collect any data … because of the discrimination involved … and the nomadic lifestyle..." [young person]

"My family who live on sites they do experience a lot more discrimination because they live on sites … I have the privilege of living in a house … I can pretend … my Mum she still lives on a site and she has to get medication … she sometimes faces discrimination … and some people in my family wont go to hospital because they fear they will get discriminated against .. [because] it happened in the past or to someone else that they know" [young person]

8.5 Implications: taking forward the feedback from the direct engagement workshops

Feedback from the direct engagement workshops is being taken forward in several ways. First, the research team and Friends, Families and Travellers (FFT) have worked a community artist, a community narrator and a creative design provider to produce visualisations, an animated video and a digital cartoon from participants' feedback. These can be accessed on the project website (Testing and demonstrating the value of the growing up in England dataset (lse.ac.uk)) and have been specifically designed for onward sharing with young people from the communities. The animation and digital cartoon were also available as part of a FFT teaching resources pack prepared for Gypsy, Traveller and Roma history month in June 2023.

Figure 17: Workshop feedback: animation and digital comic

Feedback from workshop participants has also fed into and shaped our broader project findings in multiple ways. Feedback on educational experiences has shaped our understanding and interpretation of the statistical findings on the educational journeys of young people from the communities reported in section 7. Additionally, feedback on ethnicity monitoring within administrative data collections has enhanced our awareness of the non-disclosure of Gypsy, Traveller and Roma identity within administrative data collections. The workshop feedback highlights that administrative data research and the use of de-identified datasets cannot be divorced 'upstream' concerns about administrative data collection and use, or the broader inequalities that impact on children and adults from the communities. Everyday perceptions and experiences of hostile and adverse treatment and discrimination across generations impact on the disclosure of identity and participants highlighted many different reasons why people from the communities might be reluctant to reveal their identity in official data collection and data use, including in the context of public services such as schools. This feedback has been critical in informing our overall framing of the statistical findings from the project (including our assessments of GUiE inclusivity and bias).

Participants' feedback and suggestions point also towards several measures and initiatives that could help to make things better for children and young people from the Gypsy, Traveller and Roma communities and their families (summarised below). These feed directly into the overall project policy recommendations section we set out in section 9.

Education, skills and jobs

- **Importance of practical, real world and vocationally focussed subjects and skills.** Positives in the feedback from the workshops on which to build included emphasis on the importance of maths, English and science, practical and vocational subjects and life skills. In addition, there were clearly expressed aspirations for better and more appropriate work experience opportunities and ultimately for better quality job opportunities.
- **16-18 transitions.** Much more must be done to support engagement and progression to year 11 and into 16-18 education. There was emphasis on more practical and relevant options. In terms of 16-18 educational participation and training, more options at 16 including quality and practical options without GCSE maths and English passes as entry requirements were highlighted as critical (including apprenticeships where in-work training is combined with gaining qualifications in core subjects such as GCSE maths and English).

- **There need to be more targeted opportunities aimed at children and young people at risk of dropping out (including at 14+).** Participants called for funding to always be available for students aged 14+ to complete technical qualifications alongside general qualifications, including English and maths, at key stage 4, at college.
- **Work experience.** The need for better work experience opportunities in year 11 was clearly identified. Parents and carers expressed the view that it is critical that there are more job opportunities for Gypsy, Traveller and Roma children and young people.
- **Specific measures are required to address high rates of exclusions amongst Gypsy, Traveller and Roma young people.** More broadly, everyday and repeated experiences of exclusions, isolation and fines amongst workshop participants were highlighted as demoralising whilst perceptions that punitive sanctions were applied unfairly and disproportionality fed into broader negative experiences of education. Feedback from parents suggesting an association between unsupported special educational needs and exclusions are of particular concern.
- **There should be more support for those who have dropped out of formal school and / or who are receiving home education and pandemic catch up measures.** More must also be done to ensure that pandemic catch up measures reach more children and young people from the communities.
- **Addressing unmet need for support with special educational needs and disabilities amongst young people from the Gypsy, Traveller and Roma communities** There needs to be much more appropriate support for children and young people from the communities with special educational needs and disabilities. There was clear recognition of the critical role of some SENCOs and learning assistants in the context of neuro-disabilities, autism, ADHD and other educational special needs and disabilities, including a reference to a supportive teaching assistant.
- **Specific initiatives to proactively address ethnicity related bullying and build up better relations and trust between schools, colleges and young people and parents and families from the Gypsy, Traveller and Roma communities are urgently required.** The importance of addressing and responding to bullying pro-actively before situations escalate was highlighted repeatedly. Existing good practice models for building up trust between schools and families from the communities were also highlighted. Suggestions included having teachers with specific responsibilities for Gypsy, Traveller and Roma children and young people, cultural

awareness training and specific initiatives to build cultural awareness into the curriculum and talk about Gypsy, Traveller and Roma identity, culture, experiences and history in the classroom. However, given mutigenerational concerns and experiences, children and young people and their families should always be consulted on initiatives that are taken forward that involve revealing their identity.

Additionally, we believe that the concerns about administrative data collection and use raised at the direct engagement workshops have implications for ADR UK's broader work on public engagement and the public good. Our broader project recommendations set out in section 9 highlight the importance of further engagement exercises with children and adults from the Gypsy, Traveller and Roma communities and other 'seldom heard' groups designed to listen to and these concerns and to build up trust as part of the broader administrative data research agenda going forward. Some participants at the workshops raised specific concerns about choice and children and young people from Gypsy, Traveller and Roma backgrounds having their identity presumed or disclosed without their consent. We have followed up these specific concerns with the Department for Education and further information is again provided in the project conclusions (section 9).

Finally, feedback from the engagement workshops has fed into our research plans going forward. Following on from the workshops for example, we have identified research examining unmet needs for SEND support amongst young people from the communities, and whether Gypsy, Traveller and Roma young people receive less support than their peers, as a key research priority. Again, further details of our research plans going forward are provided in the concluding section of this report (section 9).

9 Conclusions: summary of key findings, recommendations and policy implications

This concluding section draws together the key findings and recommendations from this report and considers the policy implications of our analysis. We begin with a summary of key findings. We then highlight our general recommendations on the further development of the GUiE dataset and on data inclusivity going forward. Next, we address the policy implications of our analysis. Finally, we set out our future research plans for extending our research on the lives, circumstances and needs of young people from the Gypsy, Traveller and Roma communities.

9.1 Summary of key findings

Data exclusion is the phenomenon whereby smaller or highly disadvantaged groups (sometimes referred to as hard to reach, seldom heard or vulnerable groups) have in the past been typically missing from, or invisible within, national data infrastructure and evidence. As the COVID-19 pandemic exposed, tackling data exclusion and building up more inclusive data infrastructure and evidence is critically important for effective public policy. Post-pandemic, the importance of data *inclusivity* - and of building up national data infrastructure and evidence that represents all people and groups in the UK - has moved up the national statistics agenda.

Our project highlights the key contribution that large-scale administrative data-population census linkages can play as part of these broader efforts. The project focusses on Gypsy, Traveller and Roma children and young people as a case study. The Gypsy, Traveller and Roma communities continue to be impacted by systematic gaps in data, evidence and understanding and parliamentary committees have highlighted that the communities are poorly served by multiple public services. Developing more inclusive data infrastructure and evidence that addresses the lives, circumstances and needs of children and young people from the communities is critical for Government policies in key areas such as education, skills and employment, and health.

The study assesses the research potential of the Growing Up in England (GUiE) dataset. GUiE is one of a new generation of large-scale administrative data population census linkages and links information from administrative education data systems to information from the 2011 Census. Our focus is on Wave 1 of the GUiE linkage, which includes the records of more than 7 million children and young people.

9.1.1 Key findings from the quantitative research exercise

We highlight five key findings from the quantitative analysis. First, the study establishes that the Wave 1 GUiE supports the identification of much larger analysis samples than has previously been possible in research on Gypsy, Traveller and Roma young people. This is partly a matter of scale: as one of a new generation of large-scale administrative data linkages, GUiE Wave 1 includes the records of more than 7 million children and young people. Additionally, the analysis establishes that the identification of the records of children and young people from the Gypsy, Traveller and Roma communities can be maximised by pooling the ethnicity information that is available from the 2011 Census and different parts of the education system. While the 2011 Census records Gypsy and Traveller status under the White ethnic group, Roma identity was not included as a pre-coded option. The pupil data within the education side of the GUiE Wave 1 dataset records ethnicity using Traveller of Irish Heritage and Gypsy/ Roma identifiers, while supplementary information on ethnicity is available through further education and work-based learning administrative data systems. Moreover, the longitudinal nature of the GUiE Wave 1 education data opens up a much longer time window for observing Gypsy, Traveller or Roma identity. We demonstrate that by combining and pooling ethnicity information from the different component parts of GUiE Wave 1, it is possible to further boost analysis sample size substantially.

Second, applying these methods, our analysis uses Wave 1 GUiE to identify the records of Gypsy, Traveller and Roma children and young people and to follow their educational journeys over four academic years. The three academic year cohorts we follow were carefully selected to maximise the use of the data made available in the GUiE Wave 1 linkage. Specifically, we follow the progress of children and young people who turned 15 during academic years 2008/9, 2009/10 and 2010/11. We identify children and young people who are recorded as Gypsy, Traveller and Roma within these three academic year cohorts and follow their educational journeys up to the academic year they turned 19 (academic years 2012/13, 2013/14 and 2014/15 respectively). We produce new experimental statistics reporting on 14-16 and 16-18 educational pathways and attainment of Gypsy, Traveller and Roma children and young people within the cohorts, examining their progression through further education and vocational and work-based learning as well as in schools.

Looking at 14-16 educational journeys, the analysis identifies that while a majority of Gypsy, Traveller and Roma young people were following GCSE (or IGCSE) pathways in the academic year they turned 16, young people

from the communities were substantially more likely than other young people to be participating in below Level 2 qualifications; in VRQ 2, NVQ 2 or other Level 2 qualifications; or participating in 'out of scope' qualifications. Only 60% of the Gypsy, Traveller and Roma young people in the cohorts had attained Level 1 by age 17, compared to 92% of other children in the same cohorts. Additionally, 38% of Gypsy, Traveller and Roma young people in the cohorts had attained Level 2 by age 17, compared to 78% of other young people.

Of those transitioning into 16+ education, a third of the Gypsy, Traveller and Roma young people within the cohorts were recorded as working towards qualifications which were below Level 2 or 'entry to employment' compared to 6% of other young people. Gypsy, Traveller and Roma young people were also substantially more likely to be working towards Level 2 qualifications (VRQ 2, NVQ 2 or other Level 2 qualifications) or GCSE qualifications as their main pathway and less likely to be on a A/S, A level or AVCE pathway (14%, compared to 50% for other young people). Additionally, they were slightly *more* likely to be on an apprenticeship pathway – but *less* likely to be on an *advanced* apprenticeship pathway. Looking at educational attainment by age 19, only 16% of the Gypsy, Traveller and Roma young people within the cohorts had attained Level 3 qualifications by any route, compared to 60% of their peers. Additionally, only 48% had attained Level 2 qualifications (compared to 87% of their peers) and only 65% Level 1 qualifications (compared to 94%). Nevertheless, the proportion attaining Level 1 and Level 2 qualifications via any route by age 19 was considerably higher than by age 17, demonstrating the importance of 16+ catch-up pathways at both Level 1 and Level 2.

Third, the quantitative analysis illustrates the benefits of linking administrative information about educational outcomes and progression to the rich information about individual characteristics and family and household circumstances recorded in population censuses. Applying a bespoke multidimensional child disadvantage framework developed in previous research (Vizard et al 2018, Burchardt et al 2018), we highlight the scale of the challenges facing children and young people as they go through adolescence and make transitions into young adulthood. We identify that at ages 14 or 15, the young people identified as Gypsy, Traveller and Roma within Wave 1 GUiE are substantially more likely to be living in a single parent family or to experience ill-health or a disability themselves than their peers. As many as 51% of young people identified as Gypsy, Traveller or Roma were resident in households that are assessed as educationally deprived at ages 14 or 15, compared to 18% of other

young people. Additionally, 36% were resident in households that were assessed as employment deprived compared to 16.5% of their peers; 47.7% were resident in households assessed as health and disability deprived compared to 26.7% of other young people; and 38% were resident in households that were assessed as housing deprived, compared to 14.2% of other similar young people. Looking at deprivation across multiple areas of life simultaneously, 19.8% of Gypsy, Traveller and Roma young people were deprived in three dimensions, compared to 5.1% of their peers; while young people from the communities were more than seven times as likely to be deprived in all four of these dimensions.

Fourth, the research exercise puts the spotlight on the strength of the associations between Gypsy, Traveller and Roma experiences of child multidimensional disadvantage (observed at ages 14 and 15) and their subsequent highest educational attainment via any route by age 19. The results show that educational attainment by age 19 is substantially differentiated by experiences of childhood multidimensional disadvantage at ages 14 and 15. Additionally, as educational attainment at age 19 is substantially lower for Gypsy, Traveller and Roma young people than for their peers, the lowest rates of attainment are observed at the intersection of Gypsy, Traveller and Roma identity and experiences of multidimensional child disadvantage. Moreover, the risks of low attainment are cumulative and increase even further where deprivation is recorded across multiple life domains. In total, 43% of Gypsy, Traveller and Roma young people with 3 or 4 household deprivation factors at age 14 or 15 did not go on to attain Level 1 qualifications by age 19; 62% did not go on to secure Level 2 qualifications; and 90% did not go on to secure Level 3 qualifications.

In policy terms, these findings underline the importance of tailored support packages for Gypsy, Traveller and Roma young people across all educational pathways and stages of education and training. The new evidence we produce on inter-generational cycles of disadvantage is particularly striking and puts the spotlight on the particular importance of whole family support models - particularly at earlier ages. Our analysis identifies that 89% of Gypsy, Traveller and Roma young people who experienced household educational deprivation as children aged 14 or 15 had not secured Level 3 qualifications four years later by age 19. In addition, 59% failed to go on to obtain a Level 2 qualification and 42% failed to go on to obtain a Level 1 qualification by age 19. The detailed policy recommendations we draw from the analysis are further elaborated below.

While highlighting the 'value added' of GUiE and its research potential, the quantitative analysis also highlights some limitations of the GUiE

linkage. While the Wave 1 GUiE dataset represents an important break-through in supporting the identification and analysis of the records of children and young people from the Gypsy, Traveller and Roma communities at scale, we also highlight several caveats and limitations, including the imperfections of the recording of Gypsy, Traveller and Roma identity within the 2011 Census and administrative education systems and other coverage restrictions driven by Wave 1 GUiE content and design. Additionally, we provide evidence of disproportionality in Gypsy, Traveller and Roma record matching success rates and of deprivation-related compositional effects. These factors restrict the extent to which the dataset is representative of the underlying population of Gypsy, Traveller and Roma children and young people and the generalisations and inferences that should be drawn from our statistical findings. Additionally, a change of provision in relation to the 2011 Census data from Proof of Concept to GUiE Wave 1 restricts the research potential of the dataset and we recommend that this is urgently addressed (on which, see the recommendations below).

9.1.2 Key findings from the direct engagement excise

Participants were invited to share their educational experiences and journeys. We learnt that many participants – children and young people as well as parents and carers - felt that they have worse educational and training experiences because they are from the Gypsy, Traveller or Roma communities. This included:

- Lack of respect, hostile treatment and stereotyping;
- Bullying and the inability to precipitate a response to complaints;
- Inadequate support for special educational needs and disabilities
- Repeated day to day experiences of punitive sanctions including exclusions, isolation and fines.
- Multigenerational adverse experiences of education - in some instances, similar adverse experiences were reported across multiple generations of the same families, with parents and carers recalling their own experiences of hostility, bullying, segregation and punishment from their own childhood.

Particular concerns were expressed about secondary education not meeting the needs of children and young people from the communities. There was a particular emphasis on limited subject relevance, with participants highlighting the importance of practical options focused on the

real world. Poor work experience opportunities in year 10 were highlighted as a specific concern and some participants reported their opportunities had been adversely affected by their identity. The importance of alternative pathways at age 14 was raised in discussion. Additionally, formal subject requirements (including GCSE passes in English and maths) were identified as a particular barrier to 16-18 educational participation by some parents. There were several reports of a lack of support for children and young people living with neurodiverse conditions, autism, ADHD, dyslexia and other special educational needs and disabilities. Some families felt that this lack of support and intervention was linked to their Gypsy, Traveller or Roma identity. Some parents worried about not being able to support their children adequately. Some parents and children had particularly struggled during lockdown. Some children and young people were being home educated - particularly at the secondary stage – and in some cases this was seen as preferable because of these problems and experiences. Lack of support with home education was mentioned. There were no mentions of access to pandemic catch up tutoring, even when prompted.

On data use, the discussions included many references to the needs and circumstances of Gypsy, Traveller and Roma people being ignored and not understood in the context of interactions with public services. Some participants felt that it is essential for public services to record Gypsy, Traveller and Roma identities and not having a separate Roma tickbox was raised as a specific concern. However, participants also told us that there are many different reasons why Gypsy, Traveller and Roma children and young people and their families might be reluctant to reveal their identity and background in their interactions with public services such as schools and hospitals.

- Many participants felt that revealing they are from a Gypsy, Traveller or Roma background – or having this discovered or disclosed by someone else – could result in discrimination and hostile treatment or have other adverse consequences for them, making their lives worse rather than better.
- Participants gave many examples of experiences of discriminatory and worse treatment. This included their experiences of the education system, including work experience and training opportunities at college, and broader experiences including registration with GP surgeries, hospital maternity services and use of private delivery and insurance companies. Some participants also referred to personal safety, broader surveillance and historical persecution of the communities.

- For these reasons, some participants said they might not reveal their identities in data collections. Specific concerns were expressed about children and young people having their identity assumed by schools or teachers or having their identity disclosed to others without their consent including in the context of recording ethnicity on education system administrative records.

9.2 Recommendations on dataset development

The team were the first external researchers to access the Wave 1 GUiE data in June 2021. At this point in time, the GUiE Linkage Methodology publication (ONS 2022 – which includes a detailed description of the dataset design) and GUiE metadata were not yet available. Our original use and exploration of the dataset during summer 2021 identified several issues relating to dataset design, coverage and quality that were followed up within the life-time of the project. To address these issues, the CASE team worked extensively with ONS through a system of meetings and feedback as well as through the newly created ONS GUiE technical and steering groups, broader ADR England Data Linkage Steering Group arrangements and with other stakeholders. Many of the issues we have raised have been subsequently clarified and addressed through the publication of the GUiE metadata and User Guide (ONS n.d.b); the publication of the GUiE methodology documentation and findings in Autumn 2022 (ONScdef); the release of Wave 2 GUiE; the release of the AEDE spine duplicates dataset; and the release of elements of Wave 3 GUiE (with the release of additional elements of Wave 3 being imminent at the time of writing). At the time of writing, options for the forward development of GUiE are under active consideration. Below, we set out some general recommendations for future dataset development as well as our recommendations on data inclusivity.

9.2.1 Recommendations on the further development of GUiE

Keep GUiE up to date as a policy relevant resource for the 2020s

Up to date information is critical for policy purposes, particularly in the post-pandemic world. Going forward, it will be important to the time-windows of the GUiE Waves 1-3 data. The back-series of vulnerability data derived from the English School Census already released via Wave 3 GUiE, and the imminent release of a back-series of CIN data and CLA data, are welcome additions of critical importance for understanding the lifepaths of vulnerable children. Looking forward, as well as incorporating data for more recent academic years (beyond the current 2014/15 cut off), it will be

desirable to incorporate a greater range of variables for understanding children's interactions with children's social care services, including derived variables such as age of first interaction with children's social care services and accumulated time spent interacting with different children's services. Additionally, to supplement the 2011 Census data included in Wave 1 and to ensure continued policy relevance going forward, we recommend that the new data resulting from the 2021 population census is matched into the GUiE linkage.

Substantially expand GUiE age coverage

The availability of education attainment data (highest level of qualification) is limited by the feasibility AEDE age restriction (with feasibility AEDE limited to children and young people aged 14-29 in 2014/15) as well as the age restriction that applies to the NPD education attainment file in GUiE (age >= age 14 and < 20). These age restrictions effectively create a disconnect between the conceptualisation of Wave 1 GUiE (as a resource focussing on children) and the key data that is available through GUiE. Moreover, the GUiE age restrictions in turn restrict the research potential of the Wave 2 and Wave 3 data (as linked 2011 Census and educational attainment data will only be available for those satisfying Wave 1 eligibility conditions). We recommend that any future rebuild substantially expands GUiE Wave 1 age coverage to include detailed information about younger children.

Substantially expand GUiE education coverage

The attainment data in Wave 1 is effectively a subset of the Department for Education records relating to the underlying Department for Education Young Persons Matched Administrative Dataset (YPMAD). There are many strengths of the YPMAD attainment data (Level 1, 2 and 3) including the fact that YPMAD is a linked dataset which looks across further education vocational qualifications and work-based learning, rather than being restricted to GCSE and A level attainment. The derived nature of the level 1,2 and 3 variable information made available through GUiE is an important strength and this variable should be retained in any future GUiE releases or rebuild.

However, GUiE as currently designed does not provide researcher access to much of the data currently available through the National Pupil database, including key stage 1, key stage 2, key stage 3 or key stage 4 data which are standard inputs into most educational research. Early years information held by the Department for Education is also not currently

included. It will be critical to expand the education coverage of future GUiE releases to address this.

Include a school pseudo-identifier and additional variables on school type and school moves

GUiE Wave 1 does not currently include a school identifier of any form (that is, there is no school pseudo-identifier available in the dataset). This means that it is not possible to analyse children's outcomes and progression in the context of their schools, or to take account of or control for school type or clustering effects, as is often standard in standard education research. This makes GUiE Wave 1 analysis less useful for policy purposes and for organisations such as OFSTED, as the analysis of school level effects and the 'value added' of individual schools are critical issues for understanding the school system. Additionally, the inclusion of NPD variables on school type, school moves and school exit would support analysis of school stability and drop-out from the formal education system using GUiE.

Address the change of provision relating to non-data subject (individual level) records from the 2011 Census

As noted above, Wave 1 GUiE only includes 2011 Census records for data subjects whose records were successfully matched to the bespoke extract of the feasibility AEDE. While some household level (and household reference person) variables are also available, individual level records for other household members, such as grandparents, parents and siblings who are not themselves data subjects, are not included in Wave 1 GUiE. This exclusion reflects a change in provision between the GUiE Proof of Concept dataset (which included the 2011 Census records of *all* individuals in the household of a data subject, whether or not they themselves are a data subject with a successfully matched feasibility AEDE record) and GUiE Wave 1. The exclusion of the individual level records of these 'other' (non-data subject) persons recorded in the 2011 Census restricts the current value of the GUiE Wave 1 linkage substantially (and has follow on implications in terms of restricting the research potential of Waves 2 and 3). For example, the value of examining parental information at the individual level (including characteristics such as mother's country of birth, or highest educational qualification at the individual level), was established in the PoC GUiE publication (ONS 2020). This a major limitation of Wave 1 GUiE and should be remedied at the first opportunity.

Continue to deepen and communicate understanding of GUiE linkage, coverage and quality

There remain some information gaps around GUiE linkage, content, quality and design, including more in-depth information relating to the feasibility AEDE linkage and pre-processing and the reasons underlying duplicate records in the AEDE spine. Overall linkage success rates (including overall linkage success rate information seen from the point of view of both the feasibility AEDE and Census residuals) were not included in the wave of documentation published by ONS in Autumn 2022, but are fundamental and basic to understanding dataset quality and communicating essential information on coverage, interpretation and possible sources of bias to researchers. We recommend that these are published, with residuals and matching success rates broken down by academic year. Researcher access to residuals is also increasingly viewed as good practice in administrative data research, so that researchers can identify and explore disproportionality and frame their research findings appropriately.

9.2.2 Recommendations on dataset inclusivity

Improve ethnicity recording in administrative education data systems

The analysis in section 6 identified important limitations of Gypsy, Traveller and Roma ethnicity recording in both the 2011 Census and administrative education data systems. The 2011 Census did not include a Roma tick-box. Additionally, in relation to Gypsy, Traveller and Roma recording within administrative education systems, ethnicity recording practices are different in National Pupil Database (NPD) and Individual Learner Records (ILR) datasets. The NPD uses the ethnicity classifications "Gypsy/Roma" and "Travellers of Irish Heritage" – which means there is at least some basis for identifying the Roma population. However, these pupil data ethnicity classifications are inconsistent with the 2011 Census ethnicity identifier and inappropriately aggregate Gypsy and Roma categories. Meanwhile, the ILR dataset, like the 2011 Census, does not include a separate Roma identifier.

In the current project, we have maximised the potential of the GUiE Wave 1 linkage to identify children and young people from the Gypsy, Traveller and Roma communities by pooling ethnicity information across the linked education and 2011 Census data. However, in the future, it is critical that underlying issues of good practice in ethnicity classification and data harmonisation are fully addressed, building on the high-level

recommendations set out in the Race Disparities Audit and the ONS Data Inclusivity Task Force, as well the good practice recommendations set out by civil society representative organisations such as Friends, Families and Travellers and the Traveller Movement. Going forward, more fine-grained and meaningful ethnicity classifications including separate recording of Gypsy and Irish Traveller identities, provision for recording of other Traveller status, and the inclusion of separate Roma identifiers, are all recommended good practice. At a minimum, there should be harmonisation to the 2021 Census ethnicity 18+1 standard, which includes a separate Roma coded tick-box.

Link GUiE to the 2021 Census as a means of taking harmonised and best practice ethnicity identification forward

The 2021 Census included for the first time a coded Roma tick-box option in addition to the Gypsy/Irish Traveller option first introduced in the 2011 Census. Looking forward, and in the light of the recommendations above, we highlight the importance of linking GUiE to the 2021 Census from the point of view of Gypsy, Traveller and Roma research, data inclusivity and national statistics.

Maximise the representation of 'seldom heard' groups within administrative data linkages by improving dataset design and developing bespoke methodologies that maximise the inclusivity of data linkage

In the future, in order to maximise the representation of 'seldom heard' groups within GUiE and other administrative data linkages, issues of inclusivity should be explicitly addressed when new administrative data linkages are proposed and designed. To ensure that equality and human rights concerns are fully captured and reflected in administrative dataset development, options for addressing missingness and invisibility should be considered by default at the planning stage.

The residuals exercise undertaken for our project provides evidence of disproportionality in Gypsy, Traveller and Roma record linkage rates. Building on the recommendations of the Inclusive Data Taskforce and UKSA and OSR good practice guidance relating to the post-Covid statistics agenda and the future of administrative data linkage, we recommend that disproportionality in record match success rates for disadvantaged and at-risk groups in the GUiE linkage, including but not limited to young people from the Gypsy, Traveller and Roma communities, is addressed through the development and implementation of new and bespoke inclusive data linkage methodologies. To move forward here, we recommend that a

clerical review examining how linkage rates for Gypsy, Traveller and Roma young people and other 'seldom heard groups' in the GUiE dataset can be maximised is taken forward by ONS, feeding into any broader initiatives around inclusive data linkage within ONS, UKSA and OSR working groups and teams.

Introduce a direct link between the CIN / CLA data and the 2011 Census and consider the incorporation of a future home education register into the GUiE build

To maximise GUiE research potential in relation to vulnerable children, we recommend that as part of GUiE forward development a direct link between the 2011 Census and children's social care records is introduced. This design would GUiE maximise research potential in relation to children and young people who are in contact with children's social care services (and are included in the CIN/LCA) but who do not have a school record (or another record in the feasibility AEDE).

Additionally, the development of a home-schooling register was recommended by the House of Commons Education Committee in the context of its recent inquiry into gaps in Gypsy, Traveller and Roma educational attainment. We recommend that a register of this type (or a broader register to cover children who are not in school) is considered as part of GUiE forward design. As the importance of missingness and persistent absence from school have moved up the policy agendas of organisations such as the Children's Commissioners Office (including in relation to safeguarding), this design would optimise the role of GUiE as a forward-looking resource that can address pressing policy concerns in the 2020s. Incorporating a future home-schooling register into future GUiE design would have particular value added in the context of Gypsy, Traveller and Roma research, as dropping out from school is a key concern, and a substantial proportion of children from the communities report opting for home education.

Undertake targeted engagement exercises to address concerns and maximise participation in official data collection and representation in administrative data records

We finally highlight the importance of engagement exercises to build up trust in administrative data research, address the concerns of particular groups and maximise participation in administrative data collection. The direct engagement exercise undertaken as part of this research has highlighted important concerns around the disclosure of Gypsy, Traveller and Roma identity in the context of official data collections, and as the Women and Equalities Committee have highlighted, specific initiatives

working with civil society organization are necessary. Initiatives that have already been undertaken as part of the 2011 Census and the 2021 Census are important developments and our research has highlighted the particular issues that are raised in the context of the recording of Gypsy, Traveller and Roma identity in schools.

Our engagement workshops raised specific concerns about Gypsy, Traveller and Roma identity being recorded in education administrative data systems without choice and informed consent. We have followed this issue up with Department for Education who have confirmed that while the ascription of ethnicity by schools and teachers has occurred in the past, guidance is now clear that ethnicity should be self-reported within administrative education data collections (Connell 2022b). The feedback from our engagement workshops suggests pro-active steps are required to communicate this change in approach to young people from the communities and their families, in order to build up trust and promote the disclosure of Gypsy, Traveller and Roma identity.

9.3 Policy implications: the importance of tailored support packages for young people from the Gypsy, Traveller and Roma communities and their families

Overall, the research highlights the urgent need for additional targeted and focussed efforts to provide children and young people from the Gypsy, Traveller and Roma communities and their families the support they need to secure Level 3 attainment and to break inter-generational cycles and disadvantage. Looking forward, we argue that the public policy response that is required must be life-long, multi-dimensional and multi-generational. A *lifelong* approach is required because tailored policies and support is required from early years via the new network of Family Hubs onwards, throughout the primary and secondary stages as well as in relation to post-16 education, training and work-based learning, higher education and opportunities for life-long learning and up-skilling. A *multidimensional* approach is required, as the educational disadvantages experienced by Gypsy, Traveller and Roma young people do not arise in isolation but are part of a complex overall picture of social inequalities across multiple life domains that interact and reinforce one another. A *multi-generational* approach is required because as both the quantitative and qualitative findings in this report highlight, educational disadvantage is often transmitted from one generation to another, and whole family support packages are often required.

Following on from the report of the Women and Equalities Committee (2019), which highlighted a persistent failure by national and local policy

makers to tackle long standing inequalities facing Gypsy, Roma and Traveller communities in any sustained way, the Ministry of Housing, Communities and Local Government (now the Ministry of Levelling Up, Housing and Communities), launched a new national strategy to tackle Gypsy, Roma and Traveller Inequalities (MHCLG 2019). The launch document stated that the strategy would be led by MHCLG working with several government departments and the Cabinet Office Race Disparity Unit to improve Gypsy, Traveller and Roma outcomes across health, education and employment. The announcement further stated that in order to address the serious disparities identified in the Race Disparity Audit, the Ministry of Housing, Communities and Local Government had provided £200,000 of funding between 6 projects aiming to improve outcomes for Gypsy, Roma and Traveller communities in the areas of educational attainment, health and social integration. Additionally, the announcement included funding 22 projects that support Roma communities across England through the Controlling Migration Fund and funding two projects to improve the reporting of hate crime by Gypsy, Roma and Traveller communities. However, Friends, Families and Travellers (2022b: 5) noted: "to date no visible progress has been made on the Strategy, with no published Plan or Strategy outlining any aims, actions, lines of accountability or timelines".

Below, we identify some of the ways in which tailored support packages could be embedded (or further embedded) into a range of different government policies.

9.3.1 The Government's post-Brexit skills agenda and Levelling Up

The Government's post-Brexit skills agenda and the Levelling Up White Paper put central emphasis on Level 3 educational attainment as a minimum basis for secure employment and income in the future. Following on from the independent review of post-18 education and funding (the Augar Review) in 2019, the Government's plans for the transformation of the skills, training and post-16 education landscape were set out in a White Paper in 2021 and the Skills and Post-16 Education Act (2022). This included plans for the reform of further education and vocational training, including apprenticeships, traineeships, promotion of English and maths and digital skills, targeted courses and training for adults without an existing full Level 3 (A Level equivalent) qualification, clearer progression routes for students, improved careers advice, reformed higher technical education, and provisions for a life-time skills guarantee, with emphasis on a life-long four year post-18 education entitlement that can be accessed to fund further and vocational education, similar to that available in the higher

education sector (Department for Education 2021). Other related developments include the launch and roll out of the new Technical Level (T level) programme from September 2020, providing new two-year technical courses which are taken after GCSEs and equivalent to three A levels. These are intended to provide new technically orientated routes and pathways for the required training for skilled occupations, combining practical and knowledge-based learning options at a school or college and on-the-job experience / apprenticeships.

The quantitative findings in this report highlight the substantial proportion of post-16 Gypsy, Traveller and Roma young people learning at Level 2. Transforming the post-16 qualifications landscape at Level 2 and below has received less attention in the Government's post-Brexit skills agenda which has tended to focus on Level 3 reforms. Nevertheless, the importance of qualifications at Level 2 and their role as a basis for preparation and progression to Level 3, social mobility, inclusion and life-long learning was recognised in Department for Education (2022, 2023). The emerging agenda for the reform of post-16 qualifications at Level 2 and below includes: addressing quality and standards; the need for simplified and clear progression pathways; technical qualifications that provide a basis for employment at Level 2; in-work reskilling and retraining; work-based internships, traineeships and apprenticeships; and support for SEND.

The feedback at the engagement workshops organised for this project had a high degree of alignment with this emphasis on skills and vocational options. The young people and their families at the engagement workshops expressed high aspirations in terms of education and employment, placing particular value on life-skills and work-focussed vocational and practical options and similar findings were reported in ONS (2022ab). However, key concerns expressed at the workshops included limited year 10 work experience opportunities and barriers to progression at the end of year 11, such as formal requirements in relation to English and Maths GCSEs. As noted in section 3, civil society representative organisations have identified the need for clearer tailored pathways and support to ensure continued progression through secondary education and 16+ education and training. Tailored re-engagement and catch-up pathways should be available for all young people who have dropped out of education or who are at risk of doing so. Re-engagement pathways should include improved alternative post-14 pathways where appropriate. Opportunities for intermediate as well as higher level apprenticeships are also critically important for young people with lower educational qualifications and / or who are at risk of educational

exclusion or who have been home educated but choose to re-engage with the formal education sector at age 16 (for references, see section 3).

9.3.2 Post-pandemic educational catch-up

Government proposals for pandemic catch-up put central emphasis on individual tutoring schemes, with a £1.4 billion programme of tuition announced as part of a COVID catch up plan in June 2021. The package was substantially less than that recommended by the education recovery Tsar Kevan Collins, who was appointed to lead education pandemic recovery efforts with a focus on pupils from deprived backgrounds. His recommendations included a £15b package including a broader package of measures including a systemic shift towards extended school days.

In July 2021, as part of the broader plans for a 'tutoring revolution', the Government announced a £1m pilot scheme focussing on boosting educational attainment, tackling exclusion and drop-out rates and improving pathways to employment for Gypsy, Roma and Traveller children. This was intended to reach 150 schools in Bradford, Central Bedfordshire, Essex, Hillingdon and Surrey with a focus on a range of targeted work, including: reducing drop-out rates by engaging families with pre-school children and those transitioning from primary to secondary school; targeting young people not in education, employment or training; fostering better cooperation between councils, schools and families and tutoring for catch-up lessons to support pupils whose education has been disrupted by the pandemic (Ministry of Housing, Communities and Local Government and Department for Education 2021).

However, pandemic catch up measures need to reach more children and young people from the Gypsy, Traveller and Roma communities. Tailored re-engagement and catch-up pathways should be available for all young people who have dropped out of education or who are at risk of doing so. At our engagement workshops, some participants told us that they had received support with access to some digital equipment during the pandemic. However, there were no mentions of benefiting from post-COVID-19 catch up interventions or tutoring, even when prompted.

9.3.3 Post-pandemic concern with persistent absence and home education

In March 2023, the Children's Commissioner for England raised further concerns about children falling 'off radar' and disappearing from mainstream schools in the wake of the pandemic. The latest annual data

indicated a substantial increase in severe and persistent absences with follow up analysis identifying associations with special educational needs, severe needs, mental health and anxiety. Disadvantaged children and young people (children on free school meals) and certain groups of children and young people (especially Gypsy, Traveller and Roma children and young people) are also particularly affected (De Souza 2023).

The importance of universal measures to monitor children in home education from any community - including at a very basic level creating a register of children not in school - moved up the education policy agenda prior to the pandemic but has received greater recognition in the context of increasing concerns around persistence absences and 'missingness' from school in the wake of COVID-19. A survey by the Association of Directors of Children's Services (ACDS) indicated that an additional 115,542 children and young people were known to be electively home educated at some point during the 2020-21 academic year, a 34% increase on 2019-20 (Association of Directors of Children's Services 2021).

Specific concerns in relation to OFSTED not inspecting home environments and lack of powers of local authorities in relation to accountability for children receiving an adequate education were raised in Women and Equalities Committee (2019). Concerns arise in relation to home education where parents may not have sufficient literacy to support children and no or insufficient external support/oversight is available. This concern overlaps with the issue of parents not having sufficient literacy to support children in school, with an overarching need for more support for parents, whether children are in school or home educated.

9.3.4 Digital exclusion

Covid-19 also put the spotlight on the issue of digital exclusion. For children and young people from the communities to be adequately supported in relation to Level 2 and Level 3 progression, support with all aspects of home learning, including in relation to the digital requirements of home learning within the formal education system as well as the additional requirements associated with elective home education, will need to be addressed. This is one area where broader multidimensional policies and interventions beyond the education sector will be necessary. For example, in relation to government housing and planning policies, in March 2022, the Department for Levelling Up, Housing and Communities announced £10 million of capital funding for 2022/23 to support local authorities in building new transit and permanent traveller sites, refurbishing existing permanent traveller and transit sites and providing

temporary stopping places and facilities for travellers (DLUHC 2022). Onsite provision for the 2020s should address unmet needs associated with digital exclusion and the critical importance of more adequate homework and learning environments for children, adolescents and young people. For example, access to the internet, broadband and to devices such as computers and printers are now essential needs and should be addressed locally through the provision and upgrade of sites.

9.3.5 Follow up to the Timpson Review

The Timpson Review raised concerns that exclusions were being used disproportionality to sanction socially disadvantaged children, vulnerable children, children from some ethnic minority groups and children who experience SEND. Feedback from our engagement workshops aligns with the concerns raised in the Timson Review. As we noted in section 8, this included experiences of repeated exclusions and time spent in isolation rooms and raises additional concerns about the use of exclusions in the context of unmet need for support with SEND as well as perceptions that punitive sanctions were being applied unreasonably and disproportionately. The Timpson Review set out key recommendations relating to school accountability for exclusions and off-rolling practices, highlighting the critical importance of systems that instil confidence that every exclusion is lawful, reasonable and fair and of ensuring that additional support is provided where it is needed (Timpson Review of School Exclusion, 2018). The project findings underline the particular importance of building trust with Gypsy, Traveller and Roma young people and their families and ensuring that they are adequately supported.

9.3.6 The new Government SEND strategy

The Government's new Special Educational Needs and Disabilities (SEND) and Alternative Provision (AP) Improvement Plan was set out in early 2023, outlining a strategy for a new nationally consistent SEND and alternative provision system underpinned by evidenced-based national standards. This had the two-fold aims of early identification of needs and intervention, and clear expectations for the types of support within mainstream settings; coupled with tailored provision for children and young people who require an Education and Health Care Plan and specialist provision (HM Government 2023). However, there is no mention in the plan of the need for tailored interventions or particular support for Gypsy, Traveller or Roma children and young people. The importance of addressing

unmet need for support with special educational needs and disabilities amongst young Gypsy, Traveller and Roma young people was a key theme in the feedback from the direct engagement exercise we reported in section 8. We recommend follow on research to assess the extent of unmet need for support with special educational needs and disabilities amongst young people from the communities and the development of tailored interventions and support packages as part of the new SEND strategy.

9.3.7 Early years and the new network of Family Hubs

A new Family Hubs and Start for Life programme was established in early 2023, with a focus on the first 1000 days of life and the goals of joining up and enhancing services and ensuring that parents and carers can access the support they need and that all babies have the 'best start' in life. Guidelines on Family Hub access and inclusion *do* specifically highlight that some groups may experience racism and discrimination when they access support services, and refer to Gypsy, Traveller and Roma parents as one of the groups who may be disproportionately affected by barriers to accessing support and under-represented in service provision. Collection of accurate community data on service coverage and engagement is additionally identified as a crucial step in addressing inequalities and ensuring that support services reach all those in need (National Centre for Family Hubs n.d.).

Our research puts a particular emphasis on the intergenerational transmission of educational disadvantage amongst Gypsy, Traveller and Roma families and the importance of whole family approaches to the provision of support. These findings underline the critical importance of ensuring that these commitments to inclusion and engagement with the Gypsy, Traveller and Roma communities are implemented in practice, including by forging links with outreach and civil society mechanisms where appropriate, and by incorporating support with life-long learning into the Family Hub model.

9.3.8 Inclusive learning environments

Building on the policy recommendations and advocacy of civil society organisations that represent Gypsy, Traveller and Roma people, pro-active measures are required to promote cultural understanding and to build better relationships and trust between schools, colleges and young people from the communities and their families and carers at all stages including early years environments, to schools and colleges, and higher education

institutions. Special measures are required to build better relationships and trust between schools, colleges and young people from the communities and their families and carers and to tackle identity related bullying, especially within secondary schools. As noted in section 3, good practice guidelines for promoting an inclusive teaching and learning environment for Gypsy, Traveller and Roma young people while good practice in widening participation in higher education includes the importance of adopting a whole university approach via 'the pledge'.

9.5 Future research plans

The current project was envisaged as an initial pilot study that would provide the foundations for securing larger project funding to undertake a larger and richer cohort study on Gypsy, Traveller and Roma children and young people. Going forward, we hope to deepen and radically upscale the initial study by extending our analysis of large-scale administrative datasets to address a series of research questions that have emerged from the initial round of quantitative analysis using the GUiE Wave 1 data and from the feedback from Gypsy, Traveller and Roma young people and their families at our engagement workshops.

More specifically, we plan to exploit the research potential of Wave 1-3 more fully in three ways. First, we plan to expand our existing cohort study to integrate the additional information that has been made available in Waves 2 and 3 GUiE on non-attainment outcomes including exclusions, absences, SEND. This will enable us to examine several research questions motivated by the feedback from the direct engagement workshops reported in section 8 of this report, including important questions relating to whether children and young people from the communities are less likely to receive support for special educational needs and whether intersectional inequalities impact on the exclusion rates of young people from the communities who experience SEND. Exploiting of the Waves 2 and 3 English School Census data will also enable us to examine drop-out rates and school stability amongst young people in the cohort and to address other important questions relating to receipt versus eligibility of pupil premium (there is evidence that pupil premium rates are not as high as they should be amongst Gypsy, Traveller and Roma young people). Second, again by exploiting GUiE Wave 2 and Wave 3 GUiE data, we hope to broaden our analysis of the life-paths, circumstances and needs of Gypsy, Traveller and Roma young people by incorporating additional information about the interactions of Gypsy, Traveller and Roma young people with children's social care services. Here, we will use the additional data collected through

the Children in Need Census and the Children Looked After return and made available through the Wave 2 and 3 release. Third, we aim to examine the ILR content of GUiE in more depth, exploring what it means to be a 16-18 learner at Level 2 or below and 'what works' in terms of skills and training pathways for young people from the communities (including 'what works' in terms of successful transitions to Level 3).

Looking beyond the current GUiE Wave 1-Wave 3 builds, we also hope to exploit the linkage to produce new data on Gypsy, Traveller and Roma participation in higher education data when legal arrangements for HESA to be linked in to GUiE are put into place and to make use of the new improved ethnicity classifications included in Census 2021 data if this is linked in. Finally, we are currently exploring options for supplementing our GUiE analysis with analysis of other large scale data linkages such as the Longitudinal Educational Outcomes dataset (LEO). This will enable us to extend our existing Wave 1 GUiE cohort analysis to cover additional parts of the life-course. For example, by making use of the more detailed information from the National Pupil Database included in LEO (supporting analysis of earlier school experiences, progression and school level effects) and drawing on the additional information included in LEO about labour market outcomes.

References

Administrative Data Research UK (2020) Trust Security and Public Interest: Striking the Balance. Trust_Security_and_Public_Interest-_Striking_the_Balance-_ADR_UK_2020.pdf (adruk.org)

Administrative Data Research UK (2021) ADR UK Public Engagement Strategy. ADR_UK_Public_Engagement_Strategy_2021-2026.pdf (adruk.org)

Administrative Data Research UK (2022) Growing Up in England Dataset, key messages from the ADR England Children and Young People Representative Panel GUIE_database_key_messages_from_stakeholders_report_Feb_2022.pdf (adruk.org)

Association of Directors of Children's Services (2021). Elective Home Education Survey 2021. Available at: https://adcs.org.uk/assets/documentation/ADCS_EHE_Survey_2021_Report_FINAL.pd

Blackwell L and Rogers N (2021) Longitudinal linkage of administrative data; design principles and the total error framework Longitudinal linkage of administrative data; design principles and the total error framework - GOV.UK (www.gov.uk)

Burchardt T., Obolenskaya P., Vizard P and Battaglini M., (2018) Experience of multiple disadvantage among Roma, Gypsy and Traveller children in England and Wales. Available at: Microsoft Word - Disadvantage among RGT children_CASEPAPER208 (lse.ac.uk)

Cabinet Office (n.d.) mini website on Gypsy, Traveller and Roma ethnic classifications. Available at: ONS detailed census categories 2001 plus Gypsy/Roma and Irish Traveller (ethnicity-facts-figures.service.gov.uk)

Cabinet Office (2018) Race Disparity Audit: Summary Findings from the Ethnicity Facts and Figures website. Available at: Microsoft Word - Revised RDA report March 2018.docx (publishing.service.gov.uk)

Cabinet Office (2022) Gypsy, Roma and Irish Traveller ethnicity summary. Available at: Gypsy, Roma and Irish Traveller ethnicity summary - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)

Children's Commissioner for England (2017). Childhood vulnerability in England 2017. Available at: Childhood vulnerability in England 2017 | Children's Commissioner for England (childrenscommissioner.gov.uk)

Children's Commissioner for England (2019). Childhood vulnerability in England 2019. Available at: Childhood vulnerability in England 2019 | Children's Commissioner for England (childrenscommissioner.gov.uk)

Children's Commissioner for England (2021) 'The Big Answer' The Big Ask - The Big Answer (childrenscommissioner.gov.uk)

Children's Commissioner for England (2022) Mini website: What we learnt from Gypsy, Roma, and Traveller children who responded to The Big Ask | Children's Commissioner for England (childrenscommissioner.gov.uk)

Cemlyn S, Greenfields M, Burnett S, Matthews Z and Whitwell C (2009) Inequalities experienced by Gypsy and Traveller communities: A review. Available at: research-report-12-inequalities-experienced-by-gypsy-and-traveller-communities.pdf (equalityhumanrights.com)

Commission on Race and Ethnic Disparities (2021)
The report of the Commission on Race and Ethnic Disparities - GOV.UK (www.gov.uk)

Connell G (2022a) Personal communication

Connell G (2022b) Personal communication

Department for Education (2014) mini-website: Gypsy, Roma and Traveller pupils: supporting access to education - Case study - GOV.UK (www.gov.uk)

Department for Education (2019) Unique pupil numbers Department for Education (publishing.service.gov.uk)

Department for Education (2021) Skills for Jobs: Lifelong Learning for Opportunity and Growth (publishing.service.gov.uk)

Department for Education (2022) Revised review of post-16 qualifications at level 2 and below in England - Government consultation (education.gov.uk)

Department for Education (2022b) National Pupil Database Privacy Notice

Department for Education (2023) Post-16 qualifications at level 2 and below from spring 2023 - GOV.UK (www.gov.uk)

Department for Levelling Up, Housing and Communities (DLUHC) (2022) Announcement on capital funding. Available here: Traveller Site Fund 2022/23 - GOV.UK (www.gov.uk).

Derrington, C. and Kendall, S. (2004). Gypsy Traveller Students in Secondary Schools: Culture, Identity and Achievement. Stoke on Trent: Trentham Books.

Derrington, C. and Kendall, S. (2007). 'Still in school at 16? Conclusions from a longitudinal study of Gypsy Traveller students in English secondary schools.' In: Bhatti, G., Gaine, C., Gobbo, F. and Leeman, Y. (Eds) Social Justice and Intercultural Education: an Open Ended Dialogue. Stoke-on-Trent: Trentham Books

De Souza, Dame R (2023) Interview with BBC Today Programme 22/3/2023 at 7.37.

Diamond I (2021a) Foreword. In ONS (2021a) Joined up data in government: the future of data linkage methods. Available at: Joined up data in government: the future of data linking methods - GOV.UK (www.gov.uk)

Diamond I (2021b) Ensuring everyone is counted in UK data and evidence | National Statistical (ons.gov.uk)

Doidge J.,Christen P and Harron K (2021) Quality assessment in data linkage
Quality assessment in data linkage - GOV.UK (www.gov.uk)

EHRC (2010) How fair is Britain? Available at: how-fair-is-britain.pdf (equalityhumanrights.com)

EHRC (2015) Is Britain Fairer: Available at: Is Britain Fairer? | The state of equality and human rights 2015 (equalityhumanrights.com)

EHRC (2016). England's most disadvantaged groups: Gypsies, Travellers and Roma An Is England Fairer? Review spotlight report. Available at: is-england-fairer-2016-most-disadvantaged-groups-gypsies-travellers-roma.pdf (equalityhumanrights.com)

EHRC (2018) Is Britain fairer? Available at: Is Britain Fairer? (2018) | Equality and Human Rights Commission (equalityhumanrights.com)

Friends, Families and Travellers (2021) Locked Out: 74% of GPs refused registration to nomadic patients during COVID-19 Pandemic - Friends, Families and Travellers (gypsy-traveller.org)

Friends, Families and Travellers (2022a) Briefing_Health-inequalities-experienced-by-Gypsies-and-Travellers-in-England.pdf (gypsy-traveller.org).

Friends, Families and Travellers (2022b) Shadow Report to the Universal Periodic Review Working Group and UN Member States: Focus Report on Gypsy, Roma and Traveller Communities. Available at:
FFT_GRT_Coalition_UPR_2022-1.pdf (gypsy-traveller.org)

Friends, Families and Travellers (2022c) Mini website, Police Act: What you need know - Friends, Families and Travellers (gypsy-traveller.org)

Friends, Families and Travellers (n.d.) Frequently Asked Questions - Friends, Families and Travellers (gypsy-traveller.org)

Friends, Families and Travellers (n.d.b) Mini website, Teaching Resources - Friends, Families and Travellers (gypsy-traveller.org).

Gordon E. and Humpherson E., (2022). Forewords, A UK-wide public dialogue exploring what the public perceive as 'public good' use of data for research and statistics – Office for Statistics Regulation (statisticsauthority.gov.uk)

Green F et al (2017) Who chooses private schooling in Britain and why? LLAKES research paper. Green-Anders-Henderson-Henseke.pdf (llakes.ac.uk)

Greenfields M (2019) Report of a Roundtable Meeting on Access to Higher Education for members of Gypsy, Traveller and Roma (GTR) communities. Available at:
17932_Greenfields_M.pdf (guildhe.ac.uk).

Greenfields M (2022) Personal communication

Greenfields M et al (2022) Higher-Education-Opportunities-for-Gypsy-Roma-Traveller-Young-People-in-West-Yorkshire-A-Go-Higher-West-Yorkshire-Summary-Report.pdf (gohigherwestyorks.ac.uk)

Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, Goldstein H. (2017) Challenges in administrative data linkage for research. Big Data Soc. 2017 Dec 5;4(2):2053951717745678. doi: 10.1177/2053951717745678. PMID: 30381794; PMCID: PMC6187070.

HM Government (2023) Special Educational Needs and Disabilities (SEND) and Alternative Provision (AP) Improvement Plan (publishing.service.gov.uk)

House of Commons Committee of Public Accounts (2019) Challenges in using data across Government (parliament.uk)

House of Commons Education Committee (2022a). Educational challenges of children and young people from Gypsy, Roma and Traveller backgrounds. Available at: https://committees.parliament.uk/publications/9569/documents/162030/default/

House of Commons Education Committee (2022b). Educational challenges of children and young people from Gypsy, Roma and Traveller backgrounds – call for evidence. Available at: Educational challenges faced by children from Gypsy, Roma and Traveller backgrounds: Committee launches call for written submissions - Committees - UK Parliament

Jay AJ, Mc Grath-Lone L and Gilbert R (2019) Data Resource: the National Pupil Database (NPD). Available at: https://ijpds.org/article/view/1101.

Kashef S (2023) Personal communication

Longhurst et al (n.d.) Statistical Disclosure Control for the 2011 UK Census. Office of National Statistics.

Lupton, R., Thomson S., Velthuis S., and Unwin L. (2021) Moving on from initial GCSE 'failure': Post-16 transitions for 'lower attainers' and why the English education system must do better. Available at: Microsoft Word - FINAL main report for publishing.docx (manchester.ac.uk).

Ministry of Housing, Communities and Local Government (MHCLG) (2019), New national strategy to tackle Gypsy, Roma and Traveller inequalities. Available here: New national strategy to tackle Gypsy, Roma and Traveller inequalities - GOV.UK (www.gov.uk)

Ministry of Housing, Communities and Local Government and Department for Education (2021), '£1 million education programme for Gypsy, Roma and Traveller children announced'. Available at: www.gov.uk.

Morgan J, McDonagh C and Acton T., (2023) Outsider status, and racialised habitus: the experiences of Gypsy, Roma, and Traveller students in higher education, British Journal of Sociology of Education, DOI: 10.1080/01425692.2023.2167702

Mulcahy E, Baars S, Bowen-Viner K and Menzies L (2017). The underrepresentation of Gypsy, Roma and Traveller pupils in higher education A report on barriers from early years to secondary and beyond. Available at: KINGWIDE_28494_FINAL.pdf (cfey.org)

National Centre for Family Hubs (n.d.). Access and inclusion toolkit. Access and inclusion - National Centre for Family Hubs

Obolenskaya P (2022a) Output 7 (unpublished note produced for Statistical Disclosure Control purposes dated 22/7/2022)

Obolenskaya P (2022b) Output 8 (note produced for Statistical Disclosure Control purposes dated 1/9/22)

Office for National Statistics [ONS] (2013) Beyond 2011: Matching anonymous data. Available at:
[ARCHIVED CONTENT] UK Government Web Archive - The National Archives

Office for National Statistics [ONS] (2019) Admin-based qualification statistics, feasibility research: England. Available at: Admin-based qualification statistics, feasibility research: England.

Office for National Statistics [ONS] (2019b) Freedom of Information request: 2011 Census data to attainment data
2011 Census data to attainment data - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2020a) Educational attainment and household composition, feasibility research and methodology - Office for National Statistics (ons.gov.uk).

Office for National Statistics [ONS] (2020b) Data collection transformation - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2021a) Joined up data in government: the future of data linkage methods. Available at: Joined up data in government: the future of data linking methods - GOV.UK (www.gov.uk)

Office for National Statistics [ONS] (2022a) Gypsies' and Travellers' lived experiences, overview, England and Wales. Available at: Gypsies' and Travellers' lived experiences, overview, England and Wales - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2022b) Gypsies' and Travellers' lived experiences, education and employment, England and Wales. Available at: Gypsies' and Travellers' lived experiences, education and employment, England and Wales - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2022c) Methodology: Who are the children entering care in England? - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2022d) Growing Up in England – methodology Growing Up in England (GUiE) - Office for National Statistics (ons.gov.uk)

ONS (2022e) Growing Up in England Data: personal characteristics, general breakdowns, individualised learner records, household characteristics and educational attainment, vulnerable groups and educational attainment. Available at: Growing Up in England (GUiE) - Office for National Statistics (ons.gov.uk)

ONS (2022f) Who are the children entering care in England? - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2023) Feasibility All Education Dataset for England. Available at: Feasibility All Education Dataset for England (AEDE) - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2023b) Personal Communication

Office for National Statistics [ONS] (2023c) Understanding quality of the Statistical Population Dataset in England and Wales using the 2021 Census - Demographic Index linkage - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (2023d) Future of Population and Migration Statistics Consultation Document

Office for National Statistics [ONS] (2023e) Understanding consistency of ethnicity data recorded in health-related administrative datasets in England - Office for National Statistics

Office for National Statistics [ONS] (2023f) Methods and systems used to collect ethnicity information in health administrative data sources, England: 2022 - Office for National Statistics (ons.gov.uk)

Office for National Statistics [ONS] (n.d.) ONS working paper series no 19 – An error framework for longitudinal administrative sources; its use for understanding the statistical properties of data for international migration - Office for National Statistics

Office for National Statistics [ONS] (n.d.b) GUiE User Guide, available at: MDX Browser > Growing Up in England Wave 1 @ 12169 (metadata.works)

Office for Statistics Regulation n.d.a (Quality Assurance of Administrative Data – Office for Statistics Regulation (statisticsauthority.gov.uk))

Office for Statistics Regulation n.d.b (Case example – DfE's school-level examination results statistics – Office for Statistics Regulation (statisticsauthority.gov.uk)).

Office for Statistics Regulation (2021). Defining-the-Public-Good-in-Applications-to-Access-Public-Data.pdf (statisticsauthority.gov.uk)

Office for Statistics Regulation and ADR UK (2022) A UK-wide public dialogue exploring what the public perceive as 'public good' use of data for research and statistics – Office for Statistics Regulation (statisticsauthority.gov.uk)

Parry G, Van Cleemput P, Peters J, Walters S, Thomas K, Cooper C (2007). 'Health status of Gypsies and Travellers in England' in *Journal of Epidemiological Community Health* 61(3):198-204. doi: 10.1136/jech.2006.045997.

Race Disparities Audit (2017). See Cabinet Office 2018.

Race Disparity Unit (2020). Ethnicity data: how similar or different are aggregated ethnic groups? - GOV.UK (www.gov.uk)

Race Disparity Unit (2021). Final report on progress to address COVID-19 health inequalities, Appendix D: Further data and evidence - GOV.UK (www.gov.uk).

Raleigh V (2023) Blog: When will the government and national agencies act to ensure that ethnicity coding in health records is fit for purpose? | The King's Fund (kingsfund.org.uk)

Sebba J, Berridge D, Luke N, Fletcher J, Bell K, Strand S., Thomas S, Sinclair I, O'Higgins A (2015) The Educational Attainment and Progress of Looked After Children in England: Linking Care and Education Data

Smith, S., Greenfields, M., and Rochon, R. 2021. Gypsies, Travellers, and Roma in UK Higher Education: A View from Inside the Academy. in: eds. Simmons, N. and Podrebarac, J. *Adults in the Academy: Voices of Lifelong Learners. Critical Issues in the Future of Learning and Teaching* (20). Brill, pp. 101-109. Available at: 18462_Greenfields_M.pdf (guildhe.ac.uk).

Taylor B (2019) Understanding highest educational qualification: The case for using Administrative Data | National Statistical (ons.gov.uk)

Timpson Review of School Exclusion (2018) Timpson Review of School Exclusion (publishing.service.gov.uk)

Traveller Movement (2017) The last acceptable form of racism? The pervasive discrimination and prejudice experienced by Gypsy, Roma and Traveller communities
The-Last-Acceptable-Form-of-Racism-2017.pdf (travellermovement.org.uk)

Traveller Movement (2019) Barriers in education – young Travellers in London. Available at: TTM-Barriers-in-education_web.pdf (travellermovement.org.uk)

Traveller Movement (2019b) Good Practice Education Guide
Good-Practice-Guide-Education-2019.pdf (travellermovement.org.uk)

Traveller Movement (2020) Gypsy, Roma and Traveller experiences in Secondary Education: Issues, barriers and recommendations. Available at: GRT-in-Secondary-Education-2021.pdf (travellermovement.org.uk)

Traveller Movement (2021) Gypsy, Roma and Traveller access to Further Education: 14-16 provision, vocational qualifications, apprenticeships and A Levels. Available at: GRT-Access-to-Further-Education-2021.pdf (travellermovement.org.uk)

Traveller Movement (2022) The Importance of accurate ethnic monitoring and data inclusion for GRT communities. Available at 2022.03 Ethnic Monitoring Briefing (travellermovement.org.uk)

UKDS (n.d.) Classification of household deprivation (Great Britain) 2011 - Dataset - UK Data Service CKAN

UK Statistics Authority [UKSA] (2020) Statistics for the public good 1342 UKSA Strategy 2020 v1-01.indd (statisticsauthority.gov.uk)

UKSA (2021a) Inclusive Data Taskforce Recommendations report - Leaving no one behind. How can we be more inclusive in our data?. Available at: Inclusive Data Taskforce recommendations report: Leaving no one behind – How can we be more inclusive in our data? – UK Statistics Authority

UKSA (2021b) Inclusive data taskforce: Response by the National Statistician. Available at: Inclusive data taskforce Response by the National Statistician (statisticsauthority.gov.uk)

UK Statistics Authority [UKSA] (2023) Quality issues related to administrative data. Available at EAP190 - Quality of admin data linkage (statisticsauthority.gov.uk).

UK Statistics Authority [UKSA] (n.d.) Ethical principles. Mini-website:
Ethical Principles – UK Statistics Authority

Vizard P., Burchardt T., Obolenskaya P., Shutes I. and Battaglini M. (2018) Multidimensional child poverty and disadvantage: tackling 'data exclusion' and extending the evidence base on missing and invisible children

Waind E (2019) Key questions for children's services: Identifying the gaps.
Key questions for children's services: Identifying the gaps - ADR UK

Waind E (2021) Personal communication

Wilkin, Derrington and Foster (2009) Improving the outcomes for Gypsy, Roma and Traveller pupils: literature review. Available at: Improving the outcomes for Gypsy, (publishing.service.gov.uk).

Wilkin A, Derrington C, White R, Martin K, Foster B, Kinder K, and Rutt S (2010) Improving the outcomes for Gypsy, Roma and Traveller pupils: final report. Available at:
DFE-RR043.pdf (publishing.service.gov.uk)

Women and Equalities Committee (2019) Tackling inequalities faced by the Gypsy, Traveller and Roma communities. Available here: Tackling inequalities faced by the Gypsy, Roma and Traveller communities (parliament.uk)

Appendix

GUiE legal gateways and issues of informed consent

Legal gateways for sharing data from the Department for Education with ONS are Section 45A of the Statistics and Registration Service Act (2007), Section 87 of the Education and Skills Act 2008, Section 537A of the Education Act 1996 and Section 83 of the Children Act (1989) (ONS n.d.b). The Digital Economy Act (2017) amended the Statistics and Registration Services Act (2007). The enactment of the Digital Economy Act (DEA) (2017) enables public authorities to share personal data for defined purposes for public benefit while protecting individual rights to privacy and provides a new legal gateway for the release of de-identified administrative research datasets under controlled and regulated conditions. To uphold individual rights to privacy and confidentiality, the identity of data subjects must be protected when data is shared. To comply with this requirement, administrative research datasets are de-identified and can generally only be accessed within regulated secure (trusted) environments by researchers who are accredited under the DEA (meaning that they have appropriate skills to use data ethically and to protect individual rights to privacy and confidentiality set out in the DEA, the GDPR and other data protection and human rights legislation, including training in statistical disclosure control and the 'Five Safes'). The DEA also provides the legal framework for the work of Administrative Data Research UK, the funder of the current project, whose remit is to promote administrative data research in the public good.

While the lawful basis of the creation of GUiE and its use (or processing) within the current project is not informed consent, the principle of informed consent is nevertheless central to research ethics and is one of the key ethical principles addressed by NSDEC. In terms of the informed consent framework for the Growing Up in England (GUiE) dataset, our understanding is that neither the Census 2011 nor the underlying administrative education, college or work-based data collections asked for explicit consent for data to be linked. An ONS response to a freedom of information request about the linked Census-feasibility AEDE data addressed issues of informed consent relating to both the census and feasibility AEDE components of the GUiE dataset. The ONS response notes that the privacy information provided to Census 2011 respondents did not include specifics of the linkage. However, respondents were informed that census data could be used for other statistical purposes in the 2011 Census White Paper, including the creation of linked samples for follow-up on topics that were not included in the Census itself. Additionally, this possibility was

publicised through workshops with researchers in the field of child welfare and education including BBC's Children in Need, NSPCC and Children's Commissioner for England and by publishing news items on the Children and Young People page of the ADR UK website (ONS 2019b).

In relation to the request for information about informed consent and education data, ONS (2019b) noted that it did not hold a copy of the privacy information provided to children or parents during collection of the education data, which would be a matter for DfE. A privacy notice is available on the DfE website (DfE 2022b) which sets out the lawful basis of data sharing (although this does not specifically refer to informed consent).

2011 Census: person response rates, ethnicity question item non-response and statistical disclosure control.

A useful overview of 2011 Census methodology and quality is provided at the mini-website here: Quality and methods - Office for National Statistics (ons.gov.uk). Information on response and imputation rates is archived here: [ARCHIVED CONTENT] Response and imputation rates - ONS (nationalarchives.gov.uk)

Figure 18*:* Overall 2011 Census response rate by ethnic group

The ONS document "Response Rates in the 2011 Census" includes the following information and table: "The 2011 Census person response rate is the total number of usual residents for whom individual details were provided on a returned questionnaire, divided by the estimated usually resident population … Person response rate targets were set at 94 per cent for England and Wales overall; and at least 80 per cent in every local authority area. For England and Wales, the person response rate achieved was 94 per cent (94 per cent for England and 93 per cent for Wales). Person response rates by local authority ranged from 82 per cent to 98 per cent".

Table 3 – Census response rates for England and Wales by ethnic group

Ethnic group	Persons %	Males %	Females %
White: English/Welsh/Scottish/Northern Irish/British	95.1	94.4	95.9
White: Irish	94.0	92.4	95.5
White: Gypsy or Irish Traveller	90.1	88.9	91.2
White: Other White	90.3	88.0	92.4
Mixed/multiple ethnic groups: White and Black Caribbean	83.4	81.7	85.0
Mixed/multiple ethnic groups: White and Black African	82.8	80.6	85.0
Mixed/multiple ethnic groups: White and Asian	85.4	83.3	87.5
Mixed/multiple ethnic groups: Other Mixed	82.5	79.9	85.0
Asian/Asian British: Indian	94.3	92.9	95.7
Asian/Asian British: Pakistani	93.5	92.5	94.6
Asian/Asian British: Bangladeshi	92.5	91.1	94.1
Asian/Asian British: Chinese	84.6	81.2	87.6
Asian/Asian British: Other Asian	85.1	81.9	88.1
Black/African/Caribbean/Black British: African	88.2	85.8	90.5
Black/African/Caribbean/Black British: Caribbean	91.9	90.3	93.4
Black/African/Caribbean/Black British: Other Black	64.0	60.1	68.1
Other ethnic group: Arab	72.4	68.7	77.5
Other ethnic group: Any other ethnic group	74.0	69.5	79.4

Source: ONS [Response Rates in the 2011 Census report, available at the archived mini-website: ARCHIVED CONTENT Response and imputation rates - ONS (nationalarchives.gov.uk)]

Figure 19*:* 2011 Census ethnicity question – item non-response and imputation

The ONS document "Item Edit and Imputation: Evaluation Report" (pages 58-59 and 27) includes the following information, tables and figure:

"Ethnic group was collected with 18 tick boxes and five write-in boxes. Coding rules were used to allocate each response to a single code from the ethnicity index. These have been aggregated to the 18 categories provided in [the Table below] for analysis. Non-response for the derived variable was 3%, including around 0.3% which were unresolved because they included more than three high-level ethnic groups (for example Other White, Bangladeshi and Caribbean). While there were no edit rules for ethnicity, around 21,000 values (0.04%) were changed because of inconsistency with the term-time routing filter. In the imputed values, English was 11.2% less frequent while the other categories tended to be more frequent than observed. This was mainly due to geographic effects. "

"There was one issue with the imputation, and this was for the Arab category. Due to an error in the coding of ethnicity, those that ticked 'Arab' on the questionnaire ended up with the code for 'Other'. Although these were all corrected after imputation, it meant that the Arab ethnic group was underrepresented in the donor pool, and only 2,000 persons were imputed into the Arab ethnicity. However, this did not affect the final distribution which decreased by less than 0.01%. There was a minor downward adjustment of 0.3% in the total distribution for English, while the changes to other categories were negligible (≤ 0.1%)."

Table 37: Distribution of ethnic group

All eligible responding persons, England and Wales, 2011 — Thousands

	Observed responses		Imputed responses		Difference (imputed - observed)	Total including imputed		Change (total - observed
	N	%	N	%	%	N	%	%
English	41,832	81.7	1,140	70.6	**-11.2**	42,973	81.4	**-0.3**
Irish	487	1.0	16	1.0	**<0.1**	503	1.0	**<0.1**
Gypsy/traveller	51	0.1	1	0.1	**<0.1**	52	0.1	**<0.1**
Other white	2,245	4.4	105	6.5	**2.1**	2,350	4.5	**0.1**
Mixed Caribbean	348	0.7	10	0.6	**<0.1**	358	0.7	**<0.1**
Mixed African	135	0.3	6	0.4	**0.1**	141	0.3	**<0.1**
White Asian	312	0.6	10	0.6	**<0.1**	322	0.6	**<0.1**
Other Mixed	158	0.3	5	0.3	**<0.1**	163	0.3	**<0.1**
Indian	1,329	2.6	63	3.9	**1.3**	1,392	2.6	**<0.1**
Pakistani	1,049	2.0	53	3.3	**1.2**	1,102	2.1	**<0.1**
Bangladeshi	404	0.8	19	1.2	**0.4**	423	0.8	**<0.1**
Chinese	349	0.7	22	1.4	**0.7**	372	0.7	**<0.1**
Other Asian	687	1.3	55	3.4	**2.0**	742	1.4	**0.1**
African	887	1.7	64	3.9	**2.2**	951	1.8	**0.1**
Caribbean	536	1.0	24	1.5	**0.4**	560	1.1	**<0.1**
Other Black	146	0.3	7	0.4	**0.1**	153	0.3	**<0.1**
Arab	20	<0.1	2	0.1	**0.1**	21	<0.1	**<0.1**
Other ethnicity	200	0.4	15	0.9	**0.5**	215	0.4	**<0.1**
Total	**51,175**	**100**	**1,616**	**100**	**0.0**	**52,791**	**100**	**0.0**

Figure 7: Maximum absolute percent change in any category after imputation – culture questions

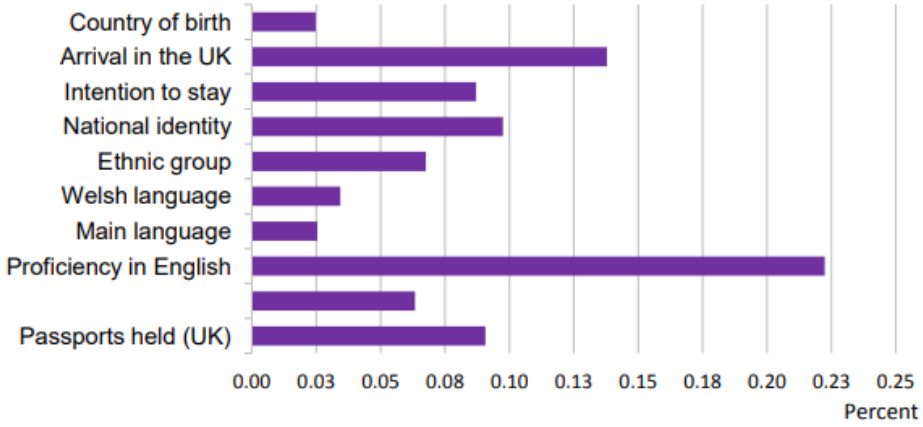


Table 8: Culture module: non-response, edit failure and imputation

All eligible** responding persons, England and Wales, 2011 — Thousands

	All** persons N	Non-response N	Edit Failure N	Total imputed N	*Non-response %*	Edit Failure %	*Total imputed %*
Country of birth[1]	52,791	800	7	**806**	*1.5*	*<0.1*	***1.5***
Arrival in the UK[1]	6,858	326	128	**453**	*4.8*	*1.9*	***6.6***
Intention to stay[1,2]	594	86	6	**92**	*14.5*	*0.9*	***15.5***
National identity[1]	52,791	1,023	5	**1,027**	*1.9*	*<0.1*	***2.0***
Ethnic group[1]	52,791	1,595	21	**1,616**	*3.0*	*<0.1*	***3.1***
Welsh language[1,3]	2,861	96	1	**97**	*3.4*	*<0.1*	***3.4***
Main language[1]	52,791	1,328	41	**1,369**	*2.5*	*0.1*	***2.6***
Proficiency in English[1]	3,929	142	<1	**143**	*3.6*	*<0.1*	***3.6***
Religion[1]	53,068	3,759	<1	**<1**	*7.1*	*<0.1*	***0.0***
Address one year ago[1,4]	52,150	2,004	97	**2,101**	*3.8*	*0.2*	***4.0***
Address one year ago country[1,5]	605	22	4	**26**	*3.6*	*0.7*	***4.4***
Address one year ago postcode[1,6]	4,074	235	1	**236**	*5.8*	*<0.1*	***5.8***
Passports held (UK)[1]	52,791	1,222	33	**1,254**	*2.3*	*0.1*	***2.4***
Passports held (other)[1]	3,963	93	1	**94**	*2.4*	*<0.1*	***2.4***

** All persons eligible to respond:

1. All persons at, or without, another address during term time
2. Where arrive to live in UK was less than 12 months ago
3. Where country of residence was Wales
4. Where aged at least one year old
5. Where address one year ago was within UK
6. Where address one year ago was outside UK

Source: ONS [Item Edit and Imputation: Evaluation Report, available at the archived mini-website: ARCHIVED CONTENT Response and imputation rates - ONS (nationalarchives.gov.uk)]

Statistical disclosure control

According to information provided by ONS, statistical disclosure control has been applied to 2011 Census data to protect confidentiality and privacy. Two complementary strategies to protect individuals while minimising damage to the results were applied, specifically targeted record swapping and restriction of detail particularly at low level geographies. ONS analysis suggests that the key strength of record swapping is that no persons or data items are removed from the census data and therefore outputs at national level and high geographies will be unaffected by record swapping.

Source: ONS [Statistical Disclosure Control for the 2011 Census, available at: ARCHIVED CONTENT] UK Government Web Archive - The National Archives]

Selected additional statistics – population estimates and breakdowns

Table 12: 2011 Census: Usual resident population by five-year age group and sex (England and Wales)

All usual residents	
Age	**Persons**
All ages	56,075,912
0 – 4	3,496,750
5 – 9	3,135,711
10 – 14	3,258,677
15 – 19	3,539,385
20 – 24	3,807,245
25 – 29	3,836,609
30 – 34	3,683,915
35 – 39	3,732,161
40 – 44	4,099,089
45 – 49	4,100,526
50 – 54	3,601,694
55 – 59	3,183,915
60 – 64	3,377,162
65 – 69	2,674,161
70 – 74	2,178,672
75 – 79	1,777,547
80 – 84	1,338,005
85 – 89	825,671
90 and over	429,017

Source: ONS 2011 Census: Population and Household Estimates for the United Kingdom - Office for National Statistics (ons.gov.uk) – Table 1 (2011 Census: Usual resident population by five-year age group and sex, United Kingdom and constituent countries Excel shsset)

Table 13: Number and percentage of individuals identified as Gypsy, Irish Traveller or Roma in the 2011 Census and the 2021 Census

	Within the English and Welsh population			
	2011 Census		2021 Census	
	Number	Percent	Number	Percent
White: Gypsy or Irish Traveller	57680	0.1	67768	0.1
White: Roma	-	-	100981	0.2

Source: ONS – Census 2021 and Census 2011 estimates, available at Ethnic group, England and Wales - Office for National Statistics (ons.gov.uk) Figure 2.
Note: There was no “Roma” ethnic group tick-box in 2011.

Selected additional statistics – number of pupils and available breakdowns

Table 14: Pupils identified as Gypsy, Traveller or Roma in state funded schools, pupil referral units and alternative provision (England)

A: January 2011 English Schools Census

	Pupils of compulsory school age and above									
	State-funded primary schools (1)(2)		State-funded secondary schools (1)(3)		Special schools (4)		Pupil Referral Units (6)		Alternative Provision (7)	
	Number	Percentage (8)	Number	Percentage (8)	Number	Percentage (8)	Number	Percentage (8)	Number	Percentage (8)
Traveller of Irish heritage	3,035	0.1	1,150	0.0	80	0.1	45	0.3	155	0.7
Gypsy/ Roma	8,425	0.3	4,320	0.1	200	0.2	100	0.7	310	1.4
All pupils (11)	**3,256,015**	**100.0**	**3,258,805**	**100.0**	**87,950**	**100.0**	**14,045**	**100**	**22,070**	**100**

Source: Schools, pupils and their characteristics: January 2011 - GOV.UK (www.gov.uk) (Main national tables, Table 4)

Notes:

1. One teacher classes as taught during a single selected period in each school on the day of the census in January.
2. Includes middle schools as deemed. Includes primary academies
3. Includes Reception classes.
4. Classes of 31 or more, excluding those large classes engaged in an excepted activity.
5. Includes only those classes where there are pupils who have been admitted as exceptions making the class size within the legal limit of 30 pupils or less.
6. From 2007 a revision to the guidance issued to schools means that information on classes engaged in an excepted activity should not be provided (information should be provided on the next ordinary teaching session).
7. Includes all pupils in large classes, including those with excepted activity.

Numbers have been rounded to the nearest 5. There may be discrepancies between the sum of constituent items and totals as shown.

B: January 2015 English Schools Census

	Pupils of compulsory school age and above									
	State-funded primary schools (1)(2)		State-funded secondary schools (1)(3)		Special schools (4)		Pupil referral units (5)		Local authority alternative provision (6)	
	Number	Percentage (8)	Number	Percentage (8)	Number	Percentage (8)	Number	Percentage (8)	Number	Percentage (8)
Traveller of Irish heritage	3,455	0.1	1,103	0.0	107	0.1	62	0.5	66	0.3
Gypsy/ Roma	12,936	0.4	6,990	0.2	328	0.3	138	1	150	0.8
All pupils (11)	**3,572,108**	**100.0**	**3,175,013**	**100.0**	**97,988**	**100.0**	**13,575**	**100.0**	**19,639**	**100.0**

Source: Schools, pupils and their characteristics: January 2015 - GOV.UK (www.gov.uk) (National tables, Table 4a)

Notes:

1. Includes middle schools as deemed.
2. Includes all primary academies, including free schools.
3. Includes city technology colleges and all secondary academies, including free schools, university technical colleges and studio schools.
4. Includes maintained special schools, special academies and non-maintained special schools. Excludes general hospital schools.
5. Includes pupils who are sole or dual main registration and pupils registered in other providers and further education colleges. Excludes 9,685 dual subsidiary registered pupils. Includes alternative provision academies and free schools.
6. Includes local authority maintained alternative provision where information on registration status is not collected for pupils in alternative provision.
7. Pupils of compulsory school age and above were classified according to ethnic group. Includes pupils who are sole or dual main registrations.
8. The number of pupils by ethnic group expressed as a percentage of all pupils of compulsory school age and above.
9. Information refused or not obtained.
10. Includes all pupils classified as belonging to an ethnic group other than White British.
11. All pupils of compulsory school age and above.

Totals may not appear to equal the sum of the component parts because numbers have been rounded to the nearest 5.

C: January English Schools Census (2021-23)

	Headcount			Percent		
	2020/21	**2021/22**	**2022/23**	**2020/21**	**2021/22**	**2022/23**
Total	8,342,004	8,418,013	8,481,287	100.0	100.0	100.0
Any other ethnic group	173,423	185,355	195,998	2.1	2.2	2.3
Asian - Any other Asian background	161,435	170,572	179,066	1.9	2.0	2.1
Asian - Bangladeshi	149,202	151,613	151,644	1.8	1.8	1.8
Asian - Chinese	37,943	48,396	58,731	0.5	0.6	0.7
Asian - Indian	275,106	289,179	311,012	3.3	3.4	3.7
Asian - Pakistani	374,031	380,781	383,457	4.5	4.5	4.5
Black - Any other Black background	64,655	65,501	65,855	0.8	0.8	0.8
Black - Black African	326,365	338,614	362,477	3.9	4.0	4.3
Black - Black Caribbean	83,712	82,402	80,353	1.0	1.0	0.9
Mixed - Any other Mixed background	200,996	213,597	225,855	2.4	2.5	2.7
Mixed - White and Asian	128,850	134,821	139,682	1.5	1.6	1.6
Mixed - White and Black African	73,202	76,262	78,375	0.9	0.9	0.9
Mixed - White and Black Caribbean	130,569	133,504	134,456	1.6	1.6	1.6
Unclassified	132,439	136,371	140,373	1.6	1.6	1.7
White - Any other White background	565,893	575,540	608,170	6.8	6.8	7.2
White - Gypsy/Roma	26,045	27,359	28,340	0.3	0.3	0.3
White - Irish	21,898	21,495	21,124	0.3	0.3	0.2
White - Traveller of Irish heritage	6,197	6,903	6,891	0.1	0.1	0.1
White - White British	5,410,043	5,379,748	5,309,428	64.9	63.9	62.6

Source: Schools, pupils and their characteristics, Academic year 2022/23 – Explore education statistics – GOV.UK (explore-education-statistics.service.gov.uk) – Ethnicity table

Notes: (1) Totals include state-funded nursery, primary, secondary, alternative provision (AP) schools and special schools, and non-maintained special schools. Does not include independent schools.

(2) Totals include pupils in General Hospital Schools, however Ethnicity and Language data is not available for these pupils. Breakdowns may not sum to totals as a result.

(3) Prior to 2017/18, Ethnicity and Language data was collected for compulsory school age children only.

Table 15: Pupil numbers by year group (England)

A: January 2011 English Schools Census

Age as at 31 August 2010				
Pupils aged:	State-funded primary schools (1)(2)	State-funded secondary schools (1)(3)(5)	Maintained special schools (4)	Non-maintained special schools
Total Under 5	**881740**	**3825**	**5010**	**65**
5	577310	1940	3290	40
6	565840	1865	3590	65
7	548185	1755	4175	85
8	527890	1660	4540	105
9	514365	18630	5090	135
10	518755	24295	5730	175
Total 5 to 10	**3252345**	**50150**	**26420**	**610**
11	3515	545720	8525	335
12	90	554820	9295	435
13	20	568525	10000	475
14	20	558285	9870	495
15	20	558050	9865	575
Total 11 to 15	**3670**	**2785405**	**47555**	**2315**
16	0	221830	4335	455
17	0	177350	3600	400
18	0	22780	2925	310
19+	0	1295	15	260
Total 16 to 19+	**0**	**423255**	**10875**	**1430**

CSA and above (8)	**3256015**	**3258805**	**84850**	**4350**
All ages	**4137755**	**3262635**	**89860**	**4415**

Source: Schools, pupils and their characteristics: January 2011 - GOV.UK (www.gov.uk) (Main national tables, Table 1a)

Notes:

1. Includes middle schools as deemed.
2. Includes primary academies.
3. Includes city technology colleges and secondary academies.
4. Includes general hospital schools.
5. Middle deemed secondary schools and all through schools deemed as secondary result in some pupils aged less than five in secondary schools.
6. Age as at 31 August 2010.
7. Includes pupils who are sole or dual main registrations. Includes boarding pupils.
8. Pupils of compulsory school age and above.

x 1 or 2 pupils.

Pupil numbers have been rounded to the nearest 5. There may be discrepancies between the sum of constituent items and totals as shown.

B: January 2015 English Schools Census

	State-funded primary schools (1)(2)	State-funded secondary schools (1)(3)(5)	State-funded special schools (4)	Non-maintained special schools
Nursery 1	46,021	697	510	22
Nursery 2	256,137	2,463	1,578	28
Reception	636,761	6,556	3,781	35
Year group 1	629,429	5,718	4,361	73
Year group 2	634,108	5,201	4,775	70
Year group 3	608,658	4,724	5,198	92
Year group 4	588,242	4,316	5,614	114
Year group 5	561,741	17,570	6,026	163
Year group 6	547,426	19,761	6,416	187
Year group 7	1,558	542,151	8,958	316
Year group 8	227	524,416	9,138	338
Year group 9	0	528,499	9,687	363
Year group 10	0	539,224	10,046	459
Year group 11	0	543,526	10,152	463

Year group 12	0	242,846	5,231	426
Year group 13	0	194,424	4,520	408
Year group 14	0	2,538	3,163	314
Not followed	0	98	863	82
	0	0	0	0
Total	**4,510,308**	**3,184,728**	**100,017**	**3,953**

Source: Schools, pupils and their characteristics: January 2015 - GOV.UK (www.gov.uk) (National tables, Table 1d)

Notes:

1. Includes middle/all through schools as deemed.
2. Includes all primary academies, including free schools.
3. Includes city technology colleges and all secondary academies, including free schools, university technical colleges and studio schools.
4. Excludes general hospital schools. Includes special academies and free schools.
5. Middle deemed secondary schools and all through schools deemed as secondary result in some pupils aged less than five in secondary schools.
6. Includes pupils who are sole or dual main registrations. Includes boarding pupils.

x 1 or 2 pupils.

Pupil numbers have been rounded to the nearest 5. There may be discrepancies between the sum of constituent items and totals as shown.

Selected additional statistics –participation in further education and available breakdowns

Table 16: Headline statistics on participation in further education by academic year (England)

A: Individual Learner Records 19+

	2015/16	2016/17	2017/18	2018/19	2019/20	2020/21
Total Participation (Aug-Jul)	2,324,700	2,236,800	2,179,100	2,068,200	1,745,800	1,640,300

Source: Further education and skills, Academic year 2022/23 – Explore education statistics – GOV.UK (explore-education-statistics.service.gov.uk)
Notes:

(1) Volumes are rounded to the nearest 100. If shown, '~' indicates a base value of between 1 and 49, whilst 0 indicates a true zero
(2) From 2016/17, some learners who would have previously been included in full level 2 and full level 3, have been reclassified. Please see the methodology document for further information
(3) Age is based on age as at 31 August of the academic year
(4) For 2017/18, there were approximately 2,400 learners undertaking community learning missing from the final data return (R14) due to a submission error for three providers. 2017/18 figures reported here does not include these learners
(5) This release does not now include offender learners for 2019/20 onwards as this data is being collected by the Ministry of Justice. In previous releases we included offender learning in our data where relevant, which accounted for around 78,000 learners aged 18+ in 2017/18 and around 57,600 learners aged 18+ in 2018/19
(6) Total participation is the count of learners that participated at any point during the stated academic period. Learners undertaking more than one course will appear only once in the grand total
(7) Total achievements are the count of learners that achieved at any point during the stated academic period. Learners achieving more than one course will appear only once in the grand total

B: Individual Learner Records - 19+ and 16-18 for state funded FE where there is an ILR record (by ethnic group, England)

	All		Asian		Black		Mixed		White		Other		Unknown	
Year	%	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%	Number
2002/03	100	4,703,310	6.6	288,230	4.5	198,310	1.1	50,220	85.7	3,754,600	2.1	90,410	N/A	321,540
2003/04	100	5,437,420	6.4	328,880	4.4	228,550	1.3	66,730	86	4,433,090	1.9	99,110	N/A	281,060
2004/05	100	5,563,150	6.6	349,790	4.8	257,440	1.5	78,290	85.1	4,529,830	2	107,760	N/A	240,040
2005/06	100	4,893,920	7	331,740	5.3	250,040	1.7	78,360	84	3,968,180	2	95,650	N/A	169,950
2006/07	100	4,202,780	7.3	297,530	5.6	230,010	1.9	77,000	83.1	3,382,000	2.1	84,790	N/A	131,450
2007/08	100	4,360,750	7.5	318,370	5.8	246,680	2	83,390	82.7	3,496,730	1.9	82,010	N/A	133,570
2008/09	100	4,837,100	7.9	373,890	5.8	274,370	2	93,630	82.3	3,873,920	1.9	89,640	N/A	131,650
2009/10	100	4,635,510	7.9	359,290	5.9	266,330	2.1	95,980	82.1	3,716,760	2	90,010	N/A	107,140
2010/11	100	4,264,860	8.4	349,980	6.2	258,100	2.3	95,430	81.2	3,393,000	1.9	81,280	N/A	87,070
2011/12	100	4,216,610	8.4	346,860	6.5	267,800	2.5	103,840	80.7	3,342,040	1.9	78,230	N/A	77,840
2012/13	100	4,320,270	8.1	340,210	6.5	272,870	2.6	111,700	81	3,420,430	1.9	80,140	N/A	94,920
2013/14	100	3,913,520	8	306,900	6.3	242,300	2.8	106,660	80.9	3,095,990	1.9	74,580	N/A	87,090
2014/15	100	3,576,860	8.2	289,230	6.5	228,440	2.9	102,010	80.3	2,821,900	2	71,980	N/A	63,300
2015/16	100	3,274,900	8.5	271,500	6.5	206,750	3	95,240	80	2,559,790	2.1	68,330	N/A	73,290
2016/17	100	3,152,110	8.8	271,760	6.7	206,600	3.1	94,650	79.1	2,448,260	2.4	72,720	N/A	58,120
2017/18	100	3,062,000	9.2	276,510	6.8	202,730	3.2	95,450	78.3	2,350,680	2.6	77,630	N/A	59,000
2018/19	100	2,924,260	9.6	276,260	6.8	195,620	3.4	96,650	77.3	2,214,220	2.8	81,600	N/A	59,910
2019/20	100	2,579,880	9.8	246,480	6.8	171,270	3.5	88,830	76.8	1,939,390	3.1	78,180	N/A	55,730
2020/21	100	2,480,570	9.5	229,260	7	169,250	4	95,810	76.5	1,853,500	3.1	75,500	N/A	57,250

Source: Further education participation - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)

Table 17: Population census derived statistics on highest educational qualification (usual residents age 16+ England)

A: 2011 Census

Highest Level of Qualification	All categories: Highest level of qualification	No qualifications	Level 1 qualifications	Level 2 qualifications	Apprenticeship	Level 3 qualifications	Level 4 qualifications and above	Other qualifications
Ethnic Group								
All categories: Ethnic group	42,989,620	9,656,810	5,714,441	6,544,614	1,532,934	5,309,631	11,769,361	2,461,829
White: Total	37,452,110	8,696,734	5,048,772	5,849,496	1,479,500	4,677,794	9,922,903	1,776,911
English/Welsh/Scottish/Northern Irish/British	34,869,724	8,279,669	4,886,964	5,642,385	1,441,196	4,461,304	8,994,376	1,163,830
Irish	487,373	140,838	36,750	47,411	13,705	47,208	163,704	37,757
Gypsy or Irish Traveller	37,425	22,286	3,883	3,266	591	2,353	3,261	1,785
Other White	2,057,588	253,941	121,175	156,434	24,008	166,929	761,562	573,539
Mixed/multiple ethnic group: Total	654,129	101,847	95,231	115,375	11,221	101,659	184,636	44,160
White and Black Caribbean	231,322	46,179	44,654	48,459	5,351	35,498	42,161	9,020
White and Black African	81,970	10,878	11,669	14,086	1,192	12,047	23,829	8,269
White and Asian	175,072	22,065	20,430	28,127	2,375	29,870	61,218	10,987
Other Mixed	165,765	22,725	18,478	24,703	2,303	24,244	57,428	15,884
Asian/Asian British: Total	3,123,249	587,709	345,423	325,273	21,193	322,880	1,075,630	445,141
Indian	1,127,677	169,826	111,063	112,565	9,072	111,444	471,036	142,671

Pakistani	745,765	190,709	105,057	87,191	5,164	75,691	183,397	98,556
Bangladeshi	284,767	80,349	42,333	34,552	1,920	29,653	56,840	39,120
Chinese	331,103	51,444	23,653	24,516	1,272	38,685	143,194	48,339
Other Asian	633,937	95,381	63,317	66,449	3,765	67,407	221,163	116,455
Black/African/Caribbean/Black British: Total	1,342,033	195,654	185,107	215,085	18,182	165,897	445,002	117,106
African	682,466	71,539	78,715	102,606	5,334	82,802	272,909	68,561
Caribbean	487,733	95,963	77,158	82,142	10,387	61,734	126,323	34,026
Other Black	171,834	28,152	29,234	30,337	2,461	21,361	45,770	14,519
Other ethnic group: Total	418,099	74,866	39,908	39,385	2,838	41,401	141,190	78,511
Arab	157,973	23,525	14,127	12,867	714	15,810	58,706	32,224
Any other ethnic group	260,126	51,341	25,781	26,518	2,124	25,591	82,484	46,287

Source: Data Viewer - Nomis - Official Census and Labour Market Statistics (nomisweb.co.uk)

Note:

1. In order to protect against disclosure of personal information, records have been swapped between different geographic areas. Some counts will be affected, particularly small counts at the lowest geographies.

2021 Census (usual residents aged 16 years and over England)

Ethnic group	Total	No qualifications	Level 1 and entry level qualifications	Level 2 qualifications	Apprenticeship	Level 3 qualifications	Level 4 qualifications or above	Other qualifications
Total	46,006,955	8,317,790	4,456,199	6,126,127	2,446,936	7,784,977	15,606,458	1,268,468
Asian, Asian British or Asian Welsh	4,134,390	818,819	359,690	416,655	80,179	535,670	1,767,774	155,603
Black, Black British, Black Welsh, Caribbean or African	1,792,286	251,785	155,588	213,703	59,231	288,240	776,350	47,389
Mixed or Multiple ethnic groups	938,209	107,055	77,262	144,431	31,900	191,246	368,267	18,048

White	38,191,422	6,910,607	3,780,834	5,272,293	2,245,371	6,661,222	12,316,887	1,004,208
White: English, Welsh, Scottish, Northern Irish or British	34,553,514	6,295,968	3,538,783	4,999,825	2,073,095	6,202,490	10,632,867	810,486
White: Irish	468,870	94,600	27,921	39,690	19,529	68,934	203,278	14,918
White: Gypsy or Irish Traveller, Roma or Other White	3,169,038	520,039	214,130	232,778	152,747	389,798	1,480,742	178,804
Other ethnic group	950,648	229,524	82,825	79,045	30,255	108,599	377,180	43,220

Source: Ready Made Tables - 2021 Census - Census of Population - Data Sources - home - Nomis - Official Census and Labour Market Statistics (nomisweb.co.uk).

Note:

1. RM049 - Highest level of qualification by ethnic group, downloaded from Nomis 6 June 2023.

2.In order to protect against disclosure of personal information, records have been swapped between different geographic areas and counts perturbed by small amounts. Small counts at the lowest geographies will be most affected.