Twin Research and Human Genetics (2026), **29**, 334–342
doi:10.1017/thg.2026.10066

CAMBRIDGE UNIVERSITY PRESS

Article

Linking the Twins Early Development Study (TEDS) With NHS Electronic Health Records

Celestine Lockhart[1], Megan Skelton[1], Shannon Bristow[1], Andy Boyd[2], Johnny Downs[3,4], Alex Dregan[5], Helen L. Fisher[6,7], Ray Leal[1], Andrew McMillan[1], Rachel Ogden[1], Louise Webster[1], Robert Plomin[1] and Thalia C. Eley[1,8]

[1]Social Genetic and Developmental Psychiatry (SGDP) Centre, King's College London, UK; [2]UK Longitudinal Linkage Collaboration, University of Bristol Medical School, UK; [3]Department of Child and Adolescent Psychiatry, King's College London Institute of Psychiatry Psychology & Neuroscience, UK; [4]South London and Maudsley NHS Foundation Trust, UK; [5]Psychological Medicine, King's College London Institute of Psychiatry Psychology & Neuroscience, UK; [6]Social Genetic and Developmental Psychiatry (SGDP) Centre, King's College London Institute of Psychiatry Psychology & Neuroscience, UK; [7]ESRC Centre for Society and Mental Health, King's College London, UK and [8]National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre (BRC), South London and Maudsley NHS Trust, London, UK

Abstract

The Twins Early Development Study (TEDS) is a longitudinal population study of over 10,000 twin pairs born in England and Wales between 1994 and 1996. As the twins enter their thirties, a primary focus of TEDS is to better understand the development of common physical and mental health problems and the relationship with the different social milestones of adulthood (e.g., employment, partnerships, and/or parenthood). With over 30 years of prospectively collected questionnaire and genetic data, the study is uniquely placed to answer questions about the health challenges facing young adults today. Incorporating linked medical records with the existing research data will provide a different data perspective on our twin's health status and outcomes and support more equitable research by helping to address both response and attrition bias. This article provides an overview of the protocol to link TEDS participants to electronic health records collected by the UK National Health Service (NHS). It will outline the linkage process, characterize the available linked study sample and NHS datasets, and describe the legal basis for this work.

Keywords: Linkage; Electronic health records; Twins; TEDS; UK LLC

(Received 12 November 2025; revise received 28 January 2026; accepted 11 March 2026; First Published online 21 May 2026)

Recent years have seen a dramatic increase in the number of people experiencing common mental health conditions, especially among young women (Chen et al., 2025). Alongside this greater mental health burden, common physical health conditions such as obesity and diabetes are also rising in young people (Taylor, 2024; Xie et al., 2022). As such, there is a growing need for high-quality research to better understand who is likely to develop a common mental health condition, how this relates to the development of common physical health conditions, and how best to intervene. Data from longitudinal population studies provide a rich source of information about the development of mental health conditions across time. However, there are additional, routinely collected sources of information that may shed light on health outcomes.

In the UK, the National Health Service (NHS) provides taxpayer-funded healthcare for those living in all four UK countries. Recent figures from the Association of British Insurers indicate that only 4.7 million people across the UK are covered by private medical insurance (Association of British Insurers, 2024), meaning that the majority of the UK population utilizes the NHS for healthcare. As such, NHS data provide valuable insights into population health across the full range of geographical and socioeconomic circumstances. The adoption of electronic health records in healthcare settings has increased the possibility of using routinely collected health data for secondary research (Casey et al., 2016). Furthermore, a major government focus in recent years has been to integrate the use of electronic health records into all NHS trusts to improve care and support the development of research (Department of Health and Social Care, 2022), as well as reduce health inequalities (Chadd et al., 2025). This article outlines the process and outcome of linking the Twins Early Development Study, a UK population-based cohort study, to routine health records collected by the UK NHS.

The Twins Early Development Study (TEDS)

TEDS is a longitudinal cohort of twins born in England and Wales between 1994 and 1996 (Lockhart et al., 2023). Data have been collected at regular intervals between ages 18 months and 30 years, allowing researchers to explore the genetic and environmental contributions to individual differences in the development of

Corresponding author: Thalia C. Eley; Email: thalia.eley@kcl.ac.uk

Cite this article: Lockhart C, Skelton M, Bristow S, Boyd A, Downs J, Dregan A, Fisher HL, Leal R, McMillan A, Ogden R, Webster L, Plomin R, Eley TC. (2026) Linking the Twins Early Development Study (TEDS) With NHS Electronic Health Records. *Twin Research and Human Genetics* **29**: 334–342, https://doi.org/10.1017/thg.2026.10066

© The Author(s), 2026. Published by Cambridge University Press on behalf of International Society for Twin Studies. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Check for updates

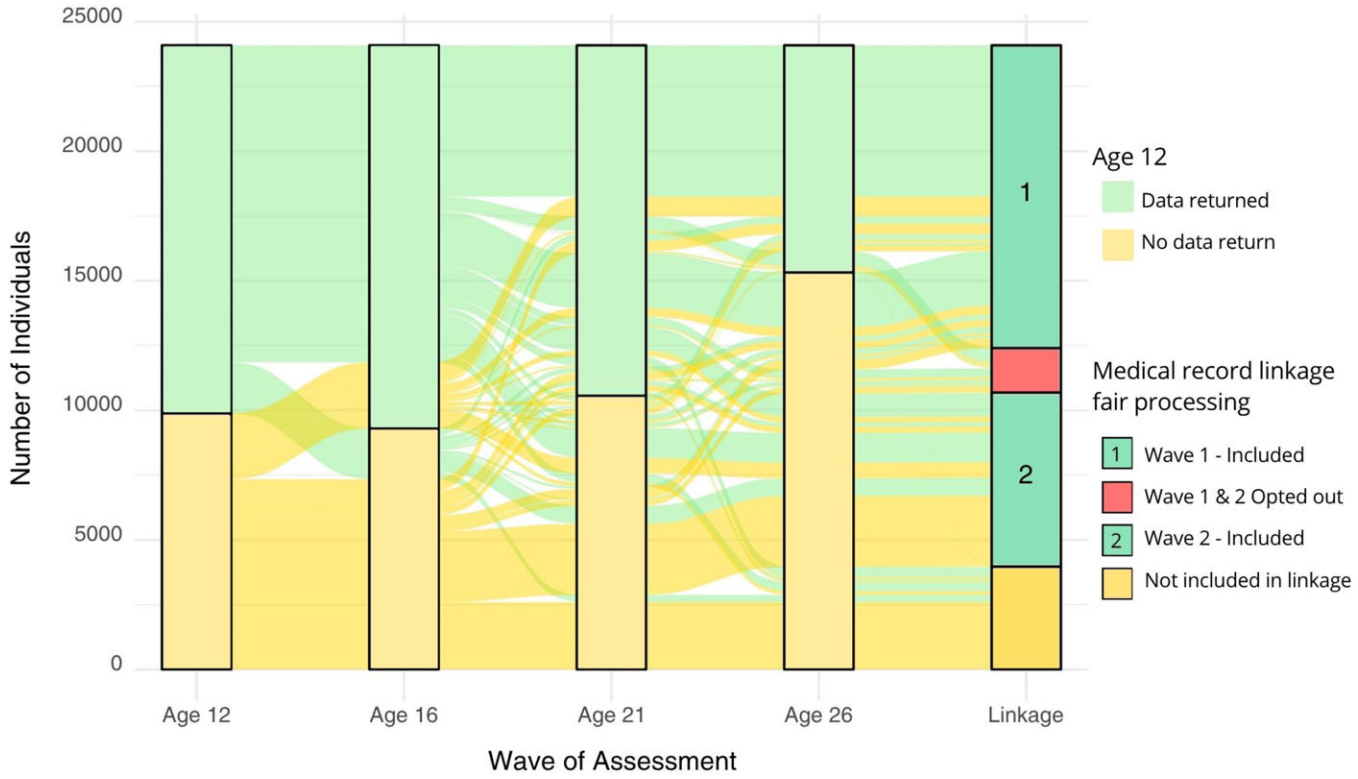


Figure 1. Returns of twin-reported data at ages 12, 16, 21 and 26.
Note: Green represents participation at each wave for participants who took part at age 12. Those not yet included in linkage (dark yellow in the linkage column) were either not traceable via the Patient Demographic Service, were traced but with an unusable address, or are part of the Environmental Risk (E-Risk) study and therefore were not sent fair processing at this time.

cognition, behavior, mental and physical health. TEDS worked with the Office for National Statistics (ONS) to identify families with twins born between January 1994 and December 1996 in England and Wales using electronic birth records. Identified families were invited to join the study, with 13,769 consenting to take part at the first data collection when their twins were 18 months (Trouton et al., 2002). Subsequently, data have been collected when twins were aged 2, 3, 4, 7, 8, 9, 10, 12, 14, 16, 18, 21, 26, and 30 years old. Information has been reported by parents, twins and teachers, with specific informants varying between waves, providing rich insight into the twins' behavior across contexts. All previous data collections have been reported in more detail elsewhere (Haworth et al., 2013; Lockhart et al., 2023; Oliver & Plomin, 2007; Rimfeld et al., 2019; Trouton et al., 2002) and on the TEDS Data Dictionary (TEDS Data Dictionary, n.d.). With participants now approximately 30 years old, TEDS is well placed to examine how early life experiences impact health and social outcomes in early adulthood.

The Benefits of Linkage

Linking to medical records increases the granularity and accuracy of health-related data by providing a wealth of new information about participants' physical and mental health conditions, access to treatment, medication dosages, and interaction with services. Used in combination with the longitudinal data collected in TEDS, these records provide new opportunities to explore the mechanisms underpinning the development of health difficulties. The linked data will also provide insight into public health, based on service use, trajectories of healthcare contact and diagnoses, alongside information about lifestyle and subclinical symptoms from longitudinal research data.

In addition to increasing data, the inclusion of medical records in research increases the fairness of cohort-based research. Not all TEDS participants contribute to every wave of data collection, but rather, many re-engage after a period of nonparticipation, leading to variability in attrition rates over time. For example, 91% of the twins who completed the assessment at age 12 responded to at least one of the subsequent waves of assessment. The varying participation between ages 12, 16, 21 and 26 of TEDS twins who have not withdrawn from the study is shown in the changes from the green to yellow sections across time in Figure 1. This inconsistency in participation can result in missing data for some participants between ages. Linking to routinely collected data is one approach to provide alternative sources of data and to assess and inform statistical approaches to address bias and wider data quality issues. As seen in Figure 1, linkage work accesses data on many participants who have missed some assessment waves or not taken part at all since age 12. In addition to those who have been included in linkage (dark green) and those who have opted out (red), the linkage column in Figure 1 also includes participants who have not been included in medical record linkage so far. This group includes participants who we could not trace through the NHS Patient Demographic Service (PDS; more detail below) or who were traced but the address provided was not usable. This group also includes individuals who are part of the current Environmental Risk (E-Risk) study (Moffitt & E-Risk Study Team, 2002) data collection but have not yet been interviewed. This group are being invited to provide consent for medical record linkage as part of that data collection.

https://doi.org/10.1017/thg.2026.10066 Published online by Cambridge University Press

Participation in longitudinal studies is easier for some groups than others. For example, in Western countries, highly educated, healthy and white individuals — reflecting a 'healthy volunteer' bias — tend to be overrepresented in research samples (Henrich et al., 2010). Furthermore, individuals facing worse health outcomes, having higher rates of behaviors associated with ill-health and complex social adversities, are more likely to be lost from follow-up waves of data collection (Cornish et al., 2021; Delfabbro et al., 2017; Watson & Wooden, 2009; Young et al., 2006). There are several reasons why research is often worse at reaching or re-engaging individuals from marginalized groups, including mistrust, perceptions that there is no benefit to their community in taking part, stigma, frequent changes in contact details or home address (Bonevski et al., 2014). As such, there are groups of individuals who are excluded from medical and public health research studies and by extension are less likely to benefit from published evidence as they may not translate across population groups. By providing data on individuals who may not have been able to take part over the years, medical record linkage increases the fairness of research. In some cases, it may also encourage re-engagement in the study for participants with whom we have lost touch (see more below about the linkage process for this group). In the long term, linkage to routinely collected records also increases participant convenience. Providing detailed information about health and treatments can be time consuming for participants. Linked records allow for such details to be accessed without lengthy questionnaires, and provide information that may be difficult to accurately recall, such as date-sequenced information on different medication regimes.

Medical Record Linkage in TEDS

Accessing medical record data from the NHS is an ongoing multistage process that has included consulting TEDS participants, securing the relevant approvals, informing study members of the research plans, and finally undertaking accurate linkage. A full timeline of the linkage program thus far is outlined in Figure 2.

Patient and Public Involvement and Engagement (PPIE) Consultations

The TEDS twins were consulted at every stage throughout the development of the medical record linkage program to ensure the research was acceptable to study participants. This was done through a mixture of in-person and online Patient and Public Involvement and Engagement (PPIE) consultations. These sessions assessed three main themes: (1) the proposed study approach and methodology, (2) the clarity and readability of the study materials, and (3) the types of research questions that would be appropriate and beneficial to investigate with linked medical record data.

To ensure that PPIE results represented as broad a group of participants as possible, we carried out consultations specifically with individuals who had been inactive in the study. Participants included in this group had been recently traced by the TEDS team and had not taken part since at least age 18, some not since age 4.

The feedback received from both active and inactive twins during the consultations was overwhelmingly positive in terms of study participants agreeing that the opt-out method was acceptable and that linkage would expand the benefit of TEDS research. Feedback from participants also improved the clarity of the study materials. For example, based on feedback that some people may struggle to process the lengthy fair processing document, the TEDS team created an animated version of the document. When asked about their views on the use of medical records in research, all participants involved in the PPIE felt it was either very important or important to explore questions relating to mental health.

Ethics and Legal Basis

Within the UK, record linkage requires a legal basis that includes either seeking explicit participant consent or using specific legislation that enables public benefit research to set aside consent requirements (subject to making all reasonable efforts to inform participants and provide a means to object). Given that a key objective of medical record linkage work is to improve research equity through capturing data on marginalized and harder-to-research participants lost to attrition, and that the same factors that drive attrition would also impact efforts to obtain consent, seeking consent was not practical in TEDS. Proceeding without consent seeks an appropriate balance between a participants' individual right to control the use of their data with the broader study aims for societal benefit.

To address this ethical issue, the NHS in England have developed a 'national data opt-out' mechanism, where NHS patients can choose to object to their NHS data being used for planning and research. Wales, Scotland and Northern Ireland do not currently have comparable national opt-out policies. However, other more localized options exist depending on the country. For example, in Wales, patients are able to opt out of their GP records being shared for planning and research purposes directly through their GP. TEDS respects all objections given directly to the study or through the national opt-out scheme. Participants can object at any point, not just when information is initially sent, and, as such, the number of participants with linked records may vary slightly over time.

TEDS was approved by the Health Research Authority Research Ethics Committee (REC; reference number: 22/SC/0313) and Confidentiality Advisory Group (CAG; reference number: 22/CAG/0038) to use an 'opt-out' method for accessing medical records, rather than a consent-based model. This opt-out objection mechanism requires studies to meet strict criteria, only being granted in circumstances where the NHS deems research to be in the public interest and obtaining consent is not practical.

Legal and Regulatory Basis for Opt-Out Method

TEDS complies with UK General Data Protection Regulation (GDPR) for holding and processing participant study data. Specifically for TEDS this is GDPR and Data Protection Act 2018 Article 6(1)(e) 'performance of a task carried out in the public interest', and where sensitive information is involved, it is Article 9(2)(j) in accordance with Article 89(1) 'scientific or historical research purposes or statistical purposes'.

However, accessing and processing personal data for linkage purposes in the UK requires Duty of Confidentiality in UK Common Law. TEDS meets this by using an opt-out methodology covered by Section 251 of the National Health Service Act 2006. The Section 251 legal basis covers the use of confidential patient information for research in the public benefit where it is not practical to obtain consent. The process of gaining Section 251 approval includes an evaluation of the feasibility of fully anonymizing the data. In the context of TEDS, the full anonymization of medical record data would render it impossible to link to existing TEDS research data and conduct ongoing participant follow-up. However, to ensure confidentiality is maintained, records are

https://doi.org/10.1017/thg.2026.10066 Published online by Cambridge University Press

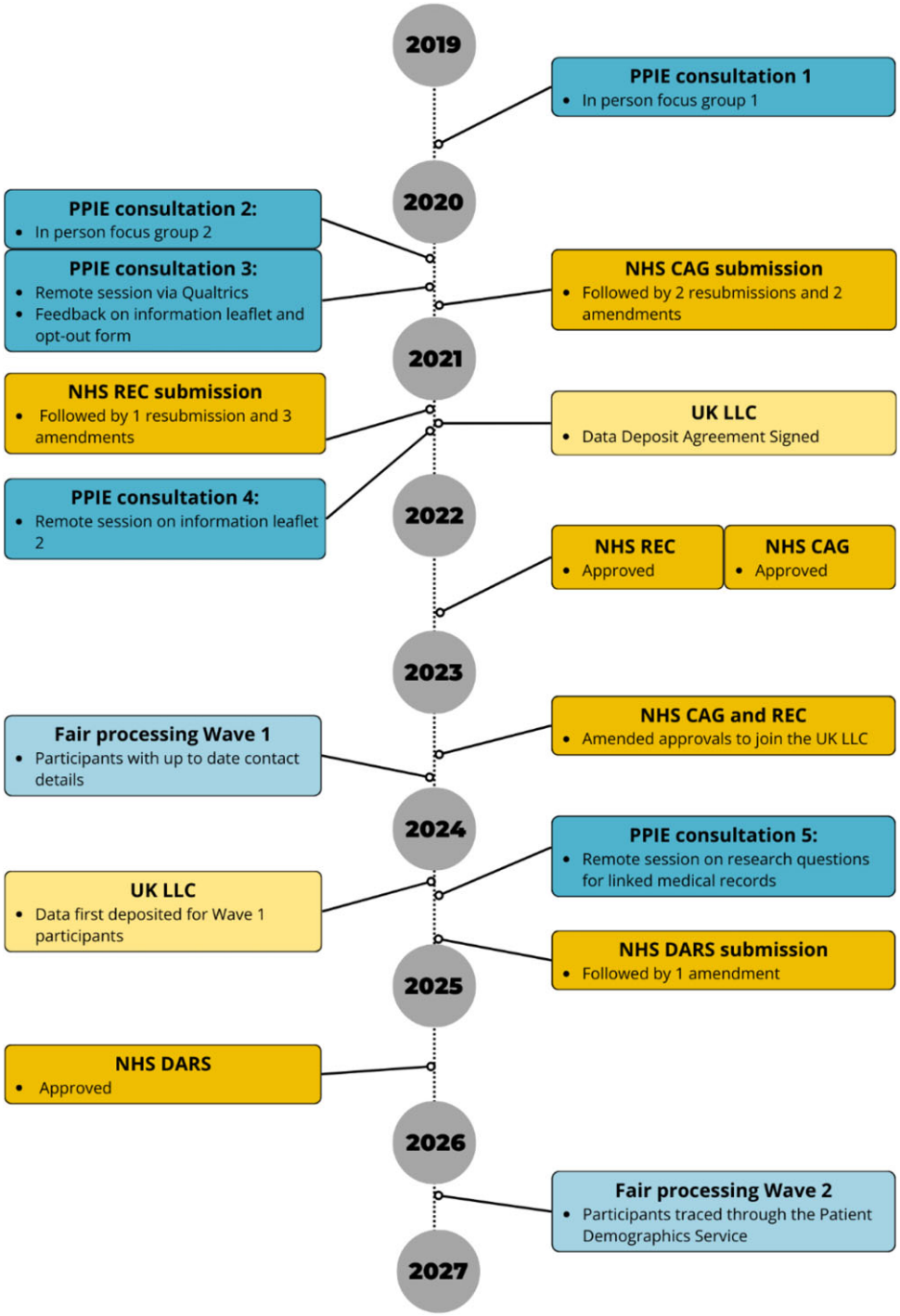


Figure 2. Timeline of medical record linkage approvals in TEDS.
Note: PPIE, Patient and public involvement and engagement; CAG, Confidentiality Advisory Group; REC, Research Ethics Committee; DARS, Data Access Request Service; UK LLC, UK Longitudinal Linkage Collaboration.

deidentified and stored within a highly controlled trusted research environment. Because Section 251 approval allows for the disclosure of confidential patient information without consent, studies must meet rigorous standards to obtain this approval.

Linkage Process

As part of the opt-out approval, it is the responsibility of the study team to provide detailed information to participants on the research plans, a process henceforth referred to as fair processing. Fair processing provides the opportunity for individuals to make an informed decision about including their data in linkage and outlines the process for opting out. To use this method, research teams must make every reasonable effort to provide fair processing information to their study participants. With this in mind, TEDS planned fair processing to be completed in two waves. Wave 1 consisted of ~11,500 participants for whom we were confident we held an up-to-date home address. These participants had completed at least one of the two most recent waves of whole cohort data collection, in which participants were asked to provide their current address. Fair processing was always sent by postal mail and also via email where email addresses were available. Additionally, any participants who had recently updated the study team with a change of address, or who had been successfully traced using

https://doi.org/10.1017/thg.2026.10066 Published online by Cambridge University Press

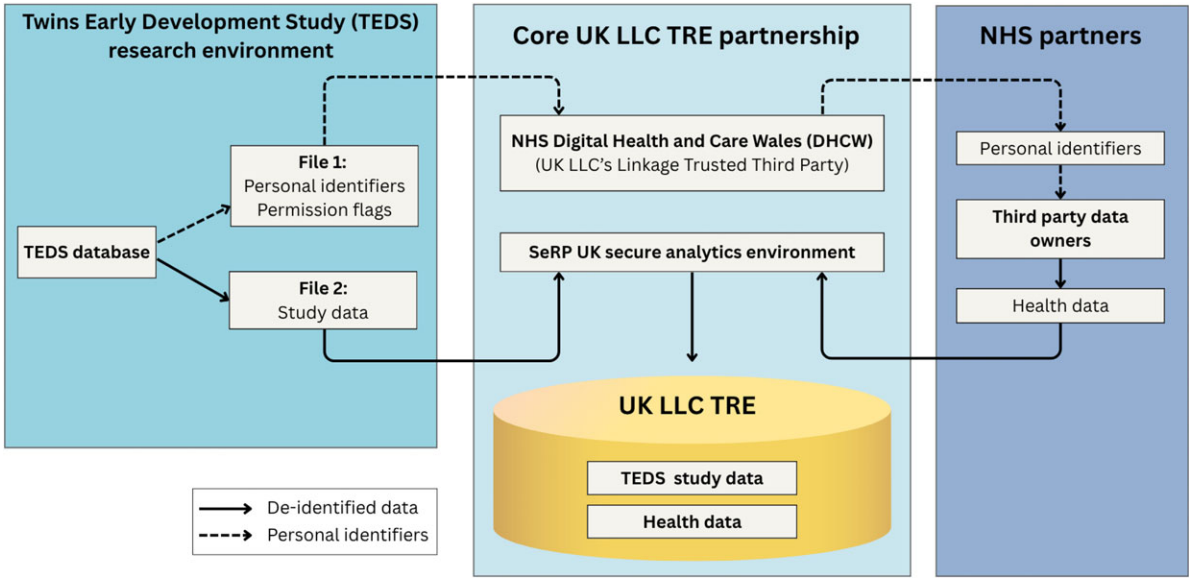


Figure 3. TEDS and UK LLC flow of participant identifiers and de-identified data.
Note: TRE, Trusted Research Environment. Figure adapted from Boyd et al., (2025).

publicly available records, were included. Subsequently, Wave 2 included participants without up-to-date contact details, many of whom had not participated in recent waves of data collection. For this group, the TEDS team gained approval from the NHS Data Access Request Service (DARS) to access these participants' current contact details held by participants' General Practice (GP) surgeries through the Patient Demographics Service (PDS). This additional step was taken to ensure that all TEDS participants had the opportunity to opt out of the medical record linkage program. As part of the fair processing campaign, every TEDS participant received the fair processing leaflet and a reminder letter approximately 2 weeks later by post. To opt out of medical record linkage, participants were asked to complete an opt-out form and return it using a prepaid postal envelope or via email. At future face-to-face data collections, either in person or online, TEDS participants will be asked for explicit consent for linkages. Consent collections are already underway for the subset of TEDS twins ($n = 2232$) included in the E-Risk Study.

To date, medical record linkage has had a positive uptake. Of those twins included in Wave 1, only 9.8% opted out of linkage activities after receiving fair processing materials. To date, the Wave 2 opt-out rate is similar at 7.7%. Of the subgroup who have been asked for consent, 85% have provided consent.

Once the fair processing process is complete (e.g., both the leaflet and reminder letter have been sent), the TEDS team sends participant identifiers and data use permission flags ('File 1') to the UK Longitudinal Linkage Collaboration (UK LLC), the national Trusted Research Environment (TRE) for the UK's longitudinal research community led by the University of Bristol and University of Edinburgh. The UK LLC is designed to support the UK's unparalleled collection of Longitudinal Population Studies (LPS) by providing record linkage and secure analysis and data curation services. The identifiers are sent directly from TEDS to the UK LLC's linkage broker, the NHS Digital Health Care Wales (DHCW), who facilitate the linkage to electronic health records through wider NHS' trusted partners across the UK's four nations. Subsequently, relevant de-identified NHS data linked with the unique TEDS identifier is transferred back into the UK LLC via their TRE infrastructure host (Secure eResearch Platform, run by Swansea University). For an extra layer of security, the associated TEDS identifier is encrypted before the data are deposited in the TRE, so that the identifier is unrecognisable. In a separate transfer, TEDS deposits relevant de-identified study data from the TEDS research database into UK LLC ('File 2'). Study data are sent alongside the same unique ID, encrypted in the same way, to ensure correct linkage with health records and future linkages. This is a two-step dynamic process with participants being added or removed as they receive fair processing or opt out. The NHS datasets are also regularly refreshed by the UK LLC data team to include updated medical records. The flow of participant identifiers and de-identified research and medical record data is outlined in Figure 3.

TEDS refreshes the identifiers and permission flags of participants in the UK LLC on a regular basis, enabling UK LLC to manage study withdrawals or objections, or to seek to establish a linkage for those who have updated their identifiers.

The protocol for TEDS linkage followed precedents set by other cohort studies, including TwinsUK (Moayyeri et al., 2013) and the Avon Longitudinal Study of Parents and Children (ALSPAC; Boyd et al., 2013).

Access to TEDS Data Using the UK LLC

The TEDS linkage program has been developed with the support of the UK Longitudinal Linkage Collaboration (Boyd et al., 2025). TEDS is now one of 22 longitudinal population studies hosting data in the UK LLC TRE. The UK LLC is a discoverable and accessible resource for public good research, that is, research that is likely to benefit a large number of people in the general population. The data available within the TRE is easy to explore through the UK LLC explore page (https://explore.ukllc.ac.uk/) and guidebook (https://guidebook.ukllc.ac.uk/docs/welcome.html).

UK-based bona fide researchers can apply to access data within the UK LLC to carry out research for public benefit through a streamlined application process that balances the needs of research studies and members of the public. Applicants must hold the Office

https://doi.org/10.1017/thg.2026.10066 Published online by Cambridge University Press

Table 1. Comparison of socio-demographic characteristics recorded at first contact for the full TEDS and linkage subsamples

	Families (*N*)	Complete twin pairs (*N*)	Female (%)	Monozygotic (%)	White ethnicity (%)	Mothers with A-levels or higher (%)	Fathers with A-levels or higher (%)	Mothers employed (%)	Fathers employed (%)
Full sample	13,722	13,722	50.1	33.4	91.9	37.4	45.4	43.1	91.6
Linkage sample – Wave 1	6953	4687	58.6	36.5	93.6	44.1	50.6	45.7	92.9
Linkage sample – Wave 2	4351	2211	39.6	29.9	91.4	32.4	39.6	42.7	91.3

Note: All twin and family socio-demographic characteristics reported in this table are based on parent-reported questionnaire data at first contact. The standard analytical exclusions have not been applied as these may not be relevant for analyses using linked data. The Wave 1 demographics have been calculated based on participants included in Wave 1 following fair processing and opt-outs. Wave 2 demographics have been calculated based on participants included in fair processing Wave 2 who have not opted out. The *N*s and demographics of the final linked sample in the UK LLC will differ slightly depending on the success rate of linkage to medical records.

of National Statistics (ONS) Accredited Researcher status prior to accessing any data, be based in the UK, and employed at an organization with sufficient capability to support the research. The application process includes completing an initial expression of interest form, which is reviewed by the UK LLC team to assess project feasibility. If approved, applicants are invited to submit a full application form that is reviewed by the UK LLC team, UK LLC Public Advisory Group, as well as relevant study teams (i.e., TEDS review applications to use TEDS data). All approved research projects are publicly catalogued in the UK LLC Data Use Register (https://ukllc.ac.uk/data-use-register).

Linkage Sample in TEDS

Record linkage to the NHS England centralized database is conducted by NHS England using a series of deterministic linkage algorithms (All linked data copyright ©2026, NHS England. Re-used with the permission of NHS England and UK LLC. All rights reserved). The first algorithm identifies exact, unique matches between all study identifiers (first name, family name, date of birth, address) and the equivalent identifiers in the NHS patient register. Where no exact match is found, a series of subsequent algorithms are applied which relax part of the criteria (e.g., only the first letter of the first name and exact matches of all other identifiers). Post linkage, the national data opt-out is checked and participants who have objected are removed. The current overall linkage rate between TEDS and the patient register is 78%. Once linked to the patient register, internal and consistent NHS identifiers are used to identify and extract the records of participants from a wide range of electronic health records databases.

Linkage failure may occur for a number of reasons. These include participants who have requested the national opt-out or who do not reside or seek care in England, errors in recorded information, changes in personal identifiers over time (e.g., name, gender, or registered address) or an insufficient number of available identifiers (Doidge et al., 2021). We note that linkage is especially challenging for twins as twin pairs often share a wide range of identifiers (e.g., social naming patterns in the UK mean they often share the first letter of their first name, as well as family name, date of birth and address). TEDS does not hold the NHS numbers of the study participants, which may partially explain the 22% of participant

s who were not successfully linked. NHS England does not provide a breakdown of reasons for linkage failure, as information on national opt-out is considered sensitive and not sharable where the opt-out is applied. Furthermore, no metrics are provided on the linkage quality (degree of matching identifiers) of the linked or nonlinked samples. Notably, although the UK LLC has approval in place for records held by both NHS Scotland and Wales, these data have not yet been deposited in the TRE. The availability of these records will likely increase the TEDS linkage rate, particularly as ~5% of the TEDS sample reside in Wales.

In terms of socio-demographic characteristics collected at first contact, the Wave 1 linkage sample has higher levels of female and monozygotic twins than the TEDS sample at first contact (Table 1). Twins included in the Wave 1 linkage sample also have parents with slightly higher levels of qualifications and employment relative to the full first contact sample. However, the Wave 1 linkage sample only includes twins who have completed a recent wave of data collection. As such, the higher rates of female and monozygotic twins and a skew towards higher family SES likely reflect participation biases seen in volunteer samples. In comparison, participants in Wave 2 have lower levels of monozygotic twin pairs, female sex, and family-level SES indicators than the full first contact sample. This indicates that the inclusion of Wave 2 participants in the linked sample helps to reduce biases arising from sample attrition. Of those who we hoped to include in Wave 2, 73.3% were successfully traced through the NHS Patient Demographic Service. Of those successfully traced, 9.5% had unusable addresses. Overall, ~7200 participants with whom we had lost touch were sent fair processing in Wave 2.

Available NHS Datasets Linked in the UK LLC

So far, the routine health datasets linked to TEDS participants include data held by NHS England. The full list of available datasets, the number of TEDS participants included, and temporal coverage is presented in Table 2.

Primary Care Datasets

Records are available from two primary care datasets. The Primary Care Medicines dataset includes information on medicines dispensed in the community resulting from prescriptions issued by GPs, clinics, dentists, and community nursing services. The General Practice Extraction Service Data for Pandemic Planning and Research (GDPPR) dataset includes a more extensive list of clinical codes relating to health conditions relevant to pandemic planning and research. These datasets have coverage across England. Projects using the GDPPR dataset are legally restricted to

https://doi.org/10.1017/thg.2026.10066 Published online by Cambridge University Press

Table 2. NHS datasets available in the UK Longitudinal Linkage Collaboration

Domain	NHS datasets	Temporal coverage	TEDS participant count
Primary care	General Practice Extraction Service (GPES) Data for Pandemic Planning and Research (GDPPR)	2020 – ongoing	8011
	Primary Care Medicines (PCM)	2015 – ongoing	6905
Hospital	HES Outpatients (HESOP)	2003 – ongoing	2268
	HES Admitted Patient Care (HESAPC)	1998 – ongoing	5530
	HES Maternity (HESAPC_MAT)	2000 – ongoing	1414
	HES Augmented Care Periods (HESAPC_ACP)	1997 – 2006	34
	HES Critical Care (HESCC)	2008 – ongoing	80
	Emergency Care Data Set (ECDS)	2017 – ongoing	3666
	HES Accident & Emergency (HESAE)	2007 – 2020	6348
Registration	Cancer Registration (CANCER)	1971 – ongoing	84
	Demographics	2004 – ongoing	8647
	Civil Registrations of Death (MORTALITY)	1993 – ongoing	<10
Mental health	Mental Health Services Data Set (MHSDS)	2006 – ongoing	1065
	Talking Therapies for Anxiety and Depression	2012 – ongoing	1309
Community	Community Services Data Set (CSDS)	2015 – ongoing	1688
COVID-19	COVID-19 Second Generation Surveillance System (COVIDSGSS)	2020 – ongoing	3471
	Enzyme-linked Immunosorbent Assay (IELISA)	2020 – ongoing	191
	National Pathology Exchange (NPEX)	2020 – ongoing	6892
	Severe Acute Respiratory Infection (SARI-Watch) surveillance system	2020 – 2022	0
	COVID-19 Vaccination Status (CVS)	2020 – ongoing	7236
	COVID-19 Vaccination Adverse Reactions (CVAR)	2020 – ongoing	0

Note: Some NHS datasets (MHSDS, Talking Therapies for Anxiety and Depression, CSDS) contain multiple data frames, some of which contain a smaller number of participants. More specific information is available in the UK LLC Guidebook (Welcome to UK LLC Guidebook — UK LLC Dataset Documentation, n.d.). Data Copyright © (2025), NHS England. Data reused with the permission of NHS England. All rights reserved.

informing COVID-19 relevant research and must not be used for any General Practice performance management.

Hospital Datasets

TEDS participants are linked to all seven of the NHS England hospital datasets included in the UK LLC. The three main hospital datasets are Outpatient care, Admitted Patient care, and Emergency care. Outpatient covers all hospital appointments that do not require an overnight hospital stay, including referrals, waiting times, and diagnoses. Admitted Patient care covers the same for inpatient care, as well as datasets relating to Augmented Care Periods, Maternity Care, and Critical Care. There are two emergency care datasets; the Accident and Emergency dataset was retired in 2020 and replaced by the Emergency Care Data Set. Both emergency datasets include information on presenting symptoms, outcomes, waiting times, diagnosis and treatments. Together these hospital datasets include records from 1998 onwards, covering the majority of the TEDS twins' lives but excluding birth and early-year data.

Mental Health Datasets

There are data available from two NHS England mental health datasets: Mental Health Services Data Set (MHSDS) and the NHS Talking Therapies for Anxiety and Depression. MHSDS includes data on individuals who have received treatment for mental health, wellbeing, learning disability, autism or other neurodevelopmental conditions in an NHS-funded secondary care service.

The NHS Talking Therapies dataset (previously known as Improving Access to Psychological Therapies) includes information about adults who have received NHS-funded psychological treatment for anxiety, depression or related disorders.

Community Datasets

The Community Services Data Set (CSDS) and Maternity Services Data Set (MSDS) are the two community datasets available for use. The CSDS includes data on a wide range of personal, demographic, social, and medical circumstances recorded in any community setting (e.g., health centers, schools, community centers). The MSDS dataset contains pregnancy and birth information for individuals who have accessed maternity care through the NHS. This will exclude data on the TEDS participants' own birth but will provide information on their pregnancies and the birth of their children, some of whom participate in the Children of TEDS Study (Ahmadzadeh *et al.*, 2019).

COVID-19 Datasets

The UK LLC hosts electronic records from six NHS England COVID-19 datasets relating to testing, vaccinations, and COVID-related hospitalizations. These datasets all began in 2020; however, some are no

https://doi.org/10.1017/thg.2026.10066 Published online by Cambridge University Press

longer collecting new information. As outlined by the NHS terms of use, projects must be related to COVID-19 in order to access these data.

Registration Datasets

TEDS participants have linked data for the Cancer, Civil Registrations of Death, and Demographics registers. The Cancer Registration includes information on cancer diagnosis and treatment. The Civil Registrations of Death contains information collected by the Office for National Statistics on deaths reported to registry offices. The Demographics dataset outlines information such as name, date of birth, and sex.

Cross-Cohort Analysis

UK LLC hosts the TEDS collected data and linked records in an integrated relational database. The TEDS data sits alongside data from over 20 other major longitudinal studies and a pooled sample of more than 400,000 participants. This provides users of TEDS data with a straightforward mechanism to either run replication analysis in other cohorts or to combine studies for meta- or pooled-analysis. While these latter models are complicated by the need to harmonize study collected data, they are facilitated by the linked data being identical in structure and methodology across all studies. This potential is further enhanced by TEDS (born 1994–1996) sitting alongside: the 'Next Steps' cohort (born 1990–1991); the Avon Longitudinal Study of Parents and Children (born 1991–1992); and (UK) Millennium Cohort Study (born 2001–2002) which currently collectively have a combined sample size of >40,000 in the LLC and which can be used to understand outcomes across diverse UK Millennial and Generation Z populations.

Conclusion

Linkages to electronic health records pose a number of challenges relating to protocol development, access and ethical approvals, data quality, analysis platforms, and generalizability of findings (Honeyford et al., 2022). Indeed, the road to linkage in the TEDS cohort has been long and winding. Perhaps the greatest challenge lay in developing a protocol that was acceptable to TEDS participants, enabled both expression of dissent and wider fairness of outcomes, and maintained long-term participant trust. Our extensive participant consultation was specifically designed to inform this decision-making process and provided confidence that an opt-out approach could be used without negatively impacting participant relationships.

Integrating medical records and longitudinal population study data has the potential to offer powerful insights into public health. The NHS is facing both longstanding and emerging challenges (Khan, 2023), and the wealth of data collected by longitudinal population studies such as TEDS will allow researchers to address many critical and timely health related research questions. For example, prediction models of physical and mental health treatment outcomes can be built using TEDS data, incorporating a broader range of longitudinal information than is available in medical records alone. The twin structure and genomic data available in TEDS could also answer novel questions, such as about genetic influences on service use. Cross-cohort analyses facilitated by the UK LLC's collaborative research environment will enable the investigation of rare and uncommon disorders. The collaboration between longitudinal population studies and integrated platforms such as the UK LLC make medical record linkage more feasible. The integration of rich, detailed longitudinal population study data with real-world data from health services used by the majority of the UK population provides a unique opportunity for TEDS to contribute to highly impactful research that will improve our understanding of health and wellbeing for many years to come.

Data availability. Access to TEDS linked data is governed by application directly to the UK LLC (https://ukllc.ac.uk/apply).

Acknowledgments. The authors gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. For the purposes of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Accepted Author Manuscript version arising from this submission. The views expressed are those of the authors and not necessarily those of the ESRC, MRC or King's College London.

This work uses data accessed within UK LLC's Trusted Research Environment (TRE). We thank the Secure eResearch Platform (SeRP UK) team at Swansea University and NHS Digital Health and Care Wales for providing the TRE's infrastructure and support. This work uses data provided by participants of the contributing LPS, which have been collected through their LPS or as part of their care and support and/or interactions with UK government services. We wish to recognise and thank the participants and each contributing to the LPS team. We thank the following LPS for contributing data that made this research possible: TEDS - Twins Early Development Study. This work uses data provided by patients and collected by the NHS as part of their care and support. NHS England collates patient data and gives permission for publicly beneficial uses via its Data Access Service.

Financial support. TEDS is supported by the UK Medical Research Council (MRC) (MR/V012878/1 and previously MR/M021475/1). The E-Risk substudy is also supported by the MRC (MR/X010791/1). H.L.F. receives salary support from the Economic and Social Research Council (ESRC) Centre for Society and Mental Health at King's College London [ES/S012567/1]; [UKRI861]. AB is funded by UK Research & Innovation (Medical Research Council MR/X021556/1, Economic and Social Research Council ES/X000567/1). TE is part-funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre (BRC).

Competing interests. None to declare.

Ethical statement. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. This project has been approved by UK LLC and its contributing data owners. Information about this project and its outputs can be accessed via UK LLC's Data Use Register. UK LLC has ethical approval from the Health Research Authority Research Ethics Committee (Haydock Committee; ref: 25/NW/0314).

References

Ahmadzadeh, Y. I., Eley, T. C., Plomin, R., Dale, P. S., Lester, K. J., Oliver, B. R., McMillan, A., & McAdams, T. A. (2019). Children of the Twins Early Development Study (CoTEDS): A Children-of-Twins Study. Twin research and human genetics : the official journal of the *International Society for Twin Studies*, *22*(6), 514–522. https://doi.org/10.1017/thg.2019.61

Association of British Insurers. (2024, November 6). Private medical insurers report record 4.7 million covered by employer health schemes. https://www.abi.org.uk/news/news-articles/2024/112/private-medical-insurance-data-2023/

Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., Brozek, I., & Hughes, C. (2014). Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Medical Research Methodology*, *14*, 42. https://doi.org/10.1186/1471-2288-14-42

Boyd, A., Evans, K. M., Turner, E. L., Flaig, R., Oakley, J., Campbell, K. C., Thomas, R., McLachlan, S., Crane, M., Whitehorn, R., Calkin, R., Hill, A., Berman, S., Ford, D., Tobin, M., Porteous, D., Gomes, D. F., Garcia, M.

P., Wong, A., . . . **Chaturvedi, N.** (2025). UK Longitudinal Linkage Collaboration (UK LLC): The national trusted research environment for longitudinal research. *International Journal for Population Data Science*, *10*, 2468. https://doi.org/10.23889/ijpds.v10i1.2468

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., & Davey Smith, G. (2013). Cohort profile: The 'Children of the 90s' — the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, *42*, 111–127. https://doi.org/10.1093/ije/dys064

Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016). Using electronic health records for population health research: A review of methods and applications. *Annual Review of Public Health*, *37*, 61–81. https://doi.org/10.1146/annurev-publhealth-032315-021353

Chadd, K., Caute, A., Pettican, A., & Enderby, P. (2025). Operationalising routinely collected patient data in research to further the pursuit of social justice and health equity: a team-based scoping review. *BMC Medical Research Methodology*, *25*, 14. https://doi.org/10.1186/s12874-025-02466-9

Chen, S., Huang, W., Zhang, M., Song, Y., Zhao, C., Sun, H., Wang, Y., Wang, J., Sun, Y., Zhou, L., Zhu, Y., Wang, H., Xu, Z., Bai, Y., & Chang, C. (2025). Dynamic changes and future trend predictions of the global burden of anxiety disorders: analysis of 204 countries and regions from 1990 to 2021 and the impact of the COVID-19 pandemic. *EClinicalMedicine*, *79*, 103014. https://doi.org/10.1016/j.eclinm.2024.103014

Cornish, R. P., Macleod, J., Boyd, A., & Tilling, K. (2021). Factors associated with participation over time in the Avon Longitudinal Study of Parents and Children: A study using linked education and primary care data. *International Journal of Epidemiology*, *50*, 293–302. https://doi.org/10.1093/ije/dyaa192

Delfabbro, P., Winefield, H., Winefield, A., Malvaso, C., & Plueckhahn, T. (2017). Factors associated with attrition in a 10-year longitudinal study of young people: Implications for studies of employment in school leavers: Factors influencing attrition. *Australian Psychologist*, *52*, 41–51. https://doi.org/10.1111/ap.12207

Department of Health and Social Care. (2022). A plan for digital health and social care. https://www.gov.uk/government/publications/a-plan-for-digital-health-and-social-care/a-plan-for-digital-health-and-social-care

Doidge, J., Christen, P., & Harron, K. (2021). Quality assessment in data linkage. https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/quality-assessment-in-data-linkage

Haworth, C. M. A., Davis, O. S. P., & Plomin, R. (2013). Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics*, *16*, 117–125. https://doi.org/10.1017/thg.2012.91

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Honeyford, K., Expert, P., Mendelsohn, E. E., Post, B., Faisal, A. A., Glampson, B., Mayer, E. K., & Costelloe, C. E. (2022). Challenges and recommendations for high quality research using electronic health records. *Frontiers in Digital Health*, *4*, 940330. https://doi.org/10.3389/fdgth.2022.940330

Khan, Z. (2023). The emerging challenges and strengths of the National Health Services: A physician perspective. *Cureus*, *15*, e38617. https://doi.org/10.7759/cureus.38617

Lockhart, C., Bright, J., Ahmadzadeh, Y., Breen, G., Bristow, S., Boyd, A., Downs, J., Hotopf, M., Palaiologou, E., Rimfeld, K., Maxwell, J., Malanchini, M., McAdams, T. A., McMillan, A., Plomin, R., & Eley, T. C. (2023). Twins Early Development Study (TEDS): A genetically sensitive investigation of mental health outcomes in the mid-twenties. *JCPP Advances*, *3*, e12154. https://doi.org/10.1002/jcv2.12154

Moayyeri, A., Hammond, C. J., Hart, D. J., & Spector, T. D. (2013). The UK Adult Twin Registry (TwinsUK Resource). *Twin Research and Human Genetics*, *16*, 144–149. https://doi.org/10.1017/thg.2012.89

Moffitt, T. E., & **E-Risk Study Team**. (2002). Teen-aged mothers in contemporary Britain. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *43*, 727–742. https://doi.org/10.1111/1469-7610.00082

Oliver, B. R., & Plomin, R. (2007). Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Research and Human Genetics*, *10*, 96–105. https://doi.org/10.1375/twin.10.1.96

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., Dale, P. S., Eley, T. C., & Plomin, R. (2019). Twins Early Development Study: A genetically sensitive investigation into behavioral and cognitive development from infancy to emerging adulthood. *Twin Research and Human Genetics*, *22*, 508–513. https://doi.org/10.1017/thg.2019.56

Taylor, L. (2024). One in eight people globally are obese, with rates in children increasing fourfold in three decades. *BMJ*, *384*, q527. https://doi.org/10.1136/bmj.q527

TEDS Data Dictionary. (n.d.). https://datadictionary.teds.ac.uk/

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins Early Development Study (TEDS): A Multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research*, *5*, 444–448. https://doi.org/10.1375/136905202320906255

Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 157–182). John Wiley & Sons.

Welcome to UK LLC Guidebook — UK LLC Dataset Documentation. (n.d.). https://guidebook.ukllc.ac.uk/docs/welcome.html

Xie, J., Wang, M., Long, Z., Ning, H., Li, J., Cao, Y., Liao, Y., Liu, G., Wang, F., & Pan, A. (2022). Global burden of type 2 diabetes in adolescents and young adults, 1990-2019: Systematic analysis of the Global Burden of Disease Study 2019. *BMJ*, *379*, e072385. https://doi.org/10.1136/bmj-2022-072385

Young, A. F., Powers, J. R., & Bell, S. L. (2006). Attrition in longitudinal studies: who do you lose? *Australian and New Zealand Journal of Public Health*, *30*, 353–361. https://doi.org/10.1111/j.1467-842x.2006.tb00849.x

https://doi.org/10.1017/thg.2026.10066 Published online by Cambridge University Press