

Linking pre- and post-adoption records for research in anonymised form in a data safe haven: legal implications and support for a social licence

Kerina Jones¹, Sharon Heys¹, Rachel Thompson¹, David Ford¹, Lucy Griffiths¹, Rhodri Johnson¹, Alex Lee¹, and Karen Broadhurst²

Submission History

Submitted:	02/06/2020
Accepted:	22/09/2020
Published:	22/10/2020

¹Population Data Science, Swansea University Medical School, Swansea, SA2 8PP, UK

²Centre for Child & Family Justice Research, Lancaster University, Lancaster, LA1 4YW, UK

Abstract

Introduction

The long-term health and wellbeing of adoptees is under-researched. One reason for this has been limited data accessibility regarding the adoption process, and another is a practice common in some UK jurisdictions of changing the National Health Service (NHS) number (or equivalent) at adoption, as part of creating the new identity. The SAIL Databank holds data on child and family court cases from Cafcass Cymru, together with children's social care data, and can link these with routine health and administrative data in anonymised form. However, because the linkage key at SAIL is based on an encryption of the NHS number, working with pre- and post-adoption records for longitudinal research remains a major challenge. We set out to explore the legal implications of, and social support for, linking these records for use in anonymised form for longitudinal research.

Methods

We reviewed the main legislation and regulations governing the use of data about adoptees in England and Wales. We gauged support for a social licence in Wales by carrying out interviews with individuals who had been involved in the adoptions process, and by engaging with general public groups for their views. We drew out the main emerging themes and, in combination with the review, propose a way forward.

Results

The legal review indicated that there are provisions in the Family Procedure Rules (England and Wales) and the General Data Protection Regulation that can be relied upon for the lawful processing of adoption data into anonymised form for research. The main points of concern about linking pre- and post-adoption records were privacy, data security, the need to limit the number of organisations involved in data sharing, and re-identification risk. The over-riding message was favourable with longitudinal research seen as strongly beneficial.

Conclusion

This study has indicated that in Wales, there is no legal impediment, nor major objection from individuals involved in the adoptions process, or members of the general public, for the use of adoption data in anonymised form in a data safe haven. This includes the linkage of pre- and post-adoption records to enable novel longitudinal research to take place. The provisos were that robust safeguards must be in place, and that the research should aim to benefit adoptees and to improve policy and practice. We conclude that it is reasonable to proceed with caution to develop practical ways to link pre- and post-adoption records in a data safe haven.

Keywords

public engagement; legal issues; adoptions data

*Corresponding Author:

Email Address: K.H.Jones@Swansea.ac.uk (Kerina Jones)

Introduction

Adoptees often experience poor mental and physical health and this can lead to sub-optimal life outcomes and greater risk of difficulties projecting into adult life, particularly as in many adoption cases, there will have been some form of abuse or neglect [1–4]. Not enough is known about the development of these difficulties and what preventive, mitigative and support measures could be put in place. To date, the majority of published research has focused on the stability of adoption placements using representative or sub-samples of records [5], with a dearth of studies providing a longer or more holistic understanding of children's adoption journeys and life chances. Research on Welsh adoptees is particularly limited. Because of due sensitivities and risks that can be connected to the need for adoption, there are strict rules around the use of data about adoptees to safeguard their identities.

Until recently, there has been a lack of individual-level data about children involved in the adoption process available for research, but this has changed in the UK with the establishment of the Nuffield Family Justice Observatory (Nuffield FJO) and the associated Family Justice Data Partnership, which is a collaboration between the universities of Lancaster and Swansea [6]. The Nuffield FJO was set up in 2019 based on the results of a detailed scoping study [7]. It aims to support better outcomes for children and families in the family justice system in England and Wales by improving the use of data and research evidence in decision-making [8]. Correspondingly, the Data Partnership aims to effect step-change in the use of core family justice administrative datasets, to supply timely, accessible outputs to the Nuffield FJO and its range of audiences [6].

Towards this aim, the Child and Family Court Advisory Support Services (Cafcass) in England and Cafcass Cymru in Wales have each made their case data available for research in anonymised form in a data safe haven. The Cafcass datasets include the nature of the case, the child/ren and families involved, the local authority making the application to initiate court proceedings, and the legal outputs. In addition, social care datasets provided to SAIL contain further information such as adoption placements. It is beyond the scope of this paper to describe the datasets in detail but they are the subject of other publications [9, 10]. The acquisition of these different datasets enables, for the first time, the scope of these datasets to be examined with respect to their value for longitudinal research about adoptees, as well as the feasibility of linking pre- and post-adoption records. However, a necessary first step is to establish the legal implications and whether stakeholders support this work.

SAIL is a national data safe haven of de-identified health and administrative datasets about the population of Wales. Data extracts are made available for research in anonymised form on a secure platform (SeRP UK), subject to control measures and approvals [11]. For children involved in family court proceedings in Wales, Cafcass and children's social care data can immediately be linked across a variety of datasets held in SAIL, such as primary and secondary care health data and education records [12]. This means that important research questions can begin to be addressed, such as associations with particular health conditions, educational outcomes, housing arrangements and indicators of deprivation. Acquisition of the

Cafcass datasets and their deposit in the SAIL Databank has been heralded as a major accomplishment [13]; it has provided the opportunity in Wales for exemplar linkage studies, and will allow similar linkage of data from England as datasets accrue.

The Data Platform research programme in SAIL includes a range of projects relating to all family court cases. However, there is a particular challenge to research about adoptees, and particularly, longitudinal research on pre- and post-adoption records. This is because it is common practice in England and Wales for an adoptee to be allocated a new National Health Service (NHS) number at the finalisation of the adoption process. The NHS number is the unique number allocated to each user of the public sector health service in England and Wales to enable correct identification and matching of records across health providers. The new NHS number becomes part of the adoptee's new identity, but it means that there is a loss of continuity between health records pre- and post-adoption. Discussions with adoption policymakers and practitioners¹ during the course of this study indicate there is meant to be transfer of all relevant details to the new care setting, but this can be patchy, leading to an information deficit for care and research.

Coupled with this, in common with other organisations that make data available for research, SAIL relies on the NHS number (or equivalent numerical identifier in other jurisdictions) for the creation of a reliable linkage key [14]. This is the case even for datasets that do not arise from a health setting and therefore do not contain an NHS number. This is because in the matching process, personal details (name, postcode, date of birth, sex) are compared to an administrative register containing the same details plus NHS number. This allows an NHS number to be assigned to the relevant record and encrypted to create the linkage key [14]. The routine change of name and address for adoptees, from pre- to post-adoption, impedes this process. Consequently, being able to study the pre- and post-adoption status of adoptees can be difficult as there is no straightforward link between the records before and after adoption.

Addressing this obstacle is urgent, as without the ability to link pre- and post-adoption records, the full benefit of the recently acquired Cafcass Cymru and children's social care data in association with health and administrative data will not be achieved. Being able to progress in this area of research stands to add valuable new knowledge about the lives of adoptees but also enables comparisons to be drawn between children who are adopted, and those who return to family or stay in foster care. How children fare over time according to their permanency placement (i.e. where they are placed to live long-term), is one of the major concerns of those making highly consequential decisions for children in the family courts.

In this paper, we focus mainly on the situation in Wales since we are already able to link data about Welsh adoptees to health and administrative datasets in the SAIL Databank. However, the findings will be relevant to England and other countries that change the NHS number (or their equivalent identity marker) at adoption. Cafcass in Wales is referred to as Cafcass Cymru and is part of Welsh Government. The implications of changing or retaining the NHS number at adoption are out of scope and will be the subject of another article.

¹Please see the acknowledgements section

Main aim

The main aim of this study was to explore the legal and social implications of linking pre- and post-adoption records for use in anonymised form for longitudinal research in a data safe haven, with a view to recommending an acceptable way forward in relation to data linkage.

Methods

Legislation and regulations

The main relevant general data protection legislation (common to England and Wales) are the *UK Data Protection Act* (DPA, 2018) [15] and the EU General Data Protection Regulation (GDPR, 2016) [16]. Adoption is a legal process that transfers all legal rights and responsibilities relating to a child from the birth parents to approved adopters. The primary legal provisions for this process are contained in the *Adoption and Children Act 2002* [17]. A solicitor² employed by Swansea University specialising in data protection law conducted this desk-based documentation review. It was based on the following questions, which were devised and agreed by the solicitor and the lead author:

1. Is there an absolute prohibition on the use of identifiable adoption data?
2. Is there a prohibition on re-identification of an adoptee who has assumed a new identity?
3. Given the advice in relation to questions 1 and 2, how does this apply to the intended research for scientific purposes in the public benefit?
4. How does this apply within the context of using anonymised data in a data safe haven? (This question uses the SAIL Databank as a case in point.)
5. Will the existing controls within a data safe haven (using the SAIL information governance model as a working example) suffice in respect of questions 1–4, and is it therefore safe to proceed with this work?

Engagement

Using convenience sampling, we carried out interviews with a range of individuals who had been involved in the adoptions process. They included adults who had been adopted (adoptees, N=6), parents who had adopted a child or children (adoptive parents, N=10), and birth parents whose child/ren had been adopted (birth parents, N=3). Participants were contacted via support groups and their coordinators. The conversations with adoptees and birth parents were face-to-face and were recorded in note form. The interviews with adoptive parents were conducted by phone and audio recorded with the permission of the participant, as well as being noted during the interview. We did not transcribe the recordings but chose to replay them as needed to retain the tone and emphasis of responses in the thematic analysis.

²On the roll of the Law Society as a Solicitor of the Senior Courts of England and Wales

We also engaged with general public groups (without involvement in the adoptions process) to gain their views for comparison. These were members (N=9) of the SAIL Consumer Panel [18] and a young parents' support group contacted via the coordinator (N=6). The Panel regularly provides a general public view on the use of data for research. We chose the latter group to add younger participants because the Panel members tend to be in the over 50s age groups. Even so, we do not claim that the participants are representative of the population in Wales. The public groups were facilitated by the lead author, and participants provided their views verbally and on written sheets.

Before taking part, potential participants were provided with a written introduction to the study by email. If they agreed to take part, they were sent a participant information sheet and consent form. The format of the interviews and group engagements followed a standard schedule, except that some slight adaptations were made depending on the knowledge base and background of the participants. For example, the Consumer Panel members knew more about data linkage, whereas adoptive parents knew more about the adoptions process than the general public. The schedule began with a description of the study, and its purpose being to gain viewpoints on the use of data about adoptees for research in anonymised form, reiterating and building on the information provided in advance. Although such data will be used for research in anonymised form, the social licence is important for good data ethics, openness and transparency. Participants were introduced to the work of the Nuffield FJO and the associated Data Partnership, along with the work of Cafcass Cymru, the nature of the dataset and examples of how it could be used in research. Examples of routine health and administrative datasets, such as general practice, in-patient, social care and school attainment were provided, with a brief description of how data can be linked at the individual level. Participants were also informed about the need to safeguard the identity of adoptees, possible risks due to disclosure, and the need to follow proper data governance in proposed data uses.

Participants were encouraged to ask questions at any point, and they were reassured that they could discontinue the engagement at any time, including to withdraw from the study completely if they wished. The personal details requested from each person were minimal to provide some context, but also to avoid the risk of identity disclosure. These were sex (male/female/other/prefer not to say), age (in 10 year bands), whether they had personal involvement with the adoptions process, and (if they had) its nature (adoptee, adoptive parent, birth parent or professional). It was possible for an individual to have been involved in more than one capacity. All the engagement was conducted with adults living in Wales and there were 34 participants in total.

The main questions of interest were:

1. How do you feel about research using anonymised linked family court records and health / administrative data taking place?
2. What are your views on the acceptability of anonymously linking pre- and post-adoption records of someone who has been assigned a new identity during family court proceedings?

Between questions 1 and 2, participants were advised about the common practice of changing the NHS number at adoption, the purpose of this being safeguarding, and the challenge this creates for linking pre- and post-adoption records when the linkage key is based on an encryption of the NHS number.

Results

Legislation and regulations

The review of the legislation and regulations yielded the following information about data pertaining to adoptees:

1. Is there an absolute prohibition on the use of identifiable adoption data?

The *Adoption and Children Act 2002* is clear that where an adoption order has been made, the new identity of the child should not be compromised. The only lawful means of undertaking this type of processing would be where explicit informed consent has been provided by all parties including the child (dependent on age).

2. Is there a prohibition on re-identification of an adoptee who has assumed a new identity?

There is an absolute prohibition in the context of identifiable data where there has been a change of identity. This is because there may be issues of personal safety at stake for both adopted child and adoptive parents. The only exception would be where explicit, informed consent has been obtained to allow for this activity.

3. Given the information under questions 1 and 2, how does this apply to intended research for scientific purposes in the public benefit?

Research in the public benefit is a lawful ground for processing within the GDPR Article 6 1(e) or 9(2)(g) and (j). Cafcass England states within its privacy notice [19, 20] that through the application of Practice Direction 12 of the Family Procedure Rules [21], and in accordance with other specific legislation³, the service is permitted to share case information with some third parties (including government departments) for research. The data may also be linked to other information that is held by third parties in order to monitor the effectiveness of family court proceedings and develop policy and good practice. Furthermore, data sharing only takes place if approved by a Cafcass Research Governance Committee. For Cafcass Cymru, the position is slightly different in that permission must be given by either Welsh Ministers or from the President of the Family Division for data sharing for similar purposes. Cafcass Cymru reports to, and is part of, Welsh Government and has separate and independent Research Governance Committees arrangements from Cafcass England. As such, these provisions provide Cafcass England with the legal grounds to permit sharing

for legitimate research, which is proportionate and within the overarching principles of the GDPR. Within Wales, the additional step of obtaining the required permission as detailed would form part of any Data Protection Impact Assessment in respect of data sharing. Therefore, the processing of identifiable data for anonymisation, and passing the data to a trusted third party to perform this task on behalf of Cafcass, is potentially both legitimate and proportionate where due process is followed.

4. How does this apply within the context of using anonymised data in a data safe haven? (This question uses the SAIL Databank as a case in point.)

Extracts of data held in SAIL are available for research in anonymised form in a virtual environment on a secure platform [11]. The model is one of legal anonymisation allowing data to be processed for research purposes outside the remit of the GDPR and DPA. In order to effect anonymisation, SAIL uses a trusted third party to remove identifying information as part of a defined and regulated process [14]. The lawful grounds for processing to de-identify data are defined within Article 6(1)(e) and Articles 9(2)(g) and (j) of the GDPR [16]. In addition to this, Cafcass has specific statutory authority under the Family Procedure Rules, Practice Direction 12G [21] that legitimises the anonymisation process. The position for Cafcass Wales is slightly different as detailed above in that permission from Welsh ministers or the President of the Family Division is a precursor to data sharing.

The only point at which linkage of Cafcass data to health and administrative data can occur is within the SAIL databank. Project-level controls and restricted access are applied in order to manage risk and to ensure the data are legally anonymised for use in research. In addition, Cafcass data are classified by SAIL as a restricted dataset, meaning that proposed uses of the data require data controller permission as well as review by an independent Information Governance Review Panel [11]. The linkage of pre- and post-adoption records from Cafcass data to other datasets, such as from health and social care, is therefore considered safe within the context of SAIL as it does not involve re-identification, but simply links the two parts of anonymised records to produce a continuum for longitudinal research. At no time do researchers have access to the data in identifiable form.

5. Will the existing controls within a data safe haven (using the SAIL information governance model as a working example) suffice in respect of questions 1–4, and is it therefore safe to proceed with this work?

As we note, SAIL will not operate to re-identify an individual who has been adopted but will simply allow two anonymised records relating to one individual to be linked together. It should be acknowledged, however, that the more variables and datasets that are linked, the greater the risk of internal re-identification. SAIL operates on the basis of data privacy-by-design, applying an array of physical, technical and procedural controls to the data and to the environment. SAIL does not release

³The Social Services & Well-being (Wales) Act 2014 (s184) <http://www.legislation.gov.uk/anaw/2014/4/contents>

Table 1: Age and sex of individuals who participated in the study.

Group	18–24	25–34	35–44	45–54	55–64	65+
Adoptees (N=6)				4 women	2 men	
Adoptive parents (N=10)			3 women	2 men, 1 woman	1 man, 3 women	
Birth parents (N=3)	2 women	1 man				
General public (N=15)	3 women	3 women	1 man, 2 women	3 women	1 woman	1 man, 1 woman

Ages are shown in banded years (N=34 in total.). The general public were drawn from a Consumer Panel [18] and a group of young parents.

row-level data outside its secure environment and carries out results screening to ensure they are safe to release externally [11, 22]. Whilst there are legal prohibitions on disclosing and processing identifiable adoption data, research carried out using anonymised data sits outside the legal provisions. It is therefore lawful provided the robust controls are rigorously applied.

Engagement

The age and sex distributions of the participants are summarised in Table 1.

We set out the findings using the two main questions as the framework. We include viewpoints from the interviews with individuals who had been involved in the adoptions process (adoptees, adoptive parents and birth parents) and the views of the general public groups. We then use these to draw out emerging themes.

1) How do you feel about research using anonymised linked family court records and health / administrative data taking place?

Involved individuals

Each of the adoptees (N=6) expressed a positive view on the research taking place. One participant raised a concern about anonymisation and security, but considered that risks were mitigated as long as the data were used in a data safe haven. Being able to carry out the research was seen as beneficial to provide evidence on health and other life outcomes, to flag educational difficulties, psychological issues and differences between adoptees and non-adoptees. One adoptee (a woman aged 45–54) said this is *'really, really interesting and sensible.'* A man (aged 55–64) said he could see the value as *'no one has done this before.'*

The adoptive parents (N=10) expressed favourable views regarding research using the data about adoptees. Three provisos were raised: one on the need for data anonymisation, another the assurance of confidentiality and one about the work needing to be focused on benefitting adoptees. The noted benefits of the research included: obviating the lack of information on outcomes for Children who are Looked After (CLA) and gaining further understanding on grief, trauma and loss. An adoptive father (45–54) said *'I can only see benefits really, but obviously any risks would be around confidentiality.'* The view of an adoptive mother (35–44) was *'I welcome any kind of study that helps the children that get adopted'*.

The birth parents (N=3) were in favour of the research taking place with no concerns raised. One member of the

group expressed that more information would help the service in making decisions to help avoid unnecessary adoption or to go ahead if that was the best option. One member (a man aged 25–34) said *'If there is data out there that can be used to benefit people, then why not use it?'* The others present agreed with these sentiments.

General public

All the members of the general public (N=15) were broadly in favour of the use of the records being used in research, providing that the data were anonymised. One person (aged 55–64) felt that finding out about the work *'may remind individuals of a distressing period in their life'*, and another (aged 25–34) made a similar point: *'if people are aware of the project it could be upsetting to bring up emotions of a period of their life which is traumatic'*. A further point of concern was that the research could lead to certain groups of people being stigmatised if particular associations were found e.g. tendency to crime. In terms of benefits, a participant (aged 35–44) said: *'I support the study as it is beneficial to general public'*; another (aged 55–64) said *'it could also assist in future planning of resources'*. A participant (aged 25–34) felt *'this is an essential project to protect our future generations'*; another (aged 45–54) said that the research *'could [be] beneficial to policy makers and general public'*. A participant (aged 25–34) highlighted the *'huge gap in this field'* and that this work provides the *'opportunity to understand how to get [the] best outcomes for children in care/in care services'*.

2) What are your views on the acceptability of anonymously linking pre- and post-adoption records of someone who has been assigned a new identity during family court proceedings?

Involved individuals

The adoptees (N=6) were generally positive in their views on anonymously linking pre- and post-adoption records. A participant raised a concern about the risks of wider data sharing and felt the work was acceptable if only one or two organisations were involved in handling the data, but highlighted the increased risk if the number of organisations was increased. A woman (aged 45–54) said it's *'common sense for studying the whole person'* and another in the same age band said *'any gaps that can be filled – very important'*. A man (aged 55–64) said *'it makes sense and it's reassuring, adding value to the research'*. He asked whether we could feed back information to individual adoptees as he saw this as a good thing, but we are unable to do this because the data

are anonymised. He recommended there should be additional studies (collecting primary data from adoptees) to add to the family court, health and administrative datasets, and gave the example of mental health. Another adoptee (a woman aged 45–54) also highlighted mental health, noting that because issues can arise later, it would be important to study adults who had been adopted as well as children.

The adoptive parents (N=10) raised a variety of points and were generally in favour of research using linked records, pre- and post-adoption. Four of the participants (two adoptive mothers aged 35–44 and 55–64, and two adoptive fathers aged 45–54 and 55–64) mentioned data protection issues: data privacy, security and re-identification of adoptees. One of the four (the mother aged 35–44) was particularly concerned about the chance of the birth parents having access to her adopted child's identity. A different adoptive mother aged 55–64 also expressed about birth parents gaining access to the new identity, but felt the risk was minute, with social media channels being a greater worry. Three of the four who had raised data protection concerns asked the interviewer about the anonymisation processes used in SAIL and her perception of the risks. The explanation included a summary of the physical technical and procedural controls used by SAIL to mitigate risk [11] acknowledging that risk is minimised but is not zero. The adoptive mother (aged 35–44) said she was reassured by the data safe haven model, that it was essential to have the necessary safeguards in place. She said that *'in the long run both children and parents will benefit'* and *'I endorse this research'*. The adoptive father (aged 45–54) said *'I'm pretty sure that would be ok'* and *'I can only go with what you're telling me in terms of risk'*. The other adoptive mother and father among the four who asked were also reassured. Among wider observations raised were the value of being able to understand more about what happened to child X before and after adoption, and the point made by an adoptive mother (aged 35–44) that having a break in the records is a *'big hindrance to the work that needs to be done'*. It was also observed by one participant that the views of adopters on the linkage of pre- and post-adoption records might differ from those of adoptees.

The birth parents (N=3) had no concerns about research taking place on anonymously linked pre- and post-adoption records. A woman (aged 18–24) said *'this is very beneficial to children so can look back and see when conditions develop'*. The others agreed with this viewpoint.

General public

The participants from among the general public (N=15) had some mixed views, but on balance were in favour of the research taking place provided that privacy was maintained. A woman (aged 25–34) had *'a gut feeling that this is an uncomfortable move'* but she went on to say *'but I do agree that it is an essential move as knowing the persons full health and wellbeing history will help services to be improved using an evidence base'*. A woman (aged 35–44) said *'doing this in a data safe haven shouldn't present any problems'* and *'I can see the value in this to allow continuity of data'*. A man (aged 35–44) felt that the linkage of pre- and post-adoption records was *'acceptable if it is for research and policy making only'*. Two people expressed concern about disclosure

of an adoptees identity where the adoption had been related to safeguarding the child. A woman (aged 45–54) stressed that *'there needs to be a clear case of benefit to children and families'*. Another (aged 55–64) said that *'as the data for the research is in anonymised form I think the risks would be minimal and may offer valuable information in terms of planning health resources'*.

Discussion

What this study adds

- This is a unique study combining a legal review and public engagement to shed new light on the use of adoption records for research.
- It advises that there is no legal impediment to anonymising and using adoption records for research in a data safe haven.
- It indicates that there may be few social implications and that individuals involved in the adoptions process (adoptees, adoptive parents and birth parents) are broadly in favour of such research taking place.
- The findings of the legal review and engagement activities include the general acceptability of linking pre- and post-adoption records.
- The conditions are that robust safeguards must in place to protect privacy, ensure confidentiality and mitigate the risk of identity disclosure.
- Further provisos are that the research should be designed to benefit adoptees and to influence policy and practice towards improvements.
- The study design and findings are relevant in jurisdictions where the NHS (or other identity) number is changed at adoption, or for another reason where a person is assigned a new identity.
- The findings have value for progressing the novel use of adoption records for research purposes in Wales and as a demonstrator for application in other countries.

Main findings

This unique study has provided new information on the legal and social implications of using adoption records for research and of linking pre- and post-adoption records in a data safe haven environment. The legal review guides on the conditions under which data about adoptees can be shared and used to enable research for public benefit and simultaneously protect adoptee identity. Although there are strict rules on the use of identifiable data about adoptees, and particularly on disclosing their new identity, there are suitable provisions to enable safe data sharing to take place. Cafcass is able to rely on Practice Directions in the Family Procedure Rules, and Articles 6 and 9 of the GDPR to enable lawful de-identification to take place. SAIL similarly relies on GDPR Articles 6 and 9 to process extracts of de-identified datasets

into anonymised form for research [11]. The research in SAIL does not involve re-identification, but would simply link two anonymised records to produce a continuum for longitudinal research. Since Cafcass currently provide their data only to the SAIL Databank, we have used the information governance model of this safe haven in this paper.

Although the data are to be used in anonymised form only, the social licence for research is important. Public engagement on the use of individual level data is an increasingly strong feature among many organisations [23, 24]. Engaging with individuals who have been involved in the adoptions process (adoptees, adoptive parents and birth parents) and members of the general public provided an opportunity for their voices to be heard and revealed broad support for the work. Some concerns were raised about the use of the data in general, which were more pronounced when considering the linkage of pre- and post-adoption records. The main themes emerging in the concerns about the first question posed, which was on the use of the anonymised linked family court records and health / administrative data taking place, were: data anonymisation; confidentiality; reminding adoptees about a traumatic period in their life; and stigmatisation if certain problematic associations were discovered. The main themes in the benefits were: being able to use the data to compare outcomes for adoptees and non-adoptees; and assisting decision-makers on when adoption was necessary. A relatively strong positive theme was the opportunity to increase understanding of mental health issues, including trauma, grief and loss.

The main themes in the concerns about linking pre- and post-adoption records for research were: privacy; data security; limiting the number of organisations handling the data; and re-identification risk, particularly if birth parents were able to gain information on the new identity of an adoptee. The main themes among the benefits were: being able to use the records for longitudinal research; and the opportunity for additional studies on mental health issues. Viewpoints were positive on balance, with the following provisos: adoptee identity must be safeguarded; the research should have a clear benefit to adoptees; and it should be focused on guiding policy and practice.

The legal review and engagement exercises were encouraging for the future of research about adoptees. The review indicated that it is lawful to use anonymised adoption records for research in a data safe haven, including linking pre- and post-adoption records, provided that robust controls are in place to control and mitigate disclosure risk. It is possible that individuals could be assigned a new identity for other reasons, such as gender reassignment, and thus the study design and findings could be relevant more widely. However, we do not comment on the legal and social implications or other issues that may arise in these scenarios. Valuable points were raised by involved individuals and members of the general public, and we are taking these on board to guide the direction of our work.

Limitations

The work we have described relates to the use of adoption records within a data safe haven operating on a similar model

to the SAIL Databank. We have not considered models where data would be released externally. The legal review is based on England and Wales and the engagement work focused on Wales since it is the Cafcass Cymru data that we are currently able to link to routine health and administrative data. SAIL does not currently hold sufficient routine health and administrative data about the English population to do this with data from Cafcass England. The NHS number (or equivalent) does not necessarily change at adoption in other jurisdictions, even within the UK. Recruiting involved individuals as participants was challenging due to sensitivities around the process of adoption. The Consumer Panel might be more *au fait* with data use than some groups of the general public because of their experience of working with SAIL. Because of the way participants were recruited, we cannot assess selection bias. We do not claim that the viewpoints we have obtained are necessarily representative of all involved individuals or of the general public as a whole. However, we have no reason to believe the findings would differ so significantly as to render them unsound. The consensus statement on public involvement and engagement with data-intensive health research lists eight good practice criteria [23]. While we have endeavoured to meet those relevant to the study, we acknowledge this may have been partial.

Next steps

We are highly aware of complexities in working with records about adoptees. For example, information in particular datasets (such as Cafcass, social care) is partial. This is because they cover various stages in the adoptions process. It also can be incomplete as Cafcass Cymru is not involved in adoption orders for all adoptees. We are currently studying the datasets to understand their scope and limitations. Based on this work, the main next steps are to consider:

1) How linkage of pre- and post-adoption records can be enacted

Since linkage within SAIL depends on an encryption of the NHS number, and a new NHS number is assigned at adoption, it is not straightforward to link pre- and post-adoption records. It is theoretically possible to select a number of variables common in the two parts of the record and link them probabilistically. However, except for cases of rare conditions or other unusual factors, the variables present are unlikely to be highly discriminatory, rendering reliable matching doubtful for the majority. Another option would be if the NHS centre [25] that allocates the new NHS number at adoption would be able to provide a linkage key between the pre- and post-adoption records to a trusted third party and onward to SAIL. We will explore all reasonable possibilities.

2) Priority topics for research

As has been noted, adoptees may have health and other difficulties during their life course over and above those experienced by non-adopted peers [1–4]. Being able to link adoption records to health and administrative datasets opens up new opportunities for research. It is important to note, however, that not all children who are subjects in family court proceedings go on to be adopted as there are other outcomes, such as returning home, staying in foster care or being placed with relatives. The questions that are critical, however, concern

the longer-term life chances for different groups of children, according to permanency decisions. Record linkage enables a more holistic view to be gained of children's interactions across a number of services, as well as longer-term life chances. These points, and the importance of relating our work to policy and practice leading to public benefit, will guide the Family Justice Data Partnership research programme.

3) Developing public engagement

During the engagement activities, it was noted that some individuals who had been involved in the adoptions process might feel worried about their data being used in research in case the identity of an adoptee was revealed, and that just becoming aware of the research taking place might remind someone of a traumatic period in their life. Furthermore, it proved valuable to hear participant views on the direction and focus of research to be undertaken. We are tuning our public engagement programme for the use of adoption records in research and the information we have gained is feeding into this. It will include regular discussion with the Consumer Panel [18] and with a group of involved individuals invited from among the participants, as well as key stakeholders in adoption agencies, fostering services, advocacy groups, legal counsel, social workers, etc. Through this engagement, we will gain further views on priorities for research, to combine with those of policy makers and practitioners. We will also prepare a written summary of how we work with the adoption records with the safeguards in place, and general reader summaries of research studies. We will make this information available to anyone who requests a copy and on our departmental Population Data Science website [26].

Conclusion

The availability of the Cafcass datasets for use with health and administrative data opens us rich new opportunities for research about adoptees, and others who have been subjects in a family court case. This study has shown that there is no legal impediment nor major objection from individuals who have been involved in the adoptions process and the general public for the use of the data in anonymised form in a data safe haven. This includes the linkage of pre- and post-adoption records to enable novel longitudinal research to take place. The provisos were that robust safeguards must be in place and the research should aim to benefit adoptees and to improve policy and practice. We conclude that it is reasonable to proceed with caution to develop practical ways to link pre- and post-adoption records.

Ethics

This study was granted ethical approval by the Swansea University Medical School research ethics committee, ref: 2019-0053.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The work conducted in this study was supported by Administrative Data Research (Wales) funded by the Economic and Social Research Council (ref: ES/S007393/1), and the SAIL Databank funded by Health & Care Research Wales (this funder does not use reference numbers). We would like to express our grateful thanks to the many policymakers and practitioners with whom we had supportive discussions about this study and the novel opportunities arising for data linkage research. These included representatives of: the Nuffield Family Justice Observatory, Welsh Government (Senedd), Department of Health, Department for Education, Adoptions UK, Cafcass Cymru, Cafcass England, the Adoption & Special Guardianship Leadership Board, and medical professionals specialising in Children who are Looked After, adoptions policy and clinical ethics. Special thanks are due to all the participants who took part in the study.

References

1. Department for Education (2019) Children looked after in England including adoption: 2018 to 2019. <https://www.gov.uk/government/statistics/children-looked-after-in-england-including-adoption-2018-to-2019>
2. Rushton A, Grant M, Feast J and Simmonds J (2013) The British Chinese Adoption Study: orphanage care, adoption and mid-life outcomes, *Journal of Child Psychology and Psychiatry*: 54(11): 1215–1222. <https://doi.org/10.1111/jcpp.12088>
3. Felitti VJ, Anda RF, Nordenberg D, *et al* (1998) Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine* [Internet]. 1998;14(4):245–58. [https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8)
4. Bellis MA, Lowey H, Leckenby N, *et al* (2014) Adverse childhood experiences: retrospective study to determine their impact on adult health behaviours and health outcomes in a UK population. *Journal of Public Health (Bangkok)* [Internet]. 2014;36(1):81–91. <https://doi.org/10.1093/pubmed/ftd038>
5. Selwyn J, Wijedasa D and Meakings S (2014) Beyond the Adoption Order: challenges, interventions and adoption disruption <https://www.gov.uk/government/publications/beyond-the-adoption-order-challenges-intervention-disruption>
6. Nuffield Family Justice Observatory for England & Wales (2019) Improving intelligence for the family justice system – A Nuffield Family Justice Observatory Data Partnership, <https://www.nuffieldfjo.org.uk/app/nuffield/files-module/local/documents/NFJO%20DP%20Summary%20Document.pdf>
7. Broadhurst K, Budd T and Williams T (2018) The Nuffield Family Justice Observatory for

- England and Wales: Making it Happen https://www.nuffieldfoundation.org/sites/default/files/files/Nuffield_Family_Justice_Observatory_making_it_happen_v_FINAL_13_02_18.pdf
8. Nuffield Family Justice Observatory (2020) <https://www.nuffieldfjo.org.uk/about>
 9. Bedston Johnson RD, Ford DV, Broadhurst K, *et al* (2020) Data Resource: population level family justice administrative data with opportunities for data linkage, IJPDS, [in press]
 10. Bedston S, Pearson R, Jay MA, *et al* (2020) Data Resource: Children and Family Court Advisory and Support Service (Cafcass) public family law administrative records in England, IJPDS, 5:1. <https://doi.org/10.23889/ijpds.v5i1.1159>
 11. Jones KH, Ford DV, Thompson S and Lyons RA (2019) A Profile of the SAIL Databank on the UK Secure Research Platform. IJPDS, 2:4. <https://doi.org/10.23889/ijpds.v4i2.1134>
 12. SAIL Databank (2020) SAIL datasets <https://saildatabank.com/saildata/sail-datasets/>
 13. Fouzder M (2020) 'Astonishing breakthrough': Munby hails access to family justice data. The law society gazette <https://www.lawgazette.co.uk/news/astonishing-breakthrough-munby-hails-access-to-family-justice-data/5103266.article#.XI9ps2aVadA.twitter>
 14. Lyons RA, Jones KH, John G, *et al* (2009) The SAIL databank: linking multiple health and social care datasets. BMC Medical Informatics and Decision Making 2009, 9:3. <https://doi.org/10.1186/1472-6947-9-3>
 15. HM Government (2018) Data Protection Act <http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>
 16. European Union (2016) General Data Protection Regulation <https://gdpr-info.eu/>
 17. HM Government (2002) Adoption and Children Act <http://www.legislation.gov.uk/ukpga/2002/38/contents>
 18. Jones KH, McNerney CL and Ford DV (2014) Involving consumers in the work of a data linkage research unit. International Journal of Consumer Studies, 38:1:45–51. <https://doi.org/10.1111/ijcs.12062>
 19. Cafcass England (2020) <https://www.cafcass.gov.uk/about-cafcass/policies/privacy-data-protection/privacy-notice-service-users/?highlight=privacy%20notice>
 20. Welsh Government (2020) Cafcass Cymru privacy notice <https://gov.wales/sites/default/files/publications/2020-01/cafcass-cymru-privacy-notice.pdf>
 21. HM Ministry of Justice (2020) https://www.justice.gov.uk/courts/procedure-rules/family/practice_directions/pd_part_12g
 22. Jones KH, Ford DV, Jones C *et al* (2014) A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: a privacy-protecting remote access system for health-related research and evaluation, Journal of Biomedical Informatics: special issue on medical data privacy. <https://doi.org/10.1016/j.jbi.2014.01.003>
 23. Aitken M, Tully, MP, Porteous C, *et al* (2019) Consensus Statement on Public Involvement and Engagement with Data-Intensive Health Research. IJPDS, 4:1:06. <https://doi.org/10.23889/ijpds.v4i1.586>
 24. Waind E (2020) Trust, Security and Public Interest: Striking the Balance. ADR UK (Administrative Data Research UK) Strategic Hub, Economic and Social Research Council https://www.adruk.org/fileadmin/uploads/adruk/Trust_Security_and_Public_Interest_Striking_the_Balance_ADR_UK_2020.pdf
 25. NHS Digital (2020) National Back Office for the Personal Demographics Service <https://digital.nhs.uk/services/national-back-office-for-the-personal-demographics-service>
 26. Population Data Science at Swansea University Medical School (2020) <https://popdatasci.swan.ac.uk/>

