**DATA LINKAGE FOR EARLY INTERVENTION IN THE UK: PARENTAL SOCIAL LICENCE AND SOCIAL DIVISIONS**

Rosalind Edwards, University of Southampton, UK – r.s.edwards@soton.ac.uk *\*corresponding author*

ORCID: 0000-0002-3512-9029

Val Gillies, University of Westminster, UK

ORCID: 0000-0001-8125-0420

Sarah Gorin, University of Southampton, UK

ORCID: 0000-0002-0721-3880

**Abstract:**

Electronic linking of public records and predictive analytics to identify families for preventive early intervention increasingly is promoted by governments. We use the concept of social licence to address questions of social legitimacy, agreement and trust in data linkage and analytics for parents of dependent children, who are the focus of early intervention initiatives in the UK. We review data-steered family policy and early intervention operational service practices. We draw on a consensus baseline analysis of data from a probability-based panel survey of parents, to show that informed consent to data linkage and use is important to all parents, but there are social divisions of knowledge, agreement and trust. There is more social licence for data linkage by services among parents in higher occupation, qualification and income groups, than among Black parents, lone parents, younger parents, and parents in larger households. These marginalised groups of parents, collectively, are more likely to be the focus of identification for early intervention. We argue that government awareness-raising exercises about the merits of data linkage are likely to bolster existing social licence among advantaged parents while running the risk of further disengagement among disadvantaged groups. This is especially where inequalities and forecasting inaccuracies are encoded into early intervention data gathering, linking and predictive practices, with consequences for a cohesive and equal society.

**Policy Significance Statement:**

Linking of administrative records and predictive analytics to identify families for preventive early intervention increasingly is promoted by governments. Informed consent to use of their data is important to parents, and there is less acceptance of public services using data linkage among marginalised social groups. Implementation of operational data linkage amongst services working with families has the potential to undermine social legitimacy and trust, with consequences for a cohesive and equal society. Addressing this through explaining the merits of data linkage is likely to bolster social licence among parents in higher occupation, qualification and income groups, while generating further disengagement and avoidance of public services among marginalised parents. Rather, meaningful dialogue that shapes the parameters of data linkage is required.

***Keywords:*** *early intervention, operational data linkage, parents, predictive analytics, social divisions, social licence*

**Introduction**

Electronic linkage of public records and predictive analytics for the operational purpose of identifying families for preventive early intervention increasingly is promoted by governments, part of a transition to data-steered social policy in the UK and internationally. Sharing and linking the separate sources of nationally and locally held information held about citizens – health, education, social care, police, housing, immigration, taxation and social security records etc., and operationalising them through algorithmic data analytics, is championed as offering the possibility of improved and more efficient public service delivery, and enabling predictive risk modelling to pre-empt problems and improve outcomes (e.g. NL Digitaal Government, 2019; Privy Council Office Canada, 2018; Stats NZ, 2018). The COVID-19 pandemic has accelerated this trend, notably boosting calls for services and agencies to share and join up their routinely-collected data. This raises questions not just about service efficacy, but importantly about the extent to which transitions to such data use occur without a democratic mandate and transgress legal, ethical and data quality norms (van Zoonen, 2020). James Shaw and colleagues (2020) argue that data sharing, linkage and the application of analytics need to earn ‘social licence’ – that is the agreement and trust of citizens, for ‘COVID-era’ data initiatives, and they advise the need for transparency and public involvement. Yet, as Liesbet van Zoonen argues, moves towards data-driven social policy largely take place out of political and social view. Data linkage and predictive analytic practices are centred on top-down monitoring, containment, and control: ‘Citizens in the system are subjected to those processes, as a group and sometimes individually, without knowing it’ (van Zoonen, 2020: e10-9). This raises questions about power, social inequalities and whose interests are serviced with these practices.

In the UK, the Government’s National Data Strategy was updated at the end of 2020 to exhort public services to share their administrative records, citing the way that data linkage has been essential for public health responses during the COVID-19 pandemic, and how this situation has underlined the need for: ‘the presumption is that, with appropriate safeguards, data should be shared to drive better outcomes. (DDCMS, 2020). The House of Lords Select Committee on Public Services similarly has emphasised data sharing for public services working with children and families specifically:

We are concerned that agencies do not share the data that they need to support vulnerable children and to determine which children need their help. The Government should issue new guidance on data-sharing powers and duties to protect vulnerable children, and, if necessary, introduce legislation to ensure that such data is shared. (2020: 43)

Yet, it is all families who are implicated in across-the-board data sharing, data linking, and application of predictive modelling, not just the families with ‘vulnerable children’. Population level data comprises data from individual families; from all families nationally or from all living in a particular local authority area. This data is linked together across different service sources at population level and subject to predictive analytics to flag up individual families for preventive intervention.

Other lessons from the pandemic are the revelation of stark social divisions and material inequalities cutting across British society (Johnson et al., 2021), and a lack of trust in public institutions among minority groups demonstrated in vaccine hesitancy, including those with lower qualifications and income, and especially among Black people (Ansell et al., 2021; ONS 2021). While supporting the improved flow of information across government, the British Academy notes potential for the exacerbation of low and unstable levels of trust among disadvantaged groups and hence challenges for social cohesion:

The steepest declines in perceptions of unity and solidarity have been in some (but not all) of the most deprived communities, among key workers and in certain ethnic minority groups … Reduced trust in national government also leads to reduced societal trust, enabling division and the targeting or scapegoating of particular groups. Thus, trust and cohesion are linked … There may also be questions about trust in governments’ use of other measures, such as technology or data linkage (British Academy, 2021: 76, 128).

Advocates of extensive joining up of public records to support operational service interventions often regard data and its collection as neutral and objective, without recognising how they can reflect discrimination and intensify the social inequalities they capture (Benjamin, 2019; Eubanks, 2018; Wachter-Boettcher, 2017). There is evidence that inequalities are encoded into the data gathering, linking and predictive practices that drive early intervention (Redden et al., 2020). The transition to data-steered social policy has subjected already marginalised groups to more disadvantage, with built-in discrimination in data bases. There are bias and errors in the data sources that are merged, and in the design of data modelling and predictive analytics applied. Particular sub-groups of parents and families are disproportionately represented in social security, social care and criminal justice systems, leading to the encoding of existing social divisions of class, race and gender in their data sets. In the UK, for example, predictive risk modelling used in child protection embeds an equation of socio-economic disadvantage with risk, discriminating against poor families (Vannier Ducasse, 2020), while attention has been drawn to the unsupported over-identification of young Black men in what amounts to digital racialised and gendered profiling in police databases that are shared with other agencies (Amnesty International United Kingdom Section, 2018; Wroe, 2021) – an injustice that is repeated in data-driven social policy systems internationally (e.g. Benjamin 2019; Eubanks 2018, Keddell 2014; Wachter-Boettcher, 2017).

In this article we address questions about trust and social cohesion in the push for more extensive data linkage for operational purposes. What are the views of those directly affected: parents of dependent children, about what is acceptable or unacceptable in relation to information about them? How far does use of data linkage and predictive analytic for operational service intervention lie within social acceptance norms? And is the extent of trust in their use by services shaped by parents’ social location, by social divisions?

We begin with a discussion of the interlocking of early intervention initiatives in families and the use of data-driven operational practices in the UK, whereby government seeks to pre-empt dysfunctional parenting and poor outcomes for children by linking together administrative records from public services and subjecting them to predictive risk modelling to identify and target families. We then outline the concept of social licence as a framing for our survey investigation of parents of dependent children before moving on to consider how transparency is regarded as a policy solution to embedding trust in the process, and parents’ views about the joining together of administrative records.

**Data linkage for early intervention in the UK**

The way that parents bring up their children is a longstanding social policy concern, identified as a cause of and solution to the state of the nation, and as a driver of social cohesion. Over past decades this focus has intensified and shifted, from implementing support for all parents in transmitting acceptable values to their children, towards a focus on children and families at risk, and early intervention in particular families’ lifestyles and behaviour. Early intervention aims to pre-empt rather than react; to prevent any risk of social, educational, health and behaviour deficiencies that, it is asserted, might otherwise occur at some point in the future, or at least address them early on when they occur to prevent them escalating (Edwards and Gillies, 2004; Gillies et al., 2017). For example, the UK’s Early Intervention Foundation was established by government in 2013 to ‘champion and support the use of effective early intervention to improve the lives of children and young people at risk of experiencing poor outcomes’[[1]](#footnote-1). Their focus on evaluating various parenting skills delivery packages has more recently been augmented by attention to the potential of administrative data for targetting and tracking early intervention (Scourfield et al., 2019). Such data-driven family policy initiatives are legitimated by what Sue White and David Wastell (2017) refer to as ‘prevention science’, a mix of technological, biological and behavioural sciences and morality that is fuelling a realignment of the relationship between families, the state and professions without open debate. Indeed there are concerns that mass data collection and automated analysis have become a governance end in itself, for top down monitoring and control (Dencik et al., 2019; van Zoonen, 2020). It is as if, in itself, collecting more information and merging data is the solution to social problems.

What goes on in families has become keyed into a wider endeavour of governance with and through data. In addition to UK government exhortations to link administrative records noted in our Introduction, the Early Intervention Foundation calls for local authorities to set up information-sharing ‘assessment hubs’: ‘It is essential that data is shared between health services and the local authority at population level and, where necessary, at an individual level to ensure that families who need services are offered them’ (Messenger & Molloy, 2014: 7). At the time of writing there are plans for local authority ‘family hubs’ and digital ‘red books’ containing information about babies from birth that can be used to identify parents deemed to need support (DHSC, 2021). Central government has initiated a Local Data Accelerator Fund for Children and Families (MHCLG, 2021a), where local authorities can bid for funding for data sharing and matching projects that support identification of families for ‘earlier intervention before risk escalates’ (MHCLG, 2021: 7). The prospectus for the Fund provides the examples of Liverpool City Council combining 35 feeds of data from children’s social services, schools, the criminal justice system, health and benefits data in order to identify those who could benefit from early intervention, and Bristol City Council Insight team’s establishment of a multiagency data ’warehouse’ as an analytic hub to help predict children at risk of criminal or sexual exploitation, becoming NEET (not in education, employment or training), or being a victim or perpetrator of serious violence (p.8).

Like Liverpool and Bristol, a number of local authorities have embarked upon linking up sets of administrative and other data in order to identify ‘high risk’ families for forms of local authority provided or contracted early intervention in the way that parents bring up their children (McIntyre & Pegg, 2018). There is some evidence however, that the majority of authorities only have ‘basic’ data matching software or the ‘building blocks’ of data linkage and analytics rather than more ‘mature’, advanced systems (MHCLG, 2021a: 7). It is impossible to discuss this with any accuracy however, since there is no central record available of which local authorities are doing what when it comes data sharing, operational linking and matching, and predictive analytics in the family policy field, and indeed no shared vocabulary between local authorities about what they are doing.

Local authority interest in multi-agency data sharing and linking to identify and intervene was heralded by the Troubled Families programme, reborn as Supporting Families in its latest phase (MHCLG, 2021b). The Troubled Families programme was set up by central government to intensively intervene in families who meet a combination of specified criteria that are treated as evidence of their current or future risk of dysfunctionality (Crossley, 2018). Local authorities are encouraged to identify families as part of the programme because is run on a Payment by Results basis – an attractive prospect for cash-strapped local authorities in the context of austerity. The suggested criteria for the Troubled Families programme range widely across the domains of social security, housing, education, health, social services, police and criminal justice, and other public provisions (MHCLG, 2020), and under the Supporting Families version of the programme the emphasis is on ‘building stronger data’, posed as part of a ‘moral mission’ to support ‘vulnerable’ families (MHCLG, 2021b).

The integration and analysis of administrative records for extracting profiles based on whole populations, and for predictive analysis flagging up particular families, may be carried out in-house by the public sector. More often, there is a range of different types and extents of involvement of commercial data analytic companies (Redden et al., 2020) in creating and operating the data hubs, data linkage, algorithmic analytics, etc. in the social policy domain. As we noted above though, there is no easily accessible means, such as a public register, for parents to find out what is happening with their own data. What little public consultation there has been about sharing and linkage of administrative records has usually focused on anonymised data for research purposes, and/or involved general population focus group discussion (e.g. Moody & Lugg, 2017; NatCen, 2018). All parents are stakeholders in the use of administrative records for data linkage and predictive analytics for targetting service intervention, but they appear to have played no part in assessments of the legitimacy of the application of data techniques to information about them and their families. The integration and outsourcing of operational practices involved in early intervention lie outside of automatic social acceptance norms, social trust and consensus, so social licence for them needs to be ascertained.

**Social licence**

Considerations of data linkage and predictive analytics for operational service intervention have been turning to the concept of social licence in a context where there is concern internationally and nationally, about the existence and sustenance of trust for these practices (e.g. Caldicot, 2016; Leonard, 2018; World Economic Forum, 2018). The concept’s growing traction is evident, notably in New Zealand, for example, in discussion of plans for data sharing between sectors (Data Futures Partnership, 2017) and Statistics New Zealand’s Integrated Data Infrastructure (Gulliver et al., 2018), but also concerning data linkage and automated intelligence in Canada (Paprica et al., 2019) and Australia (Leonard, 2018). In the UK as the focus of our discussion here, the Office for Statistics Regulation (2018: 14) asserts:

Proactively seeking to build and maintain social licence around data use should, in the long-term, also help to increase public understanding of the benefits and risks of data sharing and linking. The public will then be in a better position to engage in debates about new proposals for data use, and to judge the consequences of breaches, if they occur, either with government data or that held by private bodies.

The concept of social licence concerns social legitimacy and acceptance of practices that lie outside general norms. It draws attention to the issue that formal legal authority to share and link data does not automatically command social trust and consensus. For example, Pam Carter and colleagues (2015) drew on the concept of social licence to explain public concern about the UK ‘care.data’ initiative, involving the sharing of personal medical records for secondary purposes. They argued that the lack of public social licence for the initiative was related to poor provision of trustworthy information, the rupturing of normative GP-patient relations, and little evidence of any public good. Similar social licence points have been made about public backlashes over Australian Bureau of Census plans to link individually-identified data with other public records (Easton, 2017), and US health care organisation plans to share data with Google for advanced analytics (Wachter & Cassell, 2020). Legal licence is not necessarily a foundation for social licence.

Crucially, social licence as a conceptual approach treats broad acceptance of legitimacy and the trust that sustains it as relational, emerging from situated perceptions and understandings. As such, it is a dynamic process. The focus on a consensus of social approval draws on sociological theorising about the relationship between professionals and society. Notably, Everett Hughes (1958) conceptualised professional groups as socially licenced to carry out particular activities; that is, the public affords permission to professions to adopt practices that lie beyond normative conventions without incurring any social sanction. Another source of the concept of social licence is the examination of corporate responsibility in extraction industries, where the environment or communities may be harmed by their activities (e.g. Thomson & Boutilier, 2011). In this field there is a strand of literature that stresses the need to build and maintain a consensus of social licence among relevant stakeholders, especially the particular population affected, even if operating within the law. Social licence, then, points to the need for agencies undertaking activities that could give rise to public concern and controversy to go further than compliance with legal requirements, to make ongoing efforts to secure and maintain social licence.

In the contexts of emphasis on data linkage and analytics in the field of family policy and early intervention, and concern about the existence and sustenance of trust and ethical practice, we undertook an investigation of social licence among the population implicated: parents of dependent children. If policymakers and service providers are to begin to engage with social licence as a dynamic process, there needs to be some knowledge of the bases from which they are starting. If social licence already exists and engagement is directed towards maintaining it, then that will involve a different sort of dialogue with stakeholders than understanding where it does not exist and needs to be built. Relatedly, dialogue with those most likely to be implicated in a policy action may look very different from engagement with those who are least likely to be. Alternatively, the social licence knowledge base may mean that policymakers and service providers understand that a policy action should not be pursued because of unintended consequences.

**Our survey**

Our Parental Social Licence for Data Linkage for Service Intervention project*[[2]](#footnote-2)* investigates the dynamics of social licence and trust for the operational use of data linkage and predictive analytics to identify families for service intervention. It aims to provide an understanding of social licence for these data practices among those implicated: parents of dependent children (<16 years). As part of our research we commissioned from NatCen an online and telephone probability-based panel survey, designed to be representative of parents across the UK, to gain an understanding about what parents deem to be acceptable or unacceptable in relation to data linkage and analytics, and to assess if there is any discernible consensus indicating parental social licence. This is the source of the data that we draw on in this article, specifically relating to the elements that addressed early intervention.

The NatCen panel is recruited from the British Social Attitudes survey, a high quality random probability-based face-to-face survey. The research panel is designed to be representative of the population and produce reliable estimates of opinions. It employs a sequential mixed mode fieldwork design, and weights for non-response. Full details about the methodology of NatCen’s probability-based research panel are available at: <https://www.natcen.ac.uk/media/1484228/Developing-the-NatCen-Panel-V2.pdf>. Questions for our data linkage survey were piloted through an online parents group initially, updated through analysis of subsequent government, corporate, advocacy and media publications, and refined in discussion with NatCen. Randomisation of statement sets within questions and flipping of answer option order was used to counter mode effects.

The questions asked in the ‘attitudes to joining data’ survey covered in principle views on a range of aspects of data linkage and analytics for operational purposes. We asked about awareness of the collection and linking of administrative records, assessments of a range of early intervention rationales for data linkage and analytics, and acceptance of and trust in various bodies and services to undertake linkage of different types of information. The rationale statements (e.g. reasons for linkage, reasons for trust or lack of it) were drawn from our discourse analysis of the contents of reports and online materials from national and local government, data analytic companies, charities and advocacy groups, and mainstream media that related to early intervention (see Edwards et al., 2021).

Explanation of data linkage was treated as a process throughout the survey. It began with a description of administrative records as information collected by government departments and public service providers about people who use their services, and provided examples. Examples of data linkage and analytics drawn from our report and online materials analysis (Edwards et all, 2021) were provided as the survey progressed, including short vignettes about a local council wanting to support parenting skills, a social worker judging whether a family needs further investigation, and a police authority wanting to prevent crime and antisocial behaviour. For the most part we used a Likert scale for responses, apart from responses to questions about awareness which were dichotomous.

The probability-based sample consisted of 843 parents, of whom 57 per cent were mothers and 43 per cent fathers. Looking at the various household types that parents lived in, the majority were families of two parents with children, at 74.5 per cent. Eleven per cent were lone parents and 24 per cent of the parents lived in households comprising five or more people. Twelve per cent of the sample were younger parents, aged between 18 and 29 years. Turning to indicators of social class such as occupation, education and income, 44 per cent of the parents were in managerial and professional occupations and 50 per cent were educated to degree level or above, while 38.5 percent were in lower semi-routine and routine occupations and 10 per cent had no qualifications. The majority of parents, 65 per cent, earned £3700 per month or under, with 60 per cent of the sample owning their homes, while the remainder rented, split between local authority/housing association or privately. Ethnically, 73 per cent of parents were White British, with 19 per cent of our sample from minority ethnic groups, five per cent of whom were Black. These social divisions among the overall sample of parents obviously involve small numbers and so care must be taken here, but as will be clear from our discussion below, the profiles of the extension of (lack of) social licence for operational data linkage among Black and other marginalised groups of parents stand out.

The concept of social licence concerns social legitimacy and approval of practices that lie outside general norms of what is acceptable, as discussed above – a consensus of social agreement giving social licence. This raises the question of what level of agreement constitutes a consensus? A consensus is something more than a simple majority. In line with our conceptual approach, we adopted the consensus baseline for analysing our survey data, in which we took account of the number of response choices available to a question in order to determine what constitutes a social licence consensus (see Finch & Mason, 1991; Edwards and Gillies, 2004). For a two-response options question, rather than more than 50 per cent being taken as a consensus, if one of the options gathers half as many responses again: 75 per cent, then where this was a positive response to data linkage and predictive analytics we took that to represent a widespread granting of social licence. For a three-option question, we took a 50 per cent or greater positive response as indicative of social licence. We conducted the consensus baseline analysis at the level of the sample of parents as a whole, and also looked within this, to sub-population groups of parents (family types, social class indicators, ethnic groups) to see if they may or may not reach a consensus baseline for granting social licence at an appreciably higher or lower level than the rest.

**Transparency about data linkage**

Openness, accountability and transparency has emerged as a theme in government-commissioned reports on the use of data linkage and algorithmic analyses. For the most part, this theme is centred on openness between services in sharing data and accountability as legal compliance But there is some equation of transparency with public trust, with recommendations that people should be told about what is happening to their data and how it is used (e.g. Information Commissioner’s Office, 2020; Centre for Data Ethics and Innovation, 2020). The UK government’s Centre for Data Ethics and Innovation, for example, asserts that transparency in the use of algorithms will build public trust and recommends informing people about the process of developing and using algorithms. This is transparency of a certain kind, akin to lifting the lid off a black box so that the public may look inside but not touch. Consent on the part of those concerned to sharing and linking of their data is not a feature. Yet informed consent to operational use of their family administrative records is important for parents of dependent children.

A majority of the parents surveyed for our research said that were aware that administrative records are collected and digitally stored by government departments and public service providers (72 per cent), but only just over half knew that digital administrative records from different sources can be joined together to find out more about individual families (53 per cent). There were gradients here by ethnicity and social class indicators for both awareness of digital records and data linkage. Black and Black British parents were more aware than their White British counterparts, and parents with higher occupations, education and income more aware than other parents (see Table 1).

INSERT TABLE 1 HERE

There was consensus among the parents that families generally do not know or understand how their administrative records are used, with 60 per cent making this judgement (with a consensus threshold of 50 per cent). There was overwhelming agreement that Government should publicise that they are joining together administrative records about families, and how they use that information (81 per cent with a consensus threshold of 50 per cent). But going further than ideas about transparency engendering trust, there was also consensus that parents need to be asked permission for administrative records about their family to be joined together (60 per cent with a consensus threshold of 50 per cent). Some marginalised groups of parents had an even stronger consensus about the need for parents to give consent to data linkage, notably Black parents and lone parents (each at 66 per cent).

The hopeful notion in reports, that openness and transparency as awareness will engender public trust, is not so simple when it comes to parents of dependent children (or indeed more widely Kennedy et al., 2020). There are social divisions between the assessments of groups of parents, which raises issues about the implications of data linkage for social legitimacy and trust among the parents who are most likely to be subject to early intervention.

**Data linkage for early intervention and social divisions**

Data linkage and predictive analytics for early intervention are promoted in reports and online materials from national and local government, and data analytic companies, and other supportive bodies, as: delivering powerful control of superior knowledge in the hands of local authorities; timeliness, especially through incoming ‘real time’ data about families allowing for quick ‘early warning’ risk prevention; and economic efficiency, optimising existing resources and making savings through prevention rather than crisis management (see discourse analysis in Edwards et al., 2021). These rationales for undertaking data linkage and risk modelling are driven by austerity-tightened finances, by prevention science as the answer to straightened services, and by political and public concerns about child protection and abuse (Jupp, 2017). Under these conditions, rather than providing universal support services for families, public services providers aim to target specific families that they judge may face and cause difficulties, with dedicated early interventions to prevent the risk of problems embedding themselves, and thus to constrain costs.

Our survey put a series of dominant rationales about the benefits of joining together administrative records for early intervention to parents, drawing on our analysis of reports and online materials about data linkage and family services (see Edwards et al., 2021). We asked about the extent to which they thought that it was acceptable or unacceptable, respectively, to: identify families that might need support whether or not they have asked for it; save time and money by preventing family problems before they developed; promote efficiency by targeting services at families that have been identified as needing support; and identify families where children could be at risk of abuse to intervene and prevent it happening. At this abstract, overarching rationales level, the parents surveyed granted social licence for joining together administrative records for early intervention, but they were more circumspect when considering the specifics of trusting particular public services to do this, and private sector involvement in the process.

Among parents as a whole in the sample there was a consensus for joining families’ administrative recordsto enable early intervention at a general level, with the level of agreement representing social licence. As Table 2 shows, identifying families that might need support, catching problems early on, efficiently targeting services, and identifying risk of child abuse, etc., were seen as acceptable reasons for joining together administrative records, with over 80 per cent of parents agreeing (with a consensus baseline of 50 per cent). It is important to note, however, that, although still representing social licence, the levels of agreeing are uneven between different social groups of parents (Table 2). In particular, it is variably lower among lone parents and younger parents in the sample, and consistently lower among Black parents, especially for identification and targeting of families.

INSERT TABLE 2 HERE

The picture for the granting of social licence changes when it comes to considering the use of data linkage by specific public services. We asked parents about whether or not they trusted particular organisations to join together administrative records to identify families for targetted public services. Trust is an important element of social licence. It is bound up with considerations of information being used in legitimate and fair ways by agencies. As Peter Leonard (2018) points out, different social groups are likely to see the relationship between trust and fairness in different ways. This is because they are not all positioned in the same way in society and thus in relation to intervention services, which are targeted at particular types of families. This forms part of the wider collective contextual dynamics with which parents’ granting or withholding of social licence for data linkage and predictive analytics articulates. Indeed, while parents may grant social licence for joining families’ administrative records to enable early intervention in a generalised sense, when it comes to considering the use of data linkage by specific public services there are differences between the services concerned in whether or not they extend trust, and importantly any trust and legitimacy extended is variable between different social groups of parents.

INSERT TABLE 3 HERE

Looking at Table 3, around half of parents overall said they trusted children’s social work teams, local council education services, early years services, police and criminal justice, or immigration services to join together administrative records for targeting families, with only social work and early years services (just) achieving the social licence consensus baseline of 50 per cent or above. For the other services, there was no social licence for data linkage. Looking deeper into the patterning of this, trust in services to join together information mirrored the overall constrained social licence pattern among parents who were in managerial and professional occupations, and had higher levels of qualifications and higher incomes, but there was even less likely to be a social licence consensus for data linkage for operational purposes among marginalised social groups of parents.

In particular, Black parents do not hold a consensus of trust in any public services concerning their use of data linkage, especially not police and criminal justice and immigration services where trust drops to under a third (28 and 24 per cent respectively). These concerning figures likely reflect the far lower levels of confidence that Black people in Britain have in the police in comparison with White and Asian counterparts (ONS, 2020) – another feature of the wider contextual dynamics for social licence. But specifically in relation to data linkage, the Black parents also held an overwhelming consensus that information collected about services users is not always accurate, with 79 per cent disagreeing with the statement that joined together administrative records provide factual and unbiased information for delivering services. They also judged that data linkage leads to discrimination against some families (57 per cent agreed), and that using families’ administrative records may discourage them from accessing services when they need them (62 per cent agreed). Clearly, there is no social licence for data linkage for early intervention among Black parents.

Other marginalised social groups of parents did not extend social licence to aspects of data linkage for early intervention either, notably lone parents, younger parents, and parents in larger families. Table 3 shows that there is no social licence among lone parents for many public services to link data linkage. Rather, there is a consensus that the information collected about services users is not always accurate (63 per cent), that data linkage will lead to discrimination against some families (52 per cent), and that it can put families off accessing services when they need them (52 per cent) (all with a consensus threshold of 50 per cent). Similarly, younger parents did not extend social licence for data linkage by many public services, and held consensus that the information collected about services users is not always accurate (57 per cent) and that joined together administrative records leads to discrimination against some families (60 per cent). Parents in larger families did grant social licence for a variety of public services to use data linkage, being more trusting than the sample as a whole (see Table 3), but nonetheless hold a consensus that the information collected about services users is not always accurate (62 per cent), and that data linkage can put families off accessing services when they need them (54 per cent) (all with a consensus threshold of 50 per cent).

Early intervention relies on predictive analytics, and the necessary operational data linkage practices are often carried out as part of public-private collaborations. Data warehousing or hubs or lakes, and data integration and analytics is outsourced to multinational data analytic companies, contracted by local authorities for use of their commercial systems of profiling and algorithmic risk assessments (Redden, 2020; Edwards et al., 2021). Yet there is no parental social licence for outsourcing to commercial companies for the use of algorithms to support targetting of public services, with a consensus against this among the sample as a whole (55 per cent with a consensus threshold of 50 per cent). This consensus about data analytic company involvement holds roughly similarly across social groups, with 57 per cent of parents in the higher occupation, qualification and income group, 60 per cent of Black parents, 62 per cent of lone parents, 51 per cent of younger parents, and 53 per cent of parents in households with five or more members regarding it as unacceptable.

Overall then, data linkage for the operational purpose of early intervention is acceptable to parents of dependent children at an abstract level, but they are more circumspect when considering the specifics of trusting particular public services to do this, and there is no social licence for the involvement of commercial companies. There is less social licence for joining together families’ administrative records among marginalised social groups of parents, with some holding little trust in public services implementing data linkage. This points to a worrying level of distrust toward government and public services among those in society who are most marginalised and whose families are likely to be identified for service interventions. This lack of social licence should be a concern for policy prescriptions about sharing and linking families’ administrative records for early intervention.

**Concluding implications**

The way forward in ensuring public trust in data-steered policy through linking administrative records and algorithmic analysis, and by implication social licence for early intervention, is posed by government and associated bodies as the provision of information on the benefits and transparency about their use. There is little discussion of this as an active and sustained process; as a meaningful engagement of the subjects of data linkage in a dialogue about setting the parameters of its curation and use by and between services. Rather, public understanding strategies largely are conceived of as a one-way, top down, exercise, raising public awareness rather than reflecting people’s concerns within governance frameworks (Leonard 2018; Shaw et al., 2020; Waller & Waller, 2020). This sort of didactic explanatory ‘involvement’ is highly unlikely to lead to social licence.

Our analysis of parental social licence for preventive early intervention shows informed consent to use of their administrative records is important to parents of dependent children. Informing and asking for consent raises the possibility that consent may be withheld by some parents, given the concerns about bias and discrimination among marginalised groups of parents in particular. Indeed, there is less acceptance of public services using data linkage among parents from these groups. Government needs to be transparent about how they link and use families’ data and to gain parents’ informed consent. Addressing this through generalised policy messages about the merits of data linkage has the potential to bolster already existing social licence among parents in higher occupation, qualification and income groups, while running the risk of further disengagement, alienation and avoidance of essential public services among marginalised groups of parents who, collectively, are more likely to be the focus of identification for early intervention. Policy-makers need to realise that information about this use of data and efforts towards obtaining informed consent are likely to be received and judged quite differently among different social groups of parents[[3]](#footnote-3).

Also of concern in the embracing of transparency as an answer to the problem of social licence is that this carries its own perils in simultaneously occluding and embedding social divisions and inequalities in the assertion of control through monitoring (Monahan, 2021). Rather than data warehousing and predictive risk modelling being a neutral knowledge-generating (and cost-cutting) exercise, the concerns of marginalised groups of parents about the accuracy of data and discrimination are not groundless. Indeed, the issue of social licence among marginalised parents that we draw attention to above seems all the more pressing in a context where, as noted, there are bias and errors in the data sources that are merged, and in the design of data modelling and predictive analytics applied. Such discrimination contributes to further inequalities and lack of trust among marginalised groups of parents of dependent children. Moreover, social legitimacy is put at risk for an operational practice that has shown little evidence of efficacy. Amy Edwards and colleagues’ review of evidence-based early help in the UK (2021) leads them to suggest that what is easiest to measure is pursued at the expense of addressing the complexity and dynamics at play in family life, including poverty, that could make a difference. Further, studies drawing on extensive longitudinal data to test predictive modelling techniques (e.g. Clayton et al., 2020; Salganik et al., 2020) are finding a worrying lack of accuracy in forecasting future outcomes, with one international mass academic collaboration concluding:

Policymakers using predictive models in settings such as criminal justice and child-protective services should be concerned by these results. In addition to the many serious legal and ethical questions raised by using predictive models for decision-making, the results of the Fragile Families Challenge raise questions about the absolute level of predictive performance that is possible for some life outcomes, even with a rich dataset. (Salganik et al., 2020: 8402)

In conclusion, policy-makers need to go beyond exhortations for enacting and improving transparency to enter into robust discussions about the risks as well as the benefits of operational data linkage and predictive analytics for early intervention, and to consider and address unintended social consequences. It is vital that they pay real and meaningful attention to the extent of social licence and trust for data linkage among marginalised groups of parents in society. At a collective level, these are parents who are most likely to be implicated in such efforts towards early intervention, disproportionately affected by data linking and predictive analytic activities but lacking any mechanism to have their concerns taken seriously enough to reshape the speeding up of data-driven policy and service delivery. Implementation of sharing and linking of data amongst public services working with children and families has the potential to undermine even further social legitimacy and trust among marginalised social groups of parents, with consequences for a cohesive and equal society. Beginning from a position that asks whether data linkage and predictive analytics for early intervention will undermine social trust, then a responsible question is raised for policymakers about whether or not it should be done at all.

**Data Availability**

The data used in this article was commissioned from NatCen through their online and telephone probability-based panel survey. Data from the survey will be deposited in and available on registration with the UK Data Archive at the end of the research project, in September 2022.

**Funding Statement**

The research on which this article is based is funded by the UKRI Economic and Social Research Council under grant number ES/T001623/1.

**Competing Interests**

Competing interest: Rosalind Edwards, Val Gillies and Sarah Gorin declare none.

**Author Contributions**

Conceptualisation – Rosalind Edwards and Val Gillies

Data curation – equal contributions by all authors

Formal analysis – equal contributions by all authors

Funding acquisition – Rosalind Edwards and Val Gillies

Investigation – Rosalind Edwards and Val Gillies, support by Sarah Gorin

Methodology – Rosalind Edwards and Val Gillies, support by Sarah Gorin

Project administration – Rosalind Edwards as lead, support by other authors

Resources – Rosalind Edwards as lead, support by other authors

Software – not applicable

Supervision – not applicable

Validation – equal contributions by all authors

Visualization – not applicable

Writing original draft – Rosalind Edwards as lead, support by other authors

Writing review and editing – Rosalind Edwards as lead, support by other authors

**References**

Amnesty International United Kingdom Section (2018) *Trapped in the Matrix: Secrecy, Stigma, and Bias in the Met’s Gangs Database*: <https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf>

Answell, B., Bauer, M.W., Gingrich, J., & Stilgoe, J. (2021) Coping with Covid: two-wave survey, *RPubs*, 23 February: <https://rpubs.com/benwansell/729135>

ARI Working group 3 (2020) *Rebuilding a Resilient Britain: Trust in Public Institutions*. ARI Report 3: <https://www.upen.ac.uk/go_science/RRB3_TrustInPublicInstitutions.pdf>

Benjamin, R. (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*, Cambridge: Polity Press.

British Academy (2021) The COVID Decade: Understanding the long-term societal impacts of COVID-19: <https://www.thebritishacademy.ac.uk/documents/3238/COVID-decade-understanding-long-term-societal-impacts-COVID-19.pdf>

Caldicott, F. (2016) *Review of Data Security, Consent and Opt-Outs*, London: Department of Health

Carter, P., Laurie, G.T. & Dixon-Woods, M. (2015) The social licence for research: why *care.data* ran into trouble, *Journal of Medical Ethics*, 41: 404-409.

Centre for Data Ethics and Innovation (2020) *Review into Bias in Algorithmic Decision-Making*: <https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making>

Clayton V, Sanders M, Schoenwald E, Surkis L & Gibbons D (2020) *Machine Learning in Children’s Services*, What Works for Children’s Social Care: <https://whatworks-csc.org.uk/wp-content/uploads/WWCSC_machine_learning_in_childrens_services_does_it_work_Sep_2020_Accessible.pdf>

Crossley, S. (2018) *Troublemakers: The Construction of ‘Troubled Families’ as a Social Problem*, Bristol: Policy Press.

Data Futures Partnership (2017) *A Path to Social Licence: Guidelines for Trusted Data Use*, available at <https://trusteddata.co.nz/>.

DDCMS (Department for Digital, Cultural, Media and Sport) (2020) *National Data Strategy*: <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>

DHSC (Department of Health and Social Care) (2021) *The Best Start for Life: A Vision for the 1001 Critical Days.* The Early Years Health Development Review Report: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973112/The_best_start_for_life_a_vision_for_the_1_001_critical_days.pdf>

Dencik, L., Redden, J., Hintz, A. & Warne, H. (2019) The ‘golden view’: data-driven governance in the scoring society, *Internet Policy Review*, 8: 2: 1-24.

Easton, S. (2017) Expanding open data requires social license: information mandarins, *The Mandarin*: <https://www.themandarin.com.au/74424-two-key-federal-information-custodians-weigh-in-on-proposed-data-and-privacy-reforms/>

Edwards, A., Gharbi, G., Berry, A. & Duschinsky, R. (2021) *Support and Strengthening Families Through Provision of Early Help: A Rapid Review of Evidence*, National Children’s Bureau: <https://www.ncb.org.uk/sites/default/files/uploads/attachments/20210513_Rapid%20Review_Full%20Report%20-%20FINAL.pdf>

Edwards, R. & Gillies, V (2004) Support in parenting: values and consensus concerning who to turn to, *Journal of Social Policy* 33: 4, 627-647.

Edwards, R., Gillies, V. and Gorin, S. (2021) Problem-solving for problem-solving: data analytics to identify families for service intervention, *Critical Social Policy* <https://journals.sagepub.com/doi/full/10.1177/02610183211020294>

Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor.* New York: St Martin’s Press.

Finch, J. & Mason, J. (1991) Obligations of kinship in contemporary Britain: is there normative agreement?, *British Journal of Sociology*, 42: 3, 345–367.

Gillies, V., Edwards, R. & Horsley, N. (2017) *Challenging the Politics of Early Intervention: Who's 'Saving' Children and Why*, Bristol: Policy Press.

Gulliver, P., Jonas, M., McIntosh, T., Franslow, J. & Waayer, D. (2018) Surveys, social licence and the integrated data infrastructure, *Aotearoa New Zealand Social Work*, 30(3): 57-71.

House of Lords Public Services Committee (2020) *A Critical Juncture for Public Services: Lessons from COVID-19*: <https://committees.parliament.uk/publications/3438/documents/32865/default/>

Hughes, E.C. (1958) *Men and Their Work*, Glencoe Ill.: Free Press.

Information Commissioner’s Office (2020) *Data Sharing Code of Practice*: <https://ico.org.uk/media/for-organisations/data-sharing-a-code-of-practice-1-0.pdf>

Johnson, P., Joyce, R. & Platt, L. (2021) *The IFS Deaton Review of Inequalities: a New Year’s Message*, Institute for Fiscal Studies: <https://www.ifs.org.uk/inequality/the-ifs-deaton-review-of-inequalities-a-new-years-message/>

Jupp, E. (2017) Families, policy and place in times of austerity, *Historical Geography* 49(3): 266-272.

Keddell E (2014) The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: child abuse prevention or neo-liberal tool?, *Critical Social Policy* 35(1): 69-88.

Kennedy, H., Oman, S., Taylor, M., Bates, J. & Steedman, R. (2020) *Public Understanding and Perceptions of Data Practices*: <https://livingwithdata.org/project/wp-content/uploads/2020/05/living-with-data-2020-review-of-existing-research.pdf>

Leonard, P.G. (2018) Social Licence and Digital Trust in Data-Driven Applications and AI: A Problem Statement and Possible Solutions. Available at SSRN: [https://ssrn.com/abstract=3261228](https://ssrn.com/abstract%3D3261228)

McIntyre, N. & Pegg, D. (2018) Councils use 377,000 people’s data in efforts to predict child abuse, *The Guardian*, 16 September: <https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse>

Messenger, C. & Molloy, D. (2014) *Getting It Right For Families: A Review of Integrated Systems and Promising Practice in the Early Years Field*, Early Intervention Foundation: <https://www.eif.org.uk/report/getting-it-right-for-families-a-review-of-integrated-systems-and-promising-practice-in-the-early-years>

MHCLG (Ministry of Housing, Communities and Local Government) (2020), *Financial Framework for the Troubled Families Programme*: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/922390/Troubled_Families_Financial_Framework_2020.pdf>

MHCLG (Ministry of Housing, Communities and Local Government) (2021a), *Local Data Accelerator Fund for Children and Families: Prospectus*: <https://www.gov.uk/government/publications/local-data-accelerator-fund-for-children-and-families>

MHCLG (Ministry of Housing, Communities and Local Government) (2021b), Press release: Next phase of £165 million programme for vulnerable families launched: <https://www.gov.uk/government/news/next-phase-of-165-million-programme-for-vulnerable-families-launched>

Monahan, T. (2021) Reckoning with COVID, racial violence, and the perilous pursuit of transparency, *Surveillance & Society*, 19(1): 1-10.

Moody, G. & Lugg, F. (2017) “Public understanding of, and attitudes towards, data linkage in research: *IJPDS*, Issue 1, Vol 1:073, Proceedings of the IPDLN Conference (August 2016)”, *International Journal of Population Data Science*, 1(1). doi: 10.23889/ijpds.v1i1.92.

NatCen (2018) *Public Attitudes to Data Linkage*: A report prepared for University College London by NatCen Social Research: <https://natcen.ac.uk/our-research/research/public-attitudes-to-data-linkage/>

Nielsen Company (2018) *A Social Licence Approach to Trust. A Close-Up on Trust*: <https://www.stats.govt.nz/corporate/measuring-stats-nzs-social-licence>

NL Digitaal Government (2019) *Data Agenda Government*: <https://www.nldigitalgovernment.nl/wp-content/uploads/sites/11/2019/04/data-agenda-government.pdf>

Office for National Statistics (2020) *Confidence in the local police*: https://www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/policing/confidence-in-the-local-police/latest

Office for National Statistics (2021) *Coronavirus and vaccine hesitancy, Great Britain: 13 January to 7 February 2021*: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandvaccinehesitancygreatbritain/13januaryto7february2021>

Office for Statistics Regulation (2018) *Joining Up Data for Better Statistics*: <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/09/Data-Linkage-Joining-Up-Data.pdf>

Paprica, P.A., Nunes de Melo, M. and Schull, M.J. (2019) Social licence and the general public’s attitudes towards research based on linked administrative health data: a qualitative study, *CMAJ Open*, 7(1): E40-E46.

Privacy Council Office, Canada (2018) *A Data Strategy Roadmap for the Federal Public Service*: <https://www.canada.ca/content/dam/pco-bcp/documents/clk/Data_Strategy_Roadmap_ENG.pdf>

Redden, J., Dencik, L. & Warne, H. (2020) Datafied child welfare services: unpacking, politics, economics and power. *Policy Studies:* <https://doi.org/10.1080/01442872.2020.1724928>.

Salganik, M.J., Lundberg, I., Kindel, A.T. and 115 others (2020) Measuring the predictability of life outcomes with a scientific mass collaboration. Proceedings of the National Academy of Sciences, 117(15): 8398-8403: <https://www.pnas.org/content/pnas/117/15/8398.full.pdf>.

Scourfield, J., Corliss, C., Wijedasa, D., Robling, M. & Clayton, V. (2019*) Overview of Administrative Data on Children’s Social Care in England*, What Works for Children’s Social Care: <https://whatworks-csc.org.uk/wp-content/uploads/WWCSC_Overview_of_administrative_data_Oct_2019.pdf>

Shaw, J.A., Sethi, N. & Cassel, C.K. Social licence for use of big data in the OVID-19 era, *NPJ Digital Medicine*, 3: 128: <https://www.nature.com/articles/s41746-020-00342-y>

Stats NZ (2018) *Algorithm Assessment Report*: <https://data.govt.nz/use-data/analyse-data/governmentalgorithm-transparency>

Thomson, I. & Boutilier, R. (2011) The social licence to operate, in *SME Mining Engineering Handbook*, Englewood, Co.: Society for Mining, Metallurgy and Exploration.

Vannier Ducasse, H. (2020) Predictive risk modelling and the mistaken equation of socio-conomic disadvantage with risk of maltreatment, *British Journal of Social Work*: <https://doi.org/10.1093/bjsw/bcaa182>.

van Zoonen, L. (2020) Data governance and citizen participation in the digital welfare state, *Data and Policy*, 2: e10: <https://www.cambridge.org/core/journals/data-and-policy/article/data-governance-and-citizen-participation-in-the-digital-welfare-state/CCF2E1914C7E2D4593D550C6BC9E4C70>

Wachter, R.M. & Cassel, C.K. (2020) Sharing health care data with digital giants: overcoming obstacles and reaping benefits while protecting patients, *JAMA* 323: 507–508.

Wachter-Boettcher, S. (2017) *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*, New York, NY: WW Norton & Company.

Waller, M. & Waller, P. (2020) Why Predictive Algorithms are So Risky for Public Sector Bodies: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3716166>

White, S. & Wastell, D. (2017) The rise and rise of prevention science in UK family welfare: surveillance gets under the skin, *Families, Relationships and Societies*, 6(3): 427-446.

World Economic Forum (2018) White Paper: *How to Prevent Discriminatory Outcomes in Machine Learning*: <http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf>

Wroe, L.E. (2021) Young people and ‘county lines’: contextual and social account, *Journal of Children’s Services*, 16(1): 39-55.

**TABLES**

**Table 1: Awareness of digital administrative records and of data linkage by ethnicity, occupation, education and income**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Awareness that: | *Black and Black British (% yes)* | *White British**(% yes)* | *Managerial /**Professional* *(% yes)* | *Semi-**routine/**Routine* *(% yes)* | *Degree level +* *(% yes)* | *No qualifications**(% yes)* | *>£3700+ pm**(% yes)* | *<£1200 pm**(% yes)* |
| administrative records are collected and digitally stored by government departments and service providers | 81 | 73 | 80 | 68 | 79 | 54 | 80 | 60 |
| administrative records can be joined together to find out more about individual families | 67 | 53 | 61 | 54 | 60 | 39 | 61 | 40 |

**Table 2: Acceptability of reasons for joining together administrative records**

*50% consensus threshold*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reasons: | All sample*(%)* | Higher occupation/ qualification/ income parents *(%)* | Black parents*(%)* | Lone parents*(%)* | Younger parents (18-29)*(%)* | Parents in 5+ member households*(%)* |
| Identifying families that might need support for services even if they have not asked for it | 82 | 86 | 62 | 74 | 78 | 80 |
| Saving time and money by catching family problems early | 83 | 87 | 71 | 78 | 84 | 85 |
| Making services more efficient by identifying families that need help and targeting services at them | 84 | 87 | 69 | 84 | 73 | 79 |
| Identifying families where children could be at risk of abuse | 90 | 94 | 85 | 91 | 84 | 89 |

**Table 3: Trust in organisations to join together administrative records to identify families to target public services**

*50% consensus threshold*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Organisations: | All sample*(%)* | Higher occupation/ qualification/ income *(%)* | Black parents*(%)* | Lone parents*(%)* | Younger parents (18-29)*(%)* | Parents in 5+ member households*(%)* |
| Children’s social work teams | 55 | 55 | 38 | 52 | 50 | 55 |
| Local council education services | 47 | 48 | 39 | 43 | 48 | 51 |
| Early years services | 50 | 52 | 43 | 47 | 45 | 52 |
| Police and criminal justice | 49 | 49 | 28 | 43 | 50 | 54 |
| Immigration | 35 | 33 | 24 | 31 | 34 | 38 |

1. <https://www.eif.org.uk/about/who-is-eif#:~:text=The%20Early%20Intervention%20Foundation%20(EIF,risk%20of%20experiencing%20poor%20outcomes>. [↑](#footnote-ref-1)
2. The survey discussed in this article received ethical approval from the University of Southampton, ID 56997. [↑](#footnote-ref-2)
3. A segmented approach has been recommended with respect to different services because of the ‘chilling’ effects of mistrust in particular public institutions by marginalised and disadvantaged groups, so that distrust of one service such as the police does not ‘chill’ into distrust of data surveillance by another service such as public health – albeit its own messages about re/building trust are generalised (ARI Working Group 3, 2020). [↑](#footnote-ref-3)