

RESEARCH ARTICLE

Data linkage for early intervention in the UK: Parental social license and social divisions

Rosalind Edwards^{1,*} , Val Gillies²  and Sarah Gorin¹ 

¹Department for SSPC, University of Southampton, Southampton, United Kingdom

²Department for Social Sciences, University of Westminster, London, United Kingdom

*Corresponding author. E-mail: r.s.edwards@soton.ac.uk

Received: 22 May 2021; **Revised:** 15 September 2021; **Accepted:** 19 November 2021

Key words: early intervention; operational data linkage; parents; predictive analytics; social divisions; social license

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 license 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 license 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 marginalized 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 license 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 marginalized social groups. Implementation of operational data linkage among 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 license among parents in higher occupation, qualification, and income groups while generating further disengagement and avoidance of public services among marginalized parents. Rather, meaningful dialogue that shapes the parameters of data linkage is required.

1. 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 about citizens—health, education, social care, police, housing, immigration, taxation and social security records, and so forth—and operationalizing them through algorithmic data analytics is championed as offering the possibility of improved and more efficient public service delivery, and enabling predictive risk modeling to pre-empt problems and improve outcomes (e.g., NL Digitaal Government, 2019; Privacy 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). Shaw *et al.* (2020) argue that data sharing, data linkage, and the application of analytics need to earn “social license”—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 van Zoonen argues, moves toward data-driven social policy largely take place out of political and social view. Data linkage and predictive analytic practices are centered 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, pp. e10–e19). 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” (Department for Digital, Cultural, Media and Sport, 2020). The House of Lords Public Services Committee (2020) similarly has emphasized 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 (Department for Digital, Cultural, Media and Sport, 2020, p. 43)

Yet, it is all families who are implicated in across-the-board data sharing, data linking, and application of predictive modeling, not just the families with “vulnerable children.” Population-level data comprise data from individual families, from all families nationally or from all living in a particular local authority area. These data are 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; Office for National Statistics, 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, pp. 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 recognizing 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 marginalized groups to more disadvantage, with built-in discrimination in databases. There are bias and errors in the data sources that are merged, and in the design of data modeling and predictive analytics applied. Particular subgroups 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 datasets. In the UK, for example, predictive risk modeling used in child protection embeds an equation of socioeconomic disadvantage with risk, discriminating against poor families (Vannier Ducasse, 2020), while attention has been drawn to the unsupported overidentification of young Black men in what amounts to digital racialized 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 analytics 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 directed at 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 modeling to identify and target families. We then outline the concept of social license 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.

2. Data Linkage for Early Intervention in the UK

The way that parents bring up their children is a long-standing 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, toward a focus on children and families at risk, and early intervention in particular families' lifestyles and behavior. Early intervention aims to pre-empt rather than react; to prevent any risk of social, educational, health, and behavioral 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 Their focus on evaluating various parenting skills delivery packages has more recently been augmented by attention to the potential of administrative data for targeting and tracking early intervention (Scourfield et al., 2019). Such data-driven family policy initiatives are legitimated by what White and Wastell (2017) refer to as "prevention

¹ [https://www.eif.org.uk/about/who-is-eif#:~:text=The%20Early%20Intervention%20Foundation%20\(EIF,risk%20of%20experiencing%20poor%20outcomes.](https://www.eif.org.uk/about/who-is-eif#:~:text=The%20Early%20Intervention%20Foundation%20(EIF,risk%20of%20experiencing%20poor%20outcomes.)

science,” a mix of technological, biological, and behavioral sciences and morality that is fueling 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 endeavor 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 are 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 and Molloy, 2014, p. 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 (Department of Health and Social Care, 2021). Central government has initiated a Local Data Accelerator Fund for Children and Families (Ministry of Housing, Communities and Local Government, 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” (Ministry of Housing, Communities and Local Government, 2021a, p. 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, and 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 multi-agency data “warehouse” as an analytic hub to help predict children at risk of criminal or sexual exploitation, becoming not in education, employment, or training (NEET), 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 and 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 (Ministry of Housing, Communities and Local Government, 2021a, p. 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 multiagency data sharing and linking to identify and intervene was heralded by the Troubled Families program, reborn as Supporting Families in its latest phase (Ministry of Housing, Communities and Local Government, 2021b). The Troubled Families program 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 program, because it 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 program range widely across the domains of social security, housing, education, health, social services, police and criminal justice, and other public provisions (Ministry of Housing, Communities and Local Government, 2020), and under the Supporting Families version of the program, the emphasis is on “building stronger data,” posed as part of a “moral mission” to support “vulnerable” families (Ministry of Housing, Communities and Local Government, 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, and so forth 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 anonymized data for research purposes, and/or involved general population focus group discussion (e.g., Moody and Lugg, 2017; NatCen, 2018). All parents are stakeholders in the use of administrative records for data linkage and predictive analytics for targeting 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 license for them needs to be ascertained.

3. Social License

Considerations of data linkage and predictive analytics for operational service intervention have been turning to the concept of social license in a context where there is concern internationally and nationally, about the existence and sustenance of trust for these practices (e.g., Caldicott, 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, p. 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 license 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, Carter et al. (2015) drew on the concept of social license 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 license 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 license 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 organization plans to share data with Google for advanced analytics (Wachter and Cassel, 2020). Legal license is not necessarily a foundation for social license.

Crucially, social license 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 theorizing about the relationship between professionals and society. Notably, Hughes (1958) conceptualized professional groups as socially licensed 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 license is the examination of corporate responsibility in extraction industries, where the environment or communities may be harmed by their activities (e.g., Thomson and Boutilier, 2011). In this field, there is a strand of literature that stresses the need to build and maintain a consensus of social license among relevant stakeholders, especially the particular population affected, even if operating within the law. Social license, 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 license.

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 license among the population implicated: parents of dependent children. If policymakers and service providers are to begin to engage with social license as a dynamic process, there needs to be some knowledge of the bases from which they are starting. If social license already exists and engagement is directed toward 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 license knowledge base may mean that policymakers and service providers understand that a policy action should not be pursued because of unintended consequences.

4. Our Survey

Our Parental Social License for Data Linkage for Service Intervention project² investigates the dynamics of social license 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 license 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 license. 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 nonresponse. 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 an analysis of subsequent government, corporate, advocacy, and media publications, and refined in discussion with NatCen. Randomization 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, and 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 governments, data analytic companies, charities and advocacy groups, and mainstream media that related to early intervention (see Edwards et al., 2021b).

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 al., 2021b) 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 behavior. 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% were mothers and 43% were fathers. Looking at the various household types that parents lived in, the majority were families of two

² The survey discussed in this article received ethical approval from the University of Southampton (ID 56997).

parents with children, at 74.5%. Eleven per cent were lone parents, and 24% 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% of the parents were in managerial and professional occupations, and 50% were educated to degree level or above, while 38.5% were in lower semi-routine and routine occupations and 10% had no qualifications. The majority of parents, 65%, earned £3700 per month or under, with 60% of the sample owning their homes, while the remainder rented, split between local authority/housing association or privately. Ethnically, 73% of parents were White British, with 19% of our sample from minority ethnic groups, 5% 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 license for operational data linkage among Black and other marginalized groups of parents stand out.

The concept of social license 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 license. 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 analyzing 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 license consensus (see Edwards and Gillies, 2004; Finch and Mason, 1991). For a two-response-option question, rather than more than 50% being taken as a consensus, if one of the options gathers half as many responses again (75%), then where this was a positive response to data linkage and predictive analytics we took that to represent a widespread granting of social license. For a three-option question, we took a 50% or greater positive response as indicative of social license. We conducted the consensus baseline analysis at the level of the sample of parents as a whole, and also looked within this, to subpopulation groups of parents (family types, social class indicators, and ethnic groups) to see if they may or may not reach a consensus baseline for granting social license at an appreciably higher or lower level than the rest.

5. 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 centered 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 they are used (e.g., Centre for Data Ethics and Innovation, 2020; Information Commissioner's Office, 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 they were aware that administrative records are collected and digitally stored by government departments and public service providers (72%), 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%). 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 occupation, education, and income were more aware than other parents (see Table 1).

There was consensus among the parents that families generally do not know or understand how their administrative records are used, with 60% making this judgement (with a consensus threshold of 50%). There was overwhelming agreement that Government should publicize that they are joining together administrative records about families, and how they use that information (81% with a consensus threshold

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

of 50%). 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% with a consensus threshold of 50%). Some marginalized 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%).

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.

6. 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 governments, 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, optimizing existing resources and making savings through prevention rather than crisis management (see the discourse analysis in Edwards et al., 2021b). These rationales for undertaking data linkage and risk modeling 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 service 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., 2021b). 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 license 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 records to enable early intervention at a general level, with the level of agreement representing social license. As shown in Table 2, identifying families that might need support, catching problems early on, efficiently targeting services, and identifying risk of child abuse, and so forth, was seen as acceptable reasons for joining together administrative records, with over 80% of parents agreeing (with a consensus baseline of 50%). It is important to note, however, that, although still representing social license, 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.

The picture for the granting of social license 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 organizations to join together administrative records to identify families for targeted public services. Trust is an important element of social license. It is bound up with considerations of information being used in legitimate and fair ways by agencies. As 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

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

particular types of families. This forms part of the wider collective contextual dynamics with which parents' granting or withholding of social license for data linkage and predictive analytics articulates. Indeed, while parents may grant social license for joining families' administrative records to enable early intervention in a generalized 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.

Looking at Table 3, around half of parents overall said that 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 license consensus baseline of 50% or above. For the other services, there was no social license for data linkage. Looking deeper into the patterning of this, trust in services to join together information mirrored the overall constrained social license 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 license consensus for data linkage for operational purposes among marginalized 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%, 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 (Office for National Statistics, 2020)—another feature of the wider contextual dynamics for social license. 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% 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% agreed), and that using families' administrative records may discourage them from accessing

Table 3. Trust in organizations to join together administrative records to identify families to target public services (50% consensus threshold)

Organizations	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 services	35	33	24	31	34	38

services when they need them (62% agreed). Clearly, there is no social license for data linkage for early intervention among Black parents.

Other marginalized social groups of parents did not extend social license 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 license among lone parents for many public services to link data. Rather, there is a consensus that the information collected about services users is not always accurate (63%), that data linkage will lead to discrimination against some families (52%), and that it can put families off accessing services when they need them (52%; all with a consensus threshold of 50%). Similarly, younger parents did not extend social license for data linkage by many public services, and held consensus that the information collected about services users is not always accurate (57%) and that joined together administrative records leads to discrimination against some families (60%). Parents in larger families did grant social license 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%), and that data linkage can put families off accessing services when they need them (54%; all with a consensus threshold of 50%).

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, are outsourced to multinational data analytic companies, contracted by local authorities for use of their commercial systems of profiling and algorithmic risk assessments (Edwards et al., 2021b; Redden et al., 2020). Yet there is no parental social license for outsourcing to commercial companies for the use of algorithms to support targeting of public services, with a consensus against this among the sample as a whole (55% with a consensus threshold of 50%). This consensus about data analytic company involvement holds roughly similarly across social groups, with 57% of parents in the higher occupation, qualification, and income group, 60% of Black parents, 62% of lone parents, 51% of younger parents, and 53% 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 license for the involvement of commercial companies. There is less social license for joining together families' administrative records among marginalized 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 marginalized and whose families are likely to be identified for service interventions. This lack of social license should be a concern for policy prescriptions about sharing and linking families' administrative records for early intervention.

7. Concluding Implications

The way forward in ensuring public trust in data-steered policy through linking administrative records and algorithmic analysis, and by implication social license 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 and Waller, 2020). This sort of didactic explanatory "involvement" is highly unlikely to lead to social license.

Our analysis of parental social license 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 marginalized 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 generalized policy messages about the merits of data linkage has the potential to bolster already existing social license among parents in higher occupation, qualification, and income groups while running the risk of further disengagement, alienation, and avoidance of essential public services among marginalized groups of parents who, collectively, are more likely to be the focus of identification for early intervention. Policymakers need to realize that information about this use of data and efforts toward obtaining informed consent are likely to be received and judged quite differently among different social groups of parents.³

Also of concern in the embracing of transparency as an answer to the problem of social license 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 modeling being a neutral knowledge-generating (and cost-cutting) exercise, the concerns of marginalized groups of parents about the accuracy of data and discrimination are not groundless. Indeed, the issue of social license among marginalized 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 modeling and predictive analytics applied. Such discrimination contributes to further inequalities and lack of trust among marginalized groups of parents of dependent children. Moreover, social legitimacy is put at risk for an operational practice that has shown little evidence of efficacy. Edwards et al.'s (2021a) review of evidence-based early help in the UK 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. Furthermore, studies drawing on extensive longitudinal data to test predictive modeling techniques (e.g., Clayton et al., 2020; Salganik et al., 2020)

³ A segmented approach has been recommended with respect to different services because of the "chilling" effects of mistrust, in particular, public institutions by marginalized 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—although its own messages about re/building trust are generalized (ARI Working Group 3, 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, p. 8402).

In conclusion, policymakers 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 license and trust for data linkage among marginalized groups of parents in society. At a collective level, these are parents who are most likely to be implicated in such efforts toward 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 among public services working with children and families has the potential to undermine even further social legitimacy and trust among marginalized 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.

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. The authors declare none.

Author Contributions. Conceptualization: R.E. and V.G.; Data curation: equal contributions by all authors; Formal analysis: equal contributions by all authors; Funding acquisition: R.E. and V.G.; Investigation: R.E. and V.G., support by S.G.; Methodology: R.E. and V.G., support by S.G.; Project administration: R.E. as lead, support by other authors; Resources: R.E. as lead, support by other authors; Validation: equal contributions by all authors; Writing—original draft: R.E. as lead, support by other authors; Writing—review & editing: R.E. as lead, support by other authors.

Data Availability Statement. The data used in this article were 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.

References

- Amnesty International United Kingdom Section** (2018) *Trapped in the Matrix: Secrecy, Stigma, and Bias in the Met's Gangs Database*. Available at <https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf>.
- Ansell B, Bauer MW, Gingrich J and Stilgoe J** (2021) *Coping with Covid: Two-Wave Survey*, RPubS, 23 February. Available at <https://rpubs.com/benwansell/729135>.
- ARI Working Group 3** (2020) *Rebuilding a Resilient Britain: Trust in Public Institutions*. ARI Report 3. Available at 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*. Available at <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 GT and 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*. Available at <https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making>.

- Clayton V, Sanders M, Schoenwald E, Surkis L and Gibbons D** (2020) *Machine Learning in Children's Services*, What Works for Children's Social Care. Available at 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/>.
- Dencik L, Redden J, Hintz A and Warne H** (2019) The "golden view": Data-driven governance in the scoring society. *Internet Policy Review* 8(2), 1–24.
- Department for Digital, Cultural, Media and Sport** (2020) *National Data Strategy*. Available at <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>.
- 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. Available at 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.
- Easton S** (2017) Expanding open data requires social license: Information mandarins, *The Mandarin*. Available at <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 and Duschinsky R** (2021a) *Support and Strengthening Families Through Provision of Early Help: A Rapid Review of Evidence*, National Children's Bureau. Available at https://www.ncb.org.uk/sites/default/files/uploads/attachments/20210513_Rapid%20Review_Full%20Report%20-%20FINAL.pdf.
- Edwards R and 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** (2021b) Problem-solving for problem-solving: Data analytics to identify families for service intervention, *Critical Social Policy* <https://doi.org/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 and 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 and 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 and 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*. Available at <https://committees.parliament.uk/publications/3438/documents/32865/default/>.
- Hughes EC** (1958) *Men and Their Work*. Glencoe, IL: Free Press.
- Information Commissioner's Office** (2020) *Data Sharing Code of Practice*. Available at <https://ico.org.uk/media/for-organisations/data-sharing-a-code-of-practice-1-0.pdf>.
- Johnson P, Joyce R and Platt L** (2021) *The IFS Deaton Review of Inequalities: A New Year's Message*, Institute for Fiscal Studies. Available at <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 and Steedman R** (2020) *Public Understanding and Perceptions of Data Practices*. Available at <https://livingwithdata.org/project/wp-content/uploads/2020/05/living-with-data-2020-review-of-existing-research.pdf>.
- Leonard PG** (2018) *Social Licence and Digital Trust in Data-Driven Applications and AI: A Problem Statement and Possible Solutions*. Available at <https://ssrn.com/abstract=3261228>.
- McIntyre N and Pegg D** (2018) Councils use 377,000 people's data in efforts to predict child abuse, *The Guardian*, 16 September. Available at <https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse>.
- Messenger C and Molloy D** (2014) *Getting It Right for Families: A Review of Integrated Systems and Promising Practice in the Early Years Field*, Early Intervention Foundation. Available at <https://www.eif.org.uk/report/getting-it-right-for-families-a-review-of-integrated-systems-and-promising-practice-in-the-early-years>.
- Ministry of Housing, Communities and Local Government** (2020) *Financial Framework for the Troubled Families Programme*. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/922390/Troubled_Families_Financial_Framework_2020.pdf.
- Ministry of Housing, Communities and Local Government** (2021a) *Local Data Accelerator Fund for Children and Families: Prospectus*. Available at <https://www.gov.uk/government/publications/local-data-accelerator-fund-for-children-and-families>.
- Ministry of Housing, Communities and Local Government** (2021b) Press release: Next phase of £165 million programme for vulnerable families launched. Available at <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 and Lugg F** (2017) Public understanding of, and attitudes towards, data linkage in research: *IJPDS* (2017), Issue 1, Vol. 1: 073, Proceedings of the IPDLN Conference (August 2016). *International Journal of Population Data Science* 1(1). <https://doi.org/10.23889/ijpds.v1i1.92>.
- NatCen** (2018) *Public Attitudes to Data Linkage*: A report prepared for University College London by NatCen Social Research. Available at <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*. Available at <https://www.stats.govt.nz/corporate/measuring-stats-nzs-social-licence>.
- NL Digitaal Government** (2019) *Data Agenda Government*. Available at <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*. Available at <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*. Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandvaccinehesitancygreatbritain/13januaryto7february2021>.
- Office for Statistics Regulation** (2018) *Joining Up Data for Better Statistics*. Available at <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/09/Data-Linkage-Joining-Up-Data.pdf>.
- Paprica PA, Nunes de Melo M and Schull MJ** (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*. Available at https://www.canada.ca/content/dam/pco-bcp/documents/clk/Data_Strategy_Roadmap_ENG.pdf.
- Redden J, Dencik L and Warne H** (2020) Datafied child welfare services: Unpacking, politics, economics and power. *Policy Studies* 41(5), 507–526. <https://doi.org/10.1080/01442872.2020.1724928>
- Salganik MJ, Lundberg I, Kindel AT 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. Available at <https://www.pnas.org/content/pnas/117/15/8398.full.pdf>.
- Scourfield J, Corliss C, Wijedasa D, Robling M and Clayton V** (2019) *Overview of Administrative Data on Children's Social Care in England, What Works for Children's Social Care*. Available at https://whatworks-csc.org.uk/wp-content/uploads/WWCSC_Overview_of_administrative_data_Oct_2019.pdf.
- Shaw JA, Sethi N and Cassel CK** (2020) Social licence for use of big data in the COVID-19 era. *NPJ Digital Medicine* 3, 128. Available at <https://www.nature.com/articles/s41746-020-00342-y>.
- Stats NZ** (2018) *Algorithm Assessment Report*. Available at <https://data.govt.nz/use-data/analyse-data/governmentalgorithm-transparency>.
- Thomson I and Boutilier R** (2011) *The Social Licence to Operate*, in *SME Mining Engineering Handbook*. Englewood, CO: Society for Mining, Metallurgy and Exploration.
- van Zoonen L** (2020) Data governance and citizen participation in the digital welfare state. *Data and Policy* 2, e10. Available at <https://www.cambridge.org/core/journals/data-and-policy/article/data-governance-and-citizen-participation-in-the-digital-welfare-state/CCF2E1914C7E2D4593D550C6BC9E4C70>.
- Vannier Ducasse H** (2020) Predictive risk modelling and the mistaken equation of socio-economic disadvantage with risk of maltreatment. *British Journal of Social Work*. <https://doi.org/10.1093/bjsw/bcaa182>
- Wachter RM and Cassel CK** (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: WW Norton & Company.
- Waller M and Waller P** (2020) *Why Predictive Algorithms Are So Risky for Public Sector Bodies*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3716166.
- White S and 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*. Available at http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf.
- Wroe LE** (2021) Young people and “county lines”: Contextual and social account. *Journal of Children's Services* 16(1), 39–55.

Cite this article: Edwards R, Gillies V and Gorin S (2021). Data linkage for early intervention in the UK: Parental social license and social divisions. *Data & Policy*, 3: e34. doi:10.1017/dap.2021.34