Leon Feinstein is the Professor of Education and Children's Social Care and Director of the Rees Centre, University of Oxford. Previously, he was the Director of Evidence at the Children's Commissioner's Office, and Chief Analyst in the Prime Minister's Delivery Unit in HM Treasury. He has a PhD in Economics from University College London, is a Fellow of the British Academy of Social Sciences, and a Visiting Professor at LSE and the University of Sussex. Leon was a Trustee at the What Works Centre for Social Care and a Member of the Youth Endowment Fund Grants Committee and the NSPCC Research Advisory Group.

An argument for better data about children

Leon Feinstein, University of Oxford[†]

This essay makes a case for attempts to link data about children across periods and multiple sources and to data about the adults they live with, so as to get a better handle on levels and characteristics of child vulnerability to harm and deprivation. It contends that there is both a rights-based ethical foundation for better aggregate data about children and families as well as a practical argument that it can be used to enable government and society to improve the lives, experiences and outcomes of children and young people, although these benefits do not come without risks, preconditions and costs. The use of children's statistical data should be balanced with concerns about data protection and the prevention of misuse of data, but there is also a positive agenda about linking data to the voices of children and their families, using data to address the research concerns and issues of children and families.

The essay begins with an argument for the importance and value of aggregate data about children and their families and wider contexts as a means for influencing policy and practice in ways that can improve children's experiences and outcomes. Drawing on the example of immigration status as a risk factor causing vulnerability for some children, I first describe the weakness of current aggregate data for assessing the needs of children in the UK. I use this example because insecure immigration status is an important form of disadvantage and risk for some children, and influences their experience of education as well as other parts of their lives.

I then consider some of the real and perceived risks of improving data, and raise the issue of the importance of public trust and hence of transparency about data use and adequate mechanisms for ensuring meaningful public interest. The conclusion emphasises the vital role of government in balancing the benefits of aggregate data for holding government and society to account against the genuine risk of use of individual data in ways that would conflict with the rights of children.

A rights-based requirement for good aggregate data about children
Starting from a rights perspective, Article 3, para 2 of the United Nations Convention on the Rights of the Child (UNCRC) commits States Parties 'to ensure the child such protection and care as is necessary for his or her wellbeing, and, to this end… take all appropriate legislative and administrative measures.'

How are we to know if this is achieved and for whom? Wellbeing is not unidimensional, and the UNCRC recognises multiple aspects of protection and care, for example in relation to children who are at risk of 'physical or mental violence, injury or abuse, neglect or negligent treatment, maltreatment or exploitation, including sexual abuse' (Article 19), removed from their family environment (Article 20), adopted (Article 21), a refugee (Article 22), disabled (Article 23), economically exploited (Article 32), sexually exploited or abused (Article 34), abducted or trafficked (Article 35) or in detention (Article 37).

The vulnerability framework developed at the Children's

Commissioner's Office 2016–20 (e.g., CCO, 2017) was an attempt to measure the number of children with different characteristics related to protection and care across the domains identified above, and to assess what is known about their views and experiences through forms of qualitative research, their characteristics and outcomes in terms of statistical measures and what they receive in terms of government support. This was a deliberate attempt to span the terrain of risk and harm and to provide a general overview of issues so that the Commissioner could report to Parliament on the state of the nation's most vulnerable, hidden and invisible children and the quality of the response of government and society to them. This enabled the Commissioner to target reviews on specific issues of concern from an evidence-based and informed perspective.

The Commissioner's power to request data and visit sites where children were resident was then used to further probe experiences for groups of children for whom the data indicated high levels of concern or for whom there was inadequate information (e.g., CCO, 2020a, 2020b). This method was used to identify particular areas of concern on which subsequent lobbying by the Commissioner pushed for improvements in their care and treatment, including young carers (CCO, 2016), children in detention in mental health settings and other settings (CCO, 2020a) and homeless children (CCO, 2020b). Advocacy could have occurred without good quantitative data on the numbers of children affected, but the Commissioner found good data essential for engaging public attention. As the former Home Secretary the Rt Hon Jacqui Smith put it in a blog about the 2019 annual CCO report: 'Politics is about priorities. The Children's Commissioner has shown our politicians what's happening to our most vulnerable children. Now they must choose to do the right thing' (CCO, 2019b).

An example of the need for improved aggregate data
How many children in the UK have insecure immigration status?
One subgroup of concern in the CCO analysis was children who were vulnerable to harm by virtue of immigration status. The 2019 report included in this regard the category of refugees

and a separate category of children and young people with 'unresolved immigration status', with a further four subcategories of children and young people who were: unaccompanied asylum-seeking; arriving under Dublin regulations; in families seeking asylum; or undocumented (CCO, 2019a).

Subsequent work (Feinstein et al., 2022) has revised these categories to assess risk of harm for three distinct categories of immigration status: children and young people without leave to remain, with limited leave, or with indefinite leave. These can be distinguished from children and young people with UK citizenship in terms of level of risk of harm associated with each status, on an underlying continuum of risk, with no leave to remain the most associated with risk.

We found that it is not possible to estimate from the official statistics the number of children in the UK without leave to remain, with limited leave or with indefinite leave. This is partly because no attempt is made by the Office for National Statistics (ONS) to track children through the system across years, and partly because so many children are missing from the official statistics as undocumented, 'invisible' to the system, or below the radar (Chase, 2009; Kohli, 2006).

Much is known from case histories and qualitative research about how as a nation the UK treats children and young people with insecure immigration status (see, for example, Dexter et al., 2016; Price & Spencer, 2015). However, if we cannot even say how many children are currently going through the legal system, how they are in general and what their outcomes are, how sure can we be that we are meeting our obligations to them?

Children in need
These questions about accountability represent one kind of argument for the value of aggregate data. Arguments are also made about how to understand the effectiveness of policy. As Chancellor of the Duchy of Lancaster Michael Gove put it in the 2020 Ditchley Annual Lecture on 'The privilege of public service' (Gove, 2020):

If Government ensures its departments and agencies share and publish data far more, then data analytics specialists can help us more rigorously to evaluate policy successes and delivery failures. People's privacy of course must be protected. But once suitably anonymised, it is imperative that we learn the hugely valuable lessons that lie buried in our data.

When we try to understand levels of need more generally, for example in relation to the number of children in England living in households or families with characteristics or locations that indicate higher potential likelihood of current and future harm, we also find considerable difficulties of measurement. Some of this results from inherent empirical challenges, but the bigger difficulty comes from the lack of activity to link data about children to data about adults, so as to know how many children are living in households characterised by high levels of drug or alcohol misuse, mental health difficulties, disabilities, material deprivation, prison, abuse or other characteristics likely to increase the risk of harm or disadvantage for children. There are some important beneficial examples of linkage but, as the CCO work has indicated, very substantial gaps.

The National Audit Office (NAO) made the argument for better data in a comment in 2019 on the Department for Education (DfE): 'The Department... still does not fully understand what is driving demand for children's social care or why there is such wide variation between local authorities in their children's social care activity and costs' (HC 1868, Session 2017-19, 23 January 2019).

When the government responds to the MacAlister review of the Children's Social Care System (MacAlister, 2022) we will know more about how this administration intends to address the continued increasing pressures on the care system. The options comprise combinations of raised thresholds, reduced rights, improved prevention and increased spend. Better data on the level of need will be required if we are to know in aggregate what the results are for children and families.

The point of the CCO framework is that it is general and holistic. It is for governments to decide on priorities, but we may still wish to know the outcomes and experiences of

children impacted by the actions or inactions of government. For governments that choose to act to meet need, better data enables focus, clarity of strategy and transparency of outcome.

Data risks

Invisibility

Returning to the issue of immigration status, we must probe further into what is meant by 'invisibility' and whom it serves. A 2016 Freedom of Information request[†] by Pippa King asked 'if the Police or Home Office have requested data, or whether data has been sent to them, from the National Pupil Database and or any other Department of Education held pupil level database.' The answer indicated that: 'Since April 2012, the Police have submitted 31 requests for information to the National Pupil Database. All were granted, however only 21 resulted in information being supplied.'

As a result of legal action in 2017 the DfE stopped requesting data on nationality. However, in a country with a hostile environment for people with insecure immigration status there are evident, perceived risks to the provision of information about immigration status for children and families. Concerns about these risks increase the degree to which these families are forced to hide from official agencies, increasing their vulnerability to trafficking and other risks of harm (House of Commons, 2020).

Therefore, this essay doesn't argue for an open and blanket approval for all forms of use of children's data, but seeks to place the use of data in a context of participation and democracy. How are concerned families to know that if data is linked to enable beneficial aggregate analysis it will not be used at individual level to target children and families for punitive measures, immigration controls or other forms of sanction?

As a practical matter it is not technically hard to link data in secure, encrypted environments de-identifying individuals by removing personal, identifying information and replacing this with non-identifiable data keys that enable statistical analysis. If risks of small numbers are appropriately handled,

---

† Reference number 2016-0038372

no individuals can be re-identified. In principle, the Five Safes (UK Data Service, 2020) approach developed at the ONS enables data to be made available in non-identifying ways for research. This is similar to approaches adopted around the world (e.g., Hanafin, 2020) that provide secure, trusted and legal bases for linking administrative to longitudinal data, generating the potential for informative research. Researchers including myself have often experienced frustration at how slow and difficult it is to get access to aggregate and de-identified data held by the DfE, even when necessary safeguards are in place and a clear case has been made that the analysis is likely to add clear insight and value in support of children's interests.

However, individuals in any de-identified or anonymous dataset can be re-identified if suitable re-identifying information is provided. Moreover, it is hard for individuals and families to know how their individual data are being used and processed in the first place, and that individual data are not being shared with criminal justice or other agencies without their knowledge or agreement.

Therefore, trust is critical. However, a 2020 report by the Information Commissioner's Office (ICO) provided 'a comprehensive review of data protection practices, governance and other key control measures supporting' the National Pupil Database and other databases held by the DfE, which leads on data about children within Whitehall. The review made 139 recommendations, and found:

> There is no formal proactive oversight of any function of information governance, including data protection, records management, risk management, data sharing and information security within the DfE which along with a lack of formal documentation means the DfE cannot demonstrate accountability to the GDPR.

Members of the public might recognise the benefits of ethical and careful use of children's data for improving and evaluating policy and practice, but how can we be confident in systems for linking data about children when, according to the ICO report (2020, p 5):

Information risks are not managed in an informed or consistent manner throughout the DfE or in line with the Risk Management Framework. Information assets are not assessed with sufficient frequency to ensure that the process is effective and resulting risks are not recorded with sufficient granularity or detail on the Information Risk Log to enable meaningful control and monitoring. Not all information risks are recorded and where they are, they do not always identify actual risks or control measures.

It is of concern that the ICO finds 'There is an over reliance on using public task as the lawful basis for sharing which is not always appropriate and supported by identified legislation' (2020, Executive Summary, p 6). Thus, it seems that data management in the key government department charged to be custodian of children's data in central government is neither adequately open and transparent to inspire trust, nor adequately resourced to enable timely, secure and effective access to data.

Ethics
Ethical practice requires a further level of reflection. Ethics, as Leslie et al. (2020, p. 19) put it in their review of the ethics of using machine learning techniques in children's social care, 'is both about justifying morally correct conduct and about motivating and setting a direction of travel for that conduct.'

If we are to take a rights-based approach to requesting that data about children be available for statistical analysis, we must also recognise Article 12: 'States Parties shall assure to the child who is capable of forming his or her own views the right to express those views freely in all matters affecting the child.' This requires us both to seek the views of children about uses of data and also their views and perspectives on their circumstances, experiences and conditions as a form of information in data[†], when we use data to identify and address unmet needs, for example. This is not straightforward.

---

† I am grateful to Professor Elaine Sharland at the University of Sussex for this formulation. There is much more to be said about this.

There is an important tension between the Article 12 right to be heard and the Article 3 duty on states to ensure that 'in all actions concerning children… the best interests of the child shall be a primary consideration.'

Taylor (2016) describes the legal and operational difficulties of achieving the objectives of Article 3 when children are not at the centre of legal and policy systems. Although children may not be best placed to assess the value of aggregate data for policy decisions, this does not mean the conversation should not be had – just that Article 12 must be balanced with Article 3. As Lundy (2007) emphasises in a classic paper on the depth of meaning of Article 12, children, families and practitioners rarely have opportunities to shape the collection, interpretation and uses of data about their lives, and there is no recognised way to assess the quality or impact of this work (see also Bakketeig et al., 2020). Even where vulnerable children's voices are elicited, they are often not well heard or acted on (Kennan et al., 2018). Parents, caregivers and wider communities also have a role in setting priorities for data collection and in considering the meaning and implications from the findings of data.

In addition, to claim that there are material benefits to aggregating children's data and linking to information on parents and caregivers, we must establish both that improved data is necessary to realise the rights of children, and that appropriate safeguards are in place so that children's digital rights are not impinged. Recognising the benefits of collecting, collating and analysing children's data we must also adequately resource the work of securely, safely and transparently handling it, and of engaging children and young people, families and wider communities on the questions for analysis. On these issues there is clearly a long way to go.

Conclusion
As with children who are vulnerable by virtue of immigration status, the implicit response of government to unmet need may be to enhance invisibility and restrict rights rather than meet need. Aggregate data is a means for government and society to know something of how it is doing in meeting need. This essay argues that such data are necessary if we are to

deliver on the rights of children, but this must be balanced with a genuine and deep engagement with children, young people, families, practitioners and wider communities about what data are used and how, and more progress must be made on transparency and demonstrable safeguards.

Government can take advantage of the opportunities provided by better data about children and families, but also has a responsibility to ensure the data are used in the best interests of children, and that must involve deep, wide and meaningful dialogue that enhances trust by recognising the risks and biases of data as well as the benefits. We might recognise the limits of children and families to understand all aspects of government and data, but providing real opportunities to shape data use would not only improve legitimacy; it would also improve policymaking.

Bakketeig, E., Boddy, J., Gundersen, T., Østergaard, J., & Hanrahan, F. (2020). Deconstructing doing well: What can we learn from care experienced young people in England, Denmark and Norway? *Children and Youth Services Review, 118*

CCO (Children's Commissioner). (2016). *The support provided to young carers in England*

CCO. (2017). *Childhood vulnerability in England 2017*

CCO. (2019a). *Childhood vulnerability in England 2019*

CCO. (2019b). Former Home Secretary and member of Children's Commissioner advisory board Jacqui Smith responds to our vulnerability report

CCO. (2020a). *Who are they? Where are they? Children locked up*

CCO. (2020b). *No way out: Children stuck in B&Bs during lockdown*

Chase, E. (2009). Agency and silence: Young people seeking asylum alone in the UK. *British Journal of Social Work, 40*(7), 2050-2068

Dexter, Z., Capron, L., & Gregg, L. (2016). *Making life impossible: How the needs of destitute migrant children are going unmet.* The Children's Society

Feinstein, L., Aleghfeli, Y. K., Buckley, C., Gilhooly, R., & Kohli, R. K. S. (2022). Conceptualising and measuring levels of risk by immigration status for children in the UK. *Contemporary Social Science, 16*(5), 538-555

Gove, M. (2020). The privilege of public service [Transcript of the speech]. Ditchley Annual Lecture

Hanafin, S. (2020). *Scoping review of literature: Best international practices in linking administrative and longitudinal study data.* Dublin: Department of Children, Equality, Disability, Integration and Youth

House of Commons. (2020). *Windrush lessons learned review*. Independent review by Wendy Williams. HC 93

ICO (Information Commissioner's Office). (2020). *Data protection audit report*. Department for Education

Kennan, D., Brady, B., & Forkan, C. (2018). Supporting children's participation in decision making: A systematic literature review exploring the effectiveness of participatory processes. *British Journal of Social Work, 48*(7), 1985-2002

Kohli, R. (2006). The sound of silence: Listening to what unaccompanied asylum-seeking children say and do not say. *British Journal of Social Work, 36*(5), 707-721

Leslie, D., Holmes, L., Hitrova, C., & Ott, E. (2020). *Ethics review of machine learning in children's social care.* What Works for Children's Social Care

Lundy, L. (2007). 'Voice' is not enough: Conceptualising Article 12 of the United Nations Convention on the Rights of the Child. *British Educational Research Journal, 33*(6), 927-942

MacAlister, J. (2022). *Independent review of children's social care. Final report*

Price, J., & Spencer, S. (2015). *Safeguarding children from destitution: Local authority responses to families with 'no recourse to public funds'*. COMPAS

Taylor, R. (2016). Putting children first? Children's interests as a primary consideration in public law. *Child and Family Law Quarterly, 28*(1)

UK Data Service. (2020). What is the Five Safes framework?