



Scaling up research on family justice using large-scale administrative data: an invitation to the socio-legal community

Karen Broadhurst, Linda Cusworth, Judith Harwin, Bachar Alrouh, Stuart Bedston, Liz Trinder, Kerina Jones, David Ford & Lucy Griffiths

To cite this article: Karen Broadhurst, Linda Cusworth, Judith Harwin, Bachar Alrouh, Stuart Bedston, Liz Trinder, Kerina Jones, David Ford & Lucy Griffiths (2021) Scaling up research on family justice using large-scale administrative data: an invitation to the socio-legal community, *Journal of Social Welfare and Family Law*, 43:3, 237-255, DOI: [10.1080/09649069.2021.1953856](https://doi.org/10.1080/09649069.2021.1953856)

To link to this article: <https://doi.org/10.1080/09649069.2021.1953856>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 30 Jul 2021.



Submit your article to this journal [↗](#)



Article views: 437



View related articles [↗](#)



View Crossmark data [↗](#)



Scaling up research on family justice using large-scale administrative data: an invitation to the socio-legal community

Karen Broadhurst^a, Linda Cusworth^b, Judith Harwin^b, Bachar Alrouh^a, Stuart Bedston^a, Liz Trinder^c, Kerina Jones^d, David Ford^d and Lucy Griffiths^d

^aDepartment of Sociology, Lancaster University, Lancaster, UK; ^bLaw School, Lancaster University, Lancaster, UK; ^cLaw School, University of Exeter, Exeter, UK; ^dPopulation Data Science, Medical School, Swansea University, Swansea, UK

ABSTRACT

This article outlines the value of administrative data for family justice research. Although socio-legal scholars have extended their research beyond purely theoretical or doctrinal analyses, studies using large-scale digital datasets remain few in number. As new opportunities arise to link large-scale administrative datasets across health, education, welfare and justice, it is vital that the community of family justice researchers and analysts are supported to deliver research based on entire service or family court populations. In this context, this article provides a definition of administrative data, before outlining the potential of single, linked or blended administrative data sets for family justice research. The remaining sections of the article speak to questions that are pertinent to this particular academic community, including the distinctive contribution of the socio-legal scholar to interdisciplinary teams and the place of data providers in collaborative research. Drawing on the sociological concept of ‘publics’, the final section considers the multiple interest groups whose social licence must be secured, when personal records are used to understand the relationship between law and family life.

KEYWORDS

Administrative data; justice; socio-legal; theory; digital records

Introduction

There is a general consensus that the past two decades have witnessed a growth in empirical legal research, not least in response to sustained demand from policy-makers and practitioners (Cane and Kritzer 2010, Ho and Kramer 2013, Leeuw 2015, Creutzfeldt *et al.* 2019, Hamann 2019). However, there is also evidence of a longstanding empirical legal skills deficit internationally, due to limited research education for law students within undergraduate programmes (Genn *et al.* 2006). Arguably, academics working in the field of family justice, have been the most willing and able to conduct empirical work (Eekelaar and Maclean 1990, Huntington 2018). However, as yet, even family justice scholars have been slower to embrace new research opportunities afforded by the increasing digitisation of information within

CONTACT Karen Broadhurst  k.broadhurst@lancaster.ac.uk

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

public institutions and the courts, when compared to colleagues in health research. Digital administrative records that are machine readable, together with opportunities to link large-scale administrative datasets across health, education, welfare and justice, enable vital insights to be gained based on entire service populations. As governments nationally and internationally invest in the further digital transformation of justice (Susskind 2017), family justice scholars can take advantage of new opportunities to scale up their study of law in society and, through linking data, capture a more holistic view of child and family journeys through justice and related services.

This article sets out the opportunities and challenges of research using large-scale single, linked and blended administrative datasets for family justice research. By family justice research, we mean in this article to focus on questions about children and families involved with the family courts and related public agencies on account of public or private law matters. Although broader questions of social justice that impact on families can also be addressed using administrative data (for example questions about the distribution of resources), our focus on institutions speaks to growing interest in the UK and further afield on how the wealth of data produced by the family courts can be used to inform the delivery of justice (Brownell and Jutte 2013, Berk *et al.* 2016, Baldwin *et al.* 2019, Cusworth *et al.* 2021).

Many of the points we make in this article will be familiar to researchers in the disciplines of epidemiology and health. However, this article speaks to a community of scholars with more limited experience in the use of large-scale administrative data. Family justice scholars currently produce a wealth of empirical research, but the majority of studies have been based on relatively small samples. That is inevitably (and appropriately) the case for qualitative research that aims to take an in-depth look at processes and experiences, such as the Eekelaar and Maclean series on family justice professionals (Eekelaar *et al.* 2000, Maclean and Eekelaar 2009, Eekelaar and Maclean 2013). However, it is also the case for quantitative studies in family justice, where there are real challenges to achieving large representative samples, given resource and ethical constraints. Family justice researchers have traditionally made very effective use of court files to advance our understanding across the field. This has included financial remedies (Hitchings *et al.* 2013, Woodward and Sefton 2014), private law children (Hunt and Macleod 2008, Trinder *et al.* 2013, 2014, Harding and Newnham 2015) and public law children (Masson 2010). However, to date, this has been based on resource-intensive reading of free text files by researchers, often requiring visits to local courts, rather than the ability to download and analyse anonymised quantitative data. The result is that studies are generally restricted to single snapshots in time, based on a few hundred cases (Masson *et al.* 2008). The limits of empirical research in family law have been subject to considerable discussion (Brinig 2002, Genn *et al.* 2006, Broadhurst *et al.* 2018).

Whilst small *n* studies and theoretical work are key to our understanding of how family justice works, administrative data is one source of 'big data' (Connelly *et al.* 2016) that can extend the kinds of questions we can ask of our justice institutions. This can include large-scale profiling studies such as who engages with the family courts, and outcome studies about how law shapes lives over time. The paucity of socio-legal research at a population-level persists in a number of international contexts, despite the growth of

data safe havens that provide opportunities for the safe use of millions of linked micro-records (Holman *et al.* 2008, Ford *et al.* 2009).

Large-scale, data intensive research is an interdisciplinary challenge that requires researchers to work across new knowledge boundaries. Although there is a long history of interdisciplinarity within the family justice research community, when working with large-scale datasets that involve the mass extraction and manipulation of digital data, the interdisciplinary challenge takes on a new guise. Making effective use of large-scale and linked datasets requires multidisciplinary teams, including colleagues with quantitative or computational expertise, which is not typically held by socio-legal scholars. Here, the proliferation of data science institutes is advantageous, facilitating collaborations which enable law and science to more easily align around pressing societal questions. Although novel disciplinary mixes can be as daunting as they are exciting, clarity of roles and equal valuing of expertise makes for productive and more comfortable team work. Family justice scholars can make a distinctive contribution to interdisciplinary data science, by bringing a critical understanding of *context* to both data collection and analysis. Social theory is far from out of place in this hybrid intellectual space. Robust research using administrative data must be technically robust but also make *substantive sense* – informed by a deep understanding of the issues that face family justice systems. Learning from colleagues completing theoretical or smaller scale empirical studies, use of big(ger) datasets can address questions of prevalence, equity or trends over time – but only when analyses are infused with a keen understanding of how the law works in practice, in the real world. Empirical family justice research is not simply a technical exercise that seeks to reflect the status quo; rather policy-focused research must be informed by an awareness of the limits of our existing justice institutions and arrangements.

Tailored specifically to the international community of scholars whose work is focused on questions of family involvement and the operation of family justice systems, this article builds on an excellent generic literature on administrative data (Roos *et al.* 1987, Goerge and Lee 2002, Card *et al.* 2010, Elias 2014). The first section is largely descriptive, starting with definition, before outlining the value of research using single, linked or blended administrative datasets. The remaining sections of the article speak to questions that our own work indicates are pertinent to family justice researchers. Discussion covers the distinctive contribution of the socio-legal scholar to interdisciplinary data intensive research, before considering the place of the data providers in this collaborative mix. Drawing on the sociological concept of ‘publics’, the final sections of the article consider the multiple interest groups whose social licence must be secured, when secondary analysis uses records produced in the contested world of justice.

Administrative data: definition and value

Definition

Administrative data refers to information about persons or organisational activity, which is collected routinely by government, statutory or other agencies for their own organisational purposes (Woollard *et al.* 2014). Public institutions, such as the courts and allied services, typically maintain individual case records that not only contain demographic

details, but also details of the individual's interaction with other services. This *micro or person-level* data is hugely valuable for understanding characteristics of court users, individual pathways through justice systems, legal representation and outcomes of justice involvement. Other data includes information regarding performance, produced to meet the requirements of regulatory bodies or external auditors.

In a number of international contexts, digital administrative records are increasing in volume, quality and longevity, as are opportunities for linking administrative records across sectors of health, education, welfare and justice (Holman *et al.* 2008, Card *et al.* 2010). Indeed, a number of proponents have argued that the volume of this data now constitutes a vital source of 'big data' (Connelly *et al.* 2016). Although collected primarily for organisational purposes, this data is increasingly valuable for socio-legal research, given that the shift to digital machine-readable records now delivers millions of person-level records that span a number of years. In England, the HM Courts & Tribunal Service reform programme (The HMCTS reform programme – GOV.UK [no date](#)), which seeks to modernise the delivery of justice will deliver a wider range of new digital resources, resulting from online divorce records. The trend towards digitisation of justice is evident in a number of nation states, as developments in online dispute resolution and adjudication proliferate (Mania 2015).

Capturing family justice populations at scale

So, what are the advantages of this data for justice research? First, administrative data collected or produced centrally by government, the courts and other national agencies, is typically captured at the level of entire service populations, enabling research at a pace and scale which would be very difficult to achieve through other methods. Central to the delivery of family justice is a presumption of even-handed treatment of cases, but in the absence of population-level research, accusations of inequitable treatment of individuals or groups are difficult to either substantiate or refute (Byrom 2019). In this context, retrospective use of institutional records is a practical approach to generating system intelligence and directly addressing professional and public concerns.

To date, family justice researchers have generated a volume of rich qualitative data based largely on interview research with family justice users and professionals. There has also been some use of administrative data in the form of manual searches of court file data. However, the costs of manual collection of data and inputting into digital form for analysis is time-consuming and expensive and inevitably restricts sample sizes.

Whilst qualitative research will always be essential to understand the experiences of individuals and how the family justice system operates, small sample studies can be less persuasive in terms of their influence on policy. In contrast, when researchers combine insights based on entire service populations with qualitative research, even politically unpalatable observations are less easily dismissed (Cusworth *et al.* 2019). Moreover, whole population data helps researchers overcome difficulties in achieving representative samples. This is particularly important when dealing with contested issues where different interest groups can skew the evidence base, an ever-present issue in family justice research. It is also vital when research aims to engage hard to reach groups (Goerge and Lee 2002, Evans *et al.* 2010). For example, (Lery *et al.* 2005) used child protection administrative records to analyse sibling groups within child welfare services, advancing

knowledge about the continuation or otherwise of sibling ties, once children enter care. More recently, (Hafekost *et al.* 2017) examined maternal alcohol use disorder diagnoses and subsequent child protection contact using linked records, identifying health vulnerabilities at scale, in women involved with children's statutory services. However, examples specific to court populations are fewer in number. (Broadhurst *et al.* 2015) examined the total volume of records for women appearing in public law care proceedings in England (2007–2014; $n = 43,541$) to quantify women's repeat appearances. This latter work catalysed major reform of preventative services following child removal and illustrated, for the first time, the value of administrative family court data in England, for population-level research which seeks direct policy impacts. (Harwin *et al.* 2019) examined the total population of children subject to care proceedings to identify trends in the use of orders (supervision orders, child arrangement orders, special guardianship, care orders, and placement orders) and legal outcomes. This was the first national study to explore child outcomes capturing return to court for further care proceedings for different order types. Among several findings, it challenged myths that special guardianship is an unstable permanent placement option for children. These findings have fed directly into policy, with concomitant questioning of practice. Given large sample sizes, administrative data also offers the possibility of creating particular cohorts of individuals to study the impact of policy or legislative change on population subgroups. For example, restrictions in access to legal aid and court reform programmes are apparent in a number of international contexts but with limited analysis of the impact of cutbacks on individual lives or groups (Maclean *et al.* 2015, Byrom 2019). Powerful analyses published to date might be complemented by the different kind of questions that can be answered through population-level data (Maclean *et al.* 2015).

Equally, comparative work, which probes the influence of local or regional demographics on family justice, is also possible through disaggregation of total populations. (Harwin *et al.* 2018) study found marked regional differences in the use of supervision orders supporting reunification and of care orders. Areas that had high rates of supervision orders made proportionately fewer care orders and vice versa, suggesting that cultural practices are important, whilst raising broader issues of equity and fairness (Harwin *et al.* 2018). The large sample sizes afforded by administrative data also enable quasi-causal analyses, focussing on the different outcomes of family justice interventions on individual lives (Dunning 2012). Innovation in family justice practice is evident in a number of international contexts, but often funding is simply not available to robustly evaluate new practices. In the absence of evaluative data, new and promising practice models often falter, as proponents struggle to evidence the case for sustainable funding. Thus, administrative data is a powerful resource for a range of empirical family justice studies.

Gaining a longer view of law in society

A major advantage of administrative datasets is that they are often longitudinal or can be made longitudinal (Evans *et al.* 2010). By longitudinal we refer to studies that draw on repeated measures/collection of the same variable, but over an extended timeframe. Conducting longitudinal retrospective analyses based on population-wide administrative data that spans a number of years is a cost-effective method of gaining a long run

perspective on organisational and population patterns over time. For example, episodes of individual service engagement are date stamped; they capture an individual's engagement with services at a particular point in time and any subsequent involvement (see for example, Johnson *et al.* 2020). Hence, administrative records can be restructured to capture the pathways of individual service users or litigants over time through family justice systems, together with short and longer-term outcomes.

Although epidemiologists and social scientists generally find studies more credible when data collection is prospective and captured according to protocols designed specifically for the research in question, pragmatic considerations must also apply. For example, in the case of the youth courts, it would be very difficult to prospectively track a study population over a lengthy period of time, given the likelihood of significant attrition (Janson 1996). Individuals, youths and families move across geographic borders or may simply drop out of prospective surveys due to changing life circumstances. The likelihood of loss to follow up, which typifies prospective designs, is exacerbated in the case of individuals and families involved with justice systems given the stigmatised nature of problems and personal difficulties (Roos *et al.* 1987, Brownell and Jutte 2013, Howe *et al.* 2013). For example, (Harwin and Alrouh 2017) found that parents were reluctant to engage with researchers once children returned home from care, as their focus was on reclaiming as normal a life as possible with their children. Research served as a reminder of a turbulent period of family life that saw parents lose children from their care on account of child protection concerns. However, by extracting and analysing de-identified family court records, (Harwin *et al.* 2019) were able to answer pressing questions about the strength of the supervision order to support family reunification by calculating breakdown rates. Gaining a longitudinal perspective *retrospectively*, through the analysis of administrative datasets, can provide a longer view of the outcome of decisions taken in the family or youth courts, without running into such problems.

Joining up the dots: record linkage for justice research

Administrative data produced for non-research purposes is only useful if there are sufficient variables to address research questions (Administrative Data Taskforce 2012). Researchers will typically have little influence over data collection; that is the design, structure and therefore scope of organisational datasets. Core variables that researchers would seek to include in any research study are not necessarily relevant to organisational performance and may therefore not be included in organisational datasets. For example, the recording of domestic abuse in cases that concern disputes between parents about children's residence or time with parents is often poorly recorded, if at all (Cusworth *et al.* 2020) even though researchers would view these as critically important variables. This is a key limitation of what (Connelly *et al.* 2016) have described as 'found' rather than 'made' social science data.

In addition, data quality issues can further restrict the range of variables available for research. Missing data is common in administrative datasets, and although there is a range of principled statistical methods for imputing missing values (Perkins *et al.* 2018), in some cases the extent of missing values means that information against particular variables cannot reliably be used.

However, the scope of any single dataset can be greatly improved through linking data to other relevant sources, such as health or education datasets. By linking data we gain a less fragmented view of individual interaction with, or pathways through services and the courts (Jay *et al.* 2017). In addition, there is a wealth of excellent literature on approaches to data linkage (Harron *et al.* 2017, Zylbersztejn *et al.* 2020). In the discipline of health in particular, there are multiple examples of robust studies that have linked data across organisational sources to build holistic analyses of patients' engagement with health services and their outcomes (Harron *et al.* 2020). For example, (Orr *et al.* 2019) linked hospital and children's services records to examine the relationship between hospitalisation for maternal assault and risk of child protection involvement. More recently, Griffiths and colleagues linked Welsh family court records to a number of health records, to uncover very high levels of mental health need in pregnancy for women appearing in care proceedings (Griffiths *et al.* 2020).

Administrative data can also be linked with longitudinal survey data, such as the Millenium Cohort Study and National Child Development Study. Although historically such data has tended to be used in isolation, there is an increasing volume of published studies illustrating what can be gained by combining these data sources. By linking administrative and survey data we add richness and depth to our analyses. For example, in the US, to examine income as an important measure of wellbeing, (Medalia *et al.* 2019) combined survey data with tax records, among other sources, to produce the US Comprehensive Income Dataset. In addition, analyses of administrative data can be complemented by prospective data collection, to add depth to understanding. As above, where time and budgets allow, prospective data collection enables the researcher to have greater control over the scope and quality of data collection.

In Scandinavian countries there is a long history of the use of national registers as a major source of administrative information about populations (Connelly *et al.* 2016). However, in a number of other international contexts, a range of initiatives has now radically changed opportunities for the use of administrative data for secondary analysis. In the UK, successive policy and legislative developments have sought to reduce barriers that stand in the way of linking data across government departments as set out in the 'UK Strategy for Data Resources for Social and Economic Research' (UK Data Forum 2013). Administrative Data Research (ADR) UK is an initiative funded by the Economic and Social Research Council which aims to accelerate opportunities for safe linkage of the wealth of administrative data held by governments. Investment in ADR UK is based on the premise that secondary analysis of this data can lead to the efficient production of evidence to inform public policy. In Australia, the Research Data Infrastructure initiative similarly aims to accelerate the use of controlled data (Australian Government Department of Health no date). In Canada, the Canadian Data Platform under the Strategy for Patient-Oriented Research (SPOR) provides a single portal through which researchers can request access to health administrative, demographic and social data (Guttmann 2019).

Indeed, such is the international support for the re-use of valuable administrative data, that funding and research efforts are now being directed towards the sharing of digital micro-data across international boundaries (OECD Expert Group 2014). An increasing recognition that many social issues are global challenges (from climate change to pandemics or human trafficking), means that international co-operation regarding data

sharing is ever more pressing. Methods for linking data are also advancing, with a wealth of literature reporting a range of probabilistic techniques to maximise linkage rates and improve linkage quality (Winkler *et al.* 2015). Such developments are vital, because governments do not typically use common person identifiers across different ministerial departments.

However, it is only very recently that we have witnessed justice-focussed initiatives which aim to accelerate the safe use of linked administrative data. Although there are examples of US studies in particular that have used integrated administrative crime and health records (Parsons and Sandwick 2012), the ability to *link* person-level criminal, civil or family court records to other data sources has been limited. This may reflect the sensitivity of justice records and that policy responses are highly politicised. However, important new initiatives are emerging in the UK. First, the Ministry of Justice (MoJ) led the creation of an important data-share, joining education and justice data, developing the WATCH Tool (MoJ analysis – Profile | Tableau Public [no date](#)) which provides an overview of all children who entered the family justice system between 2010 and 2016. This was followed by major investment from UKRI ESRC in the MoJ's *Data First* programme as part of ADR UK (2020–2023), which has already made available a number of important justice datasets through the Office for National Statistics, with more to follow (Data First: Harnessing the potential of linked administrative data for the justice system – ADR UK [no date](#)). The overarching aim of the Data First initiative is to enable independent academic research, to add value and capacity to statistics produced routinely by the MoJ. A key objective of the recently established Nuffield Family Justice Observatory (FJO) is to support the use of linked data for applied family justice research (Broadhurst *et al.* 2018). Funded by the Nuffield Foundation, the Nuffield FJO is already making its mark in terms of generating vital intelligence about how the family justice system is working in England and Wales. Linked to this initiative, interdisciplinary teams are beginning to produce data resource profiles (Bedston *et al.* 2020, Johnson *et al.* 2020) and create data user groups designed to specifically support the family justice research community in England and Wales. Such resource profiles provide practical illustrations of data documentation that can be replicated in other jurisdictions. A number of exemplar studies are evidencing the value of this data to address pressing policy questions (Alrouh *et al.* 2019, Cusworth *et al.* 2021).

The relationships of interdisciplinary, data intensive research

The socio-legal contribution

Exploiting large-scale administrative data is an interdisciplinary challenge – it requires subject specialists to join forces with computer and data scientists, statisticians and epidemiologists. Although interdisciplinarity typifies socio-legal scholarship, the particular mix of disciplines required to produce reliable and meaningful research using large-scale, often messy administrative data, is more novel. The linking of data requires knowledge about the architecture of datasets and linkage methodologies. Using this data is far more than a search and count exercise. However, meaningful and accurate use of administrative data also requires specialist knowledge of the specific domain in question. For example, it is difficult for non-specialists to understand the difference

between legal orders made in the family courts and their implications for children and families, or to know how to condense the multiplicity of available legal orders for research purposes. Specialist knowledge is also required to identify errors in the data – for example, the combinations of orders recorded for children that are *not* possible within the law. Equally, subject specialists are needed to identify what administrative datasets cannot answer or what issues may be obscured by gaps in the data (Bryson *et al.* 2017). The restructuring of any data for research purposes must make both technical and *substantive* sense.

Working across knowledge boundaries is not always comfortable. Researchers must absorb new vocabularies and norms of working which can leave them less assured of their place or competence. Data intensive interdisciplinary research is not for the faint-hearted, scholars must tolerate a level of *not* knowing as they adapt their knowledge and understanding (Klein 1990, Lélé and Norgaard 2005). Specialists may also feel that they risk abandoning their core knowledge because keeping abreast of relentless social, political and legislative change is already a huge challenge. However, interdisciplinary research does not typically involve absorbing an entirely new discipline – new or peripheral knowledge complements core knowledge (Palmer 2013). This is a hybrid intellectual space, but one in which the primary discipline will not necessarily suffer as a consequence of integrating new learning.

In this context, it is vital that family justice scholars recognise the distinctive contribution they can make to data intensive research, by sharing their advanced knowledge of the specialist domain. Data do not speak for themselves – questions must be asked of data that are informed by a keen understanding of pressing policy and legal issues. Meaningful interpretation requires in-depth knowledge of a given domain. Where the naïve data scientist simply *prospects* a new subject domain (Slota *et al.* 2020), he or she may overlook the impact of key policy and legislative milestones on any time series analysis. As (Slota *et al.* 2020) write, data science can appear ‘curiously empty’ – where interdisciplinary research is approached as an act of solicitation or simple extraction in a new domain. New developments in population data science aim to avoid such pitfalls. Driven by an over-arching interest in public data for public good, such research groupings bring together individuals with different disciplinary backgrounds to deliver meaningful secondary analyses of administrative data.

Moreover, received concepts in statistics and social science gain new meaning when read through a family justice lens. For example, ‘population reconstruction’ is not simply a practical task of restructuring person-level records. Creating analytic categories is a critical conceptual exercise, which can reinforce or challenge normative structures of society (Fledderjohann and Roberts 2018). For example, social class is on the one hand a way of grouping individuals, but equally is a system of stratification that ranks people by their differential access to material, social and cultural resources. Class shapes lives in important and unequal ways – social theory brings a critical understanding of *context* to all stages of the design and conduct of research.

From a socio-legal family justice perspective, data is never actually *raw* – data are collected for specific purposes. Systematic biases in terms of missing data reflect issues in society – note the troublesome category ethnicity. Ethnicity is a loaded category, often poorly completed in administrative records because of sensitivities regarding definition, but also because it is a category that is continually contested and revised. Missing values

within datasets can also reflect absence of populations within justice institutions themselves. For example, the recent work of (Bedston *et al.* 2019), which quantified fathers' absence in child protection court records in England, uncovered the gendered nature of parents' recurrent appearances in public law children cases. The frequency with which fathers were absent from family court records could not simply be reduced to problems of recording; rather this absence reflected the marginal role that fathers can play in public law proceedings. The records unintentionally reproduce gender asymmetries. Moreover, no exercise of imputation will remedy this underlying gender bias, embedded in family justice systems.

Large-scale administrative data affords the opportunity of transforming the kinds of questions we can ask of our justice institutions, but use of this data is *not* and should not be atheoretical (Sarat and Silbey 1988, Monroe *et al.* 2015, Margetts 2017). The risk of an overly empiricist approach is that it reduces what it means to 'know' and to 'understand' to levels that are not helpful. Anderson's (2008) assertion that massive volumes of data 'forces us to view data mathematically first and establish a context for it later' is to misunderstand the research process. However, as Boyd and Crawford (2012) caution 'older forms of intellectual craft' (p.666) have a central place in data-intensive research, but risk being side-lined where it is simply the sheer volume of data that counts. Working across disciplinary boundaries requires a mindset of mutual respect, which is best fostered through durable teams, rather than fleeting or opportunistic engagements.

Collaborative engagement with data providers

Although administrative data is routinely produced, it is produced by public institutions, which are dynamic in nature. Shifting technical, policy and legislative landscape impact on how data is collected and coded (UK Statistics Authority, 2019). For example, modifications in the classification of offences in criminal justice can cause significant problems where data is simply taken at face value. Boundary changes are also typical in public agencies in regard to their geographic reach. Therefore, effective use of even the most standardised administrative data requires a keen understanding of the *operational* contexts in which this data is produced (Administrative Data Taskforce 2012). This is a particularly important consideration in the design and piloting of retrospective longitudinal studies that aim to mine data over a number of years, as described above. Thus, researchers must work very closely with those managing organisational data to understand changes in how data is categorised and stored – data owners are essential actors in the collaborative mix.

The interface between disciplines in collaborative teams has been subject to discussion. However, the data provider-academic interface remains far less visible in the published literature. Data providers hold key insights into the unique histories of their institutions' recording practices, but without opportunities for detailed discussion of methods of data collection and recording, mis-estimation is likely. Although metadata tables produced by institutions enable researchers to unpack an institution's coding methodology and any changes over time, the devil is in the detail. Close examination of records finds both formal changes to systems but also 'coding on the fly'. In response to legislative and policy change, agencies need time to make changes to categories of data collection. The immediate management of operations must take priority over concerns

about the reliability of archived data for research purposes. Pending changes in data infrastructures, interim strategies or workarounds may be implemented. Without knowledge of informal changes in data recording, the reliable tracking of individuals within systems over time is undermined (UC Data 1999, UK Statistics Authority 2015). Family justice scholars, with specialist expertise of major policy and legislative milestones and awareness of how the law works in practice, are well placed to spot system workarounds where technical infrastructures lag behind formal changes to policy or the letter of the law.

In addition, government departments and related institutions typically produce their own benchmarking data – published both internally and externally, as annual snapshots regarding demand and or performance (Ministry of Justice 2020). Where effective relationships are established with data producers, both sources are important in validating findings. Using an agency's internal statistics to establish the plausibility of initial descriptive statistics, is an important step in cross-validation (UK Statistics Authority 2015). Checking out observations with data producers also provides an opportunity for augmenting mutual understanding of important administrative data assets and improving their quality and scope. Researchers should not passively accept the limitations of administrative data; rather they should seek out opportunities to improve national data assets.

However, relationships with data owners extend beyond issues of data quality and reliability. With increasing data science capability within government departments and other legal service organisations, research is best conducted in the spirit of collaboration. The academy no longer has a monopoly on data or its analysis. Within organisations, we are witnessing a concerted effort to extract the maximal value from their data. Data cannot be viewed as simply the by-product of routine case management; rather it is a vital source of intelligence for agencies, equipping them to examine their own practices. Thus, an inclusive approach, which explicitly sets out opportunities for mutual exchange and capacity building, will achieve sustainable relationships of co-production. Data acquisition cannot be a simple 'smash and grab' exercise. Data producers hold keen intelligence about policy and legislative landscapes upon which they aim to make their mark, and this ensures their central place in any collective research effort.

Administrative data: mixing methods

The use of administrative data for quantitative research does not preclude complementary, qualitative or ethnographic enquiry. Where methods are mixed, considerable impact has followed research, because insights gained from quantitative analysis of large-scale data are enhanced through in-depth interviewing, direct observation of practice or manual review of representative case files. Although examples are limited, specific to family justice research, (Broadhurst *et al.* 2017, Trinder *et al.* 2017), in the wider policy literature there are many examples of research which combine methods, adding depth to insights gained from large-scale administrative data, through qualitative approaches. For example, (Grobe *et al.* 2017) in the US linked administrative records on families and children to data collected through surveys and in-depth interviews to examine employment instability and job characteristics of parents using child care subsidies.

Whilst use of population-level data can identify, for example, stark regional differences in rates of care orders for children, without further qualitative enquiry and deep consultation with stakeholders, explanation of phenomena falls short. Often those using quantitative or computational methods are accused of too much appeal to scientism – however, we would argue that setting different approaches to knowledge generation in opposition is counterproductive. Analyses that make use of large-scale datasets can be complemented by more in-depth qualitative inquiry. Thus, embracing large-scale data intensive research is about the *re-assembly* of new and ‘old’ methods, rather than the relegation of established repertoires to the history books.

Legal and ethical considerations: towards a social licence for justice research

Increasing access to person-level data raises multiple questions about incursions of privacy. Although use of population-level data can deliver vital societal benefits, these must be carefully balanced with citizens’ privacy concerns. Regarding institutions of justice, if trust is eroded in the security of personal data, this can undermine citizen-state interactions far beyond the life of any particular research programme. Citizen-state interactions in the context of justice, are arguably, fraught with tensions around rights, entitlements, privacy and disclosure, which create particular sensitivities regarding the secondary analysis of personal records.

Considerable progress has been made in the development of robust and transparent models of governance, which control all aspects of data sharing/re-use and storage, data access and analysis, publication and disclosure. The development of secure data sharing platforms, where personal data is anonymised prior to its handling by researchers, demonstrates best practice internationally in privacy protection and information governance (Ford *et al.* 2009, Lyons *et al.* 2009, Jones *et al.* 2019). The socio-legal research community can be assured of best practice, where approved and secure environments support their research.

However, ensuring legal compliance regarding data protection is not the same as securing public support, or what Carter *et al.* (2015) have referred to as a social licence for the re-use of personal data. An ethical approach to the secondary use of personal data requires more than attention to the legalities of data sharing. Writing on the subject of the *care.data* debacle in the UK, Carter *et al.* (2015) argued that where the proponents of this project failed was in securing public trust and public consensus for the secondary use of personal health records. Readers may recall that *care.data* aimed to extract data from NHS primary care records in England to provide accurate, timely information with the aims of informing all stakeholders about the treatment and care provided by the NHS. The objectives of this service were to improve transparency and accountability within the NHS and, ultimately, outcomes for patients. However, the programme folded due to public objection to the use of personal records. This project illustrates very clearly that having a legal mandate for data sharing, is not in itself sufficient. Legal authority alone does not guarantee social legitimacy. Achieving social legitimacy lies in careful and meaningful engagement with public audiences – over and above simple public relations exercises.

Regarding the community of justice stakeholders, we would add that the sociological concept of ‘publics’ more aptly describes the range of audiences that make up this community (Newman and Clarke 2009). Professionals, advocates, experts, academics and the media are all powerful interest groups with a range of strong opinions. All too often

references to public engagement fail to differentiate the diversity of publics who engage with questions of research value and data ethics. Different audiences bring different interests and agendas to their *listening* and have conflicting interests and values (Lehoux *et al.* 2012) – nowhere is this more so, than in the adversarial world of justice. While there is clearly a strong moral and economic argument for the safe, anonymous reuse of personal data for research, for all the reasons set out above, the challenge is in engaging with relevant publics, addressing their concerns and demonstrating the value of applied research (Waind 2020). Where public support is not garnered from diverse stakeholder groups, the progress that has been made in harnessing administrative data for research, could readily be reversed. A study conducted by Cameron *et al.* (2014) commissioned by the ESRC and the Office for National Statistics, in relation to public views on use of administrative data for research purposes, indicated that in general public had very limited awareness of the purpose or value of research, but were concerned about the security of personal data, and worried about data being leaked, lost or sold (Cameron *et al.* 2014). Therefore, the pursuit of a social licence for the sharing of highly sensitive personal justice data, albeit in safe, fully anonymised formats, must embrace a range of publics – in particular those in receipt of justice services, whose interests are often delegitimised in policy debates.

Caution: pathways to impact

The primary use of large-scale administrative data is to undertake policy and practice focussed analyses. For example, researchers use data generated by our public institutions to study their effects on lives, both in the short and longer term. However, no matter how novel or robust our research derived from administrative data, there is no ‘natural’ pathway between empirical observation and policy or legislative change. Policy makers may be eager for social science solutions to immediate policy issues, but the possibilities for revision are limited – particularly in the context of relentless downward pressure on public service budgets or more recently, economic or epidemiological shocks. Moreover, policy audiences are typically impatient and struggle to comprehend the timeframes for academic delivery, as well as the need to ask more fundamental questions about justice. The pursuit of efficiency gains in the delivery of justice has dominated systems in a number of international contexts, creating practice that is forever in crisis, reactive and, in some instances, exhausted. In search of solutions, the policy lens is therefore downstream on recommendations that may ease this strain, with limited interest in research method. For all these reasons, we cannot assume unproblematic relationships between the producers and receivers of knowledge, or indeed, the parties as set out above.

However, given the academy’s commitment to the public good, there is much value in achieving a compromise between the strong pragmatic impulse of policy and the academic preference for more open-ended critical questions (Campaign for Social Science [no date](#)). This requires a balance on all sides between affiliation and distinction, which positions the academic as not in the *service* of policy, but nevertheless, attuned to the workaday concerns of policy makers and practitioners. Increasingly, research intermediaries seek to address the gap between the worlds of academia, policy and practice. For example, the Nuffield Family Justice Observatory’s (England and Wales) primary focus is to support the production of actionable knowledge and/or to produce accessible summaries for frontline practice.

Discussion and conclusions

Justice systems underpin democracy – they serve to protect but also enable – and play a vital role in combatting abuse and challenging inequality. However, where basic questions about the impact of justice systems remain unanswered, it is difficult to argue that justice systems are doing their job. Using administrative data, particularly where different data sources can be linked, allows the researcher to address critical questions about the impact of justice systems on individual and family lives – and over time.

Although the family justice research community has limited experience in the use of administrative data at scale, scholars can learn from advances within the disciplines of epidemiology and health, in particular. Moreover, family justice scholars can make a fundamental contribution to interdisciplinary data intensive research, because all elements of research design, data collection and analysis require substantive as well as methodological expertise. As stated above, teams that can combine methodological knowledge with theoretically informed social science, are best placed to generate meaningful research about the family justice system.

With the further digitisation of justice through a range of client-facing solutions such as self-help apps, the development of online dispute resolution through to automated adjudication and algorithmic decision-making, it is imperative that family justice scholars keep pace with change. The observations of Savage and Burrows (Savage and Burrows 2007), in their articulation of a crisis in empirical sociology, are all the more prescient. They argued that sociologists risked losing their foothold in the world of social analysis given the proliferation of organisations, which both produce *and* are equipped to perform sophisticated analyses of their own social data. Moreover, are family justice scholars happy to leave analysis of ‘big data’ to those who contend like Anderson (Anderson, 2008) that theory is dead? Or do they wish to ensure that explanation is as important as actions of quantification and the like? Forming strong alliances with data producers as well as colleagues across a range of disciplines will enable the family justice community to make its mark, as research opportunities continue to transform in the context of an expanding landscape of digital justice data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Nuffield Foundation [FJO/43766].

Funder

Nuffield Foundation: grant number FJO/43,766

No involvement in any stage of the article production – only funding

References

- Administrative Data Taskforce, 2012. *The UK administrative data research network: improving access for research and policy*. London: Economic and Social Research Council.
- Alrouh, B., et al., 2019. *Born into care Wales: newborns and infants in care proceedings in Wales | nuffield foundation*. London: Nuffield Family Justice Observatory.
- Anderson, C., 2008. 'The end of theory: the data deluge makes the scientific method obsolete'. *Wired magazine*, 16 (7), 7–16.
- Australian Government Department of Health (no date) *Research data infrastructure initiative*. Available at: <https://www.health.gov.au/initiatives-and-programs/research-data-infrastructure-initiative> (Accessed: 6 July 2020).
- Baldwin, H., et al. 2019. Disentangling the effect of out-of-home care on child mental health. *Child abuse & neglect*, 88, 189–200. doi:10.1016/j.chiabu.2018.11.011
- Bedston, S., et al. 2019. Linked lives: gender, family relations and recurrent care proceedings in England. *Children and youth services review*, 105, 104392. doi:10.1016/j.childyouth.2019.104392
- Bedston, S., et al. 2020. Data resource: children and family court advisory and support service (Cafcass) public family law administrative records in England. *International Journal of Population Data Science*, 5 (1). doi:10.23889/ijpds.v5i1.1159
- Berk, R.A., Sorenson, S.B., and Barnes, G., 2016. Forecasting domestic violence: a machine learning approach to help inform arraignment decisions. *Journal of empirical legal studies*, 13 (1), 94–115. doi:10.1111/jels.12098
- Boyd, D. and Crawford, K., 2012. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15 (5), 662–679. doi:10.1080/1369118X.2012.678878
- Brinig, M.F. (2002) *Empirical work in family law, U. Ill. L. Rev.* Available at: https://scholarship.law.nd.edu/law_faculty_scholarship (Accessed: 5 March 2021).
- Broadhurst, K., et al. 2015. Connecting events in time to identify a hidden population: birth mothers and their children in recurrent care proceedings in England. *British journal of social work*, 45 (8), 2241–2260. doi:10.1093/bjsw/bcv130
- Broadhurst, K. et al. (2017) *Vulnerable birth mothers and recurrent care proceedings final summary report*. Available at: www.nuffieldfoundation.org (Accessed: 9 March 2021).
- Broadhurst, K., Budd, T., and Williams, T., 2018. *The nuffield family justice observatory for england and wales: making it happen*. London: Nuffield Foundation.
- Brownell, M.D. and Jutte, D.P., 2013. Administrative data linkage as a tool for child maltreatment research. *Child abuse & neglect*, 37 (2), 120–124. doi:10.1016/j.chiabu.2012.09.013
- Bryson, C., et al., 2017. *Understanding the lives of separating and separated families in the UK: what evidence do we need?* London: Nuffield Foundation.
- Byrom, N., 2019. *Developing the detail: evaluating the impact of court reform in england and wales on access to justice*. Guildford: Legal Education Foundation.
- Cameron, D., Pope, S., and Clemence, M., 2014. *Dialogue on data: exploring the public's views on using administrative data for research purposes*. London: Ipsos Mori Social Research Institute. *Campaign for Social Science* (no date). Available at: <https://campaignforsocialscience.org.uk/> (Accessed: 6 July 2020).
- Cane, P. and Kritzer, H., 2010. *The oxford handbook of empirical legal research*. Oxford: Oxford University Press.
- Card, D., et al., 2010. Expanding access to administrative data for research in the United States. In: *American economic association, ten years and beyond: economists answer NSF's call for long-term research agendas*. pp 1–4.
- Carter, P., Laurie, G.T., and Dixon-Woods, M., 2015. The social licence for research: why care data ran into trouble. *Journal of medical ethics*, 41 (5), 404–409. doi:10.1136/medethics-2014-102374
- Connelly, R., et al. 2016. The role of administrative data in the big data revolution in social science research. *Social science research*, 59, 1–12. doi:10.1016/j.ssresearch.2016.04.015

- Creutzfeldt, N., Mason, M., and McConnachie, K., 2019. *Routledge handbook of socio-legal theory and methods*. Abingdon: Routledge.
- Cusworth, L., et al., 2019. *Children looked after away from home aged five and under in Scotland: experiences, pathways and outcomes*. Stirling: University of Stirling.
- Cusworth, L., et al., 2020. *Who's coming to court? private family law applications in wales*. London: Nuffield Family Justice Observatory.
- Cusworth, L., et al., 2021. *Uncovering private family law: who's coming to court in England?* London: Nuffield Family Justice Observatory.
- Data First: Harnessing the potential of linked administrative data for the justice system - ADR UK (no date). Available at: <https://www.adruk.org/our-work/browse-all-projects/data-first-harnessing-the-potential-of-linked-administrative-data-for-the-justice-system-169/> (Accessed: 28 May 2020).
- Data, U.C., 1999. *An inventory of research uses of administrative data in social services programs in the United States*. Berkeley, CA: University of California. Available: http://ucdata.berkeley.edu/new_web/inventory/intro.html.
- Dunning, T., 2012. *Natural experiments in the social sciences: a design-based approach*. Cambridge: Cambridge University Press.
- Eekelaar, J. and Maclean, M., 1990. Divorce law and empirical studies: a reply. *Law quarterly review*, 106, 621–631.
- Eekelaar, J. and Maclean, M., 2013. *Family justice: the work of family judges in uncertain times*. London: Bloomsbury Publishing.
- Eekelaar, J., Maclean, M., and Beinart, S., 2000. *Family lawyers: the divorce work of solicitors*. Oxford: Hart Publishing.
- Elias, P., 2014. Administrative data. In: (Eds) Adrian Duşa, Dietrich Nelle, Günter Stock, and Gert G. Wagner *Facing the future: european research infrastructures for the humanities and social sciences*. Berlin: Scivero Verlag, 47.
- Evans, E., et al. 2010. Using administrative data for longitudinal substance abuse research. *The journal of behavioral health services & research*, 37 (2), 252–271. doi:10.1007/s11414-008-9125-3
- Fledderjohann, J. and Roberts, C., 2018. Missing men, missing infertility: the enactment of sex/gender in surveys in low-and middle-income countries. *Population horizons*, 15 (2), 66–87. doi:10.1515/pophzn-2018-0003
- Ford, D.V., et al. 2009. The SAIL databank: building a national architecture for e-health research and evaluation. *BMC health services research*, 9 (1), 157. doi:10.1186/1472-6963-9-157
- Genn, H.G., Wheeler, S., and Partington, M., 2006. *Law in the real world: improving our understanding of how law works: final report and recommendations*. London: Nuffield Foundation.
- Goerge, R.M. and Lee, B.J., 2002. Matching and cleaning administrative data. *New Zealand economic papers*, 36 (1), 63–64. doi:10.1080/00779950209544351
- Griffiths, L., et al. 2020. Maternal health, pregnancy and birth outcomes for women involved in care proceedings in Wales: a linked data study. *BMC pregnancy and childbirth*, 20 (1), 697. doi:10.1186/s12884-020-03370-4
- Grobe, D., et al. 2017. Using policy-relevant administrative data in mixed methods: a study of employment instability and parents' use of child care subsidies. *Journal of family and economic issues*, 38 (1), 146–162. doi:10.1007/s10834-016-9501-8
- Guttman, A., 2019. The SPOR Canadian data platform: opportunity for multi-provincial research. *Canadian medical association journal*, 191 (40), e1091–e1092. doi:10.1503/cmaj.191040
- Hafekost, K., et al. 2017. Maternal alcohol use disorder and subsequent child protection contact: a record-linkage population cohort study. *Child abuse & neglect*, 72, 206–214. doi:10.1016/j.chiabu.2017.08.010
- Hamann, H., 2019. The German federal courts dataset 1950–2019: from paper archives to linked open data. *Journal of empirical legal studies*, 16 (3), 671–688. doi:10.1111/jels.12230
- Harding, M. and Newnham, A. (2015) 'How do county courts share the care of children between parents', *Reading and Coventry: University of Reading and University of Warwick*. Available

online: www.nuffieldfoundation.org/sites/default/files/files/Full%20report.pdf . accessed 15. August 2020

- Harron, K., *et al.* 2020. Preterm birth, unplanned hospital contact, and mortality in infants born to teenage mothers in five countries: an administrative data cohort study. *Paediatric and perinatal epidemiology*, 24 (6), 645–654. doi:10.1111/ppe.12685
- Harron, K.L., *et al.* 2017. A guide to evaluating linkage quality for the analysis of linked data. *International journal of epidemiology*, 46 (5), 1699–1710. doi:10.1093/ije/dyx177
- Harwin, J., *et al.*, 2018. *Care demand and regional variability in England: 2010/11 to 2016/17*. Lancaster: Centre for Child and family Justice Research.
- Harwin, J., *et al.*, 2019. *The contribution of supervision orders and special guardianship to children's lives and family justice*. Lancaster: Centre for Child and family Justice Research.
- Harwin, J. and Alrouh, B., 2017. New entrants and repeat children: continuity and change in care demand over time. *Family law*, 47 (4) 404–411.
- Hitchings, E., Miles, J., and Woodward, H., 2013. *Assembling the jigsaw puzzle: understanding financial settlement on divorce*. Bristol: University of Bristol.
- Ho, D. and Kramer, L., 2013. Introduction: the empirical revolution in law. *Stanford law review*, 65 (6), 1195–1202.
- Holman, C.D.J., *et al.* 2008. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Australian health review*, 32 (4), 766–777. doi:10.1071/AH080766
- Howe, L.D., *et al.* 2013. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology*, 24 (1), 1–9. doi:10.1097/EDE.0b013e31827623b1
- Hunt, J. and Macleod, A., 2008. *Outcomes of applications to court for contact orders after parental separation or divorce*. London: Ministry of Justice.
- Huntington, C., 2018. The empirical turn in family law. *Columbia law review*, 118 (1), 227–314.
- Janson, C.G., 1996. Retrospective and prospective design and data. *European child & adolescent psychiatry*, 5 (1), 11–17. doi:10.1007/BF00538537
- Jay, M.A. *et al.* (2017) *Who cares for children? population data for family justice research*. Available at: <http://wp.lancs.ac.uk/observatory-scopingstudy/> .
- Johnson, R.D., *et al.* 2020. Data resource: population level family justice administrative data with opportunities for data linkage. *International journal of population data science*, 9, 1339.
- Jones, K.H., *et al.* 2019. A profile of the SAIL databank on the UK secure research platform. *International journal of population data science*, 4, 2.
- Klein, J.T., 1990. *Interdisciplinarity: history, theory, and practice*. Detroit, MI: Wayne state university press.
- Leeuw, F.L., 2015. Empirical legal research: the gap between facts and values and legal academic training. *Utrecht law review*, 11 (2), 19–33. doi:10.18352/ulr.315
- Lehoux, P., Daudelin, G., and Abelson, J., 2012. The unbearable lightness of citizens within public deliberation processes. *Social science & medicine*, 74 (12), 1843–1850. doi:10.1016/j.socscimed.2012.02.023
- Lélé, S. and Norgaard, R.B., 2005. Practicing interdisciplinarity. *BioScience*, 55 (11), 967–975. doi:10.1641/0006-3568(2005)055[0967:PI]2.0.CO;2
- Lery, B., Shaw, T.V., and Magruder, J., 2005. Using administrative child welfare data to identify sibling groups. *Children and youth services review*, 27 (7), 783–791. doi:10.1016/j.childyouth.2004.12.012
- Lyons, R.A., *et al.* 2009. The SAIL databank: linking multiple health and social care datasets. *BMC medical informatics and decision making*, 9 (1), 3. doi:10.1186/1472-6947-9-3
- Maclean, M. and Eekelaar, J., 2009. *Family law advocacy: how barristers help the victims of family failure*. London: Bloomsbury Publishing.
- Maclean, M., Eekelaar, J., and Bastard, B., 2015. *Delivering family justice in the 21st century*. London: Bloomsbury Publishing.
- Mania, K., 2015. Online dispute resolution: the future of justice. *International comparative jurisprudence*, 1 (1), 76–86. doi:10.1016/j.icj.2015.10.006

- Margetts, H., 2017. The data science of politics. *Political studies review*, 15 (2), 201–209. SAGE Publications Sage UK: London, England. doi:10.1177/1478929917693643.
- Masson, J., 2010. A new approach to care proceedings. *Child & family social work*, 15 (3), 369–379. doi:10.1111/j.1365-2206.2010.00685.x
- Masson, J.M., Pearce, J.F., and Bader, K.F., 2008. *Care profiling study (ministry of justice research series 4/08)*. Ministry of Justice and Department of Children, Schools and Families. London.
- Medalia, C., et al. 2019. Linking survey and administrative data to measure income, inequality, and mobility. *International journal of population data science*, 4 (1), 1. doi:10.23889/ijpds.v4i1.939
- Ministry of Justice (2020) *Family court statistics quarterly: july to September 2020* - GOV.UK. Available at: <https://www.gov.uk/government/statistics/family-court-statistics-quarterly-july-to-september-2020/family-court-statistics-quarterly-july-to-september-2020> (Accessed: 5 March 2021).
- MoJ analysis - Profile | Tableau Public (no date). Available at: <https://public.tableau.com/profile/moj.analysis#!/> (Accessed: 23 March 2021).
- Monroe, B.L., et al. 2015. No! formal theory, causal inference, and big data are not contradictory trends in political science. *PS, political science & politics*, 48 (1), 71–74.
- Newman, J. and Clarke, J., 2009. *Publics, politics and power: remaking the public in public services*. London: Sage Publications.
- OECD Expert Group, 2014. *OECD expert group for international collaboration on microdata access: final report*. Paris: Organisation for Economic Co-operation and Development.
- Orr, C., et al. 2019. Hospitalisations for maternal assault are associated with increased risk of child protection involvement. *Child abuse & neglect*, 95, 104014. doi:10.1016/j.chiabu.2019.05.007
- Palmer, C.L., 2013. *Work at the boundaries of science: information and the interdisciplinary research process*. Springer Science & Business Media. Dordrecht.
- Parsons, J. and Sandwick, T., 2012. *Closing the gap using criminal justice and public health data to improve the identification of mental illness*. New York: Vera Institute of Justice.
- Perkins, N.J., et al. 2018. Principled approaches to missing data in epidemiologic studies. *American journal of epidemiology*, 187 (3), 568–575. doi:10.1093/aje/kwx348
- Roos, L.L., Nicol, J.P., and Cageorge, S.M., 1987. Using administrative data for longitudinal research: comparisons with primary data collection. *Journal of clinical epidemiology*, 40 (1), 41–49.
- Sarat, A. and Silbey, S., 1988. The pull of the policy audience. *Law & policy*, 10(2-3) (2–3), 97–166. doi:10.1111/j.1467-9930.1988.tb00007.x
- Savage, M. and Burrows, R., 2007. The coming crisis of empirical sociology. *Sociology*, 41 (5), 885–899. doi:10.1177/0038038507080443
- Slota, S.C., et al. 2020. Prospecting (in) the data sciences. *Big data & society*, 7 (1), 1. doi:10.1177/2053951720906849
- Susskind, R.E., 2017. *Tomorrow's lawyers: an introduction to your future*. Oxford: Oxford University Press.
- The HMCTS reform programme - GOV.UK (no date). Available at: <https://www.gov.uk/guidance/the-hmcts-reform-programme#our-reform-programme> (Accessed: 5 March 2021).
- Trinder, L., et al., 2013. *Enforcing contact orders: problem-solving or punishment*. Exeter: Exeter Law School.
- Trinder, L., et al. 2014. Litigants in person in private family law cases. In: *Litigants in person in private family law cases*. London: Ministry of Justice. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/380479/litigants-in-person-in-private-family-law-cases.pdf
- Trinder, L. et al. (2017) *Finding fault? divorce law and practice in England and wales summary report*. Available at: www.nuffieldfoundation.org (Accessed: 9 March 2021).
- UK Data Forum, 2013. *UK strategy for data resources for social and economic research 2013-2018*. London.

- UK Statistics Authority (2015) *Quality assurance of administrative data*. Available at: https://osr.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-settingthestandar_tcm97-44370.pdf. (Accessed: 24 March 2021).
- UK Statistics Authority(2019) Administrative Data Quality Assurance Tool Kit. https://osr.statisticsauthority.gov.uk/wp-content/uploads/2019/02/qualityassurancetoolkit_updated_Feb19_2.pdf
- Waind, E., 2020. *Trust, security and public interest: striking the balance A review of previous literature on public attitudes towards the sharing and linking of administrative data for research*. London: Economic and Social Research Council.
- Winkler, W.E., et al. 2015. Probabilistic linkage. In: K. Harron, eds. *Methodological developments in data linkage*. Hoboken, New Jersey: John Wiley & Sons. 8–35.
- Woodward, H.D. and Sefton, M., 2014. *Pensions on divorce: an empirical study*. Cardiff: Cardiff Law School, Cardiff University.
- Woollard, M., et al. 2014. Administrative data: problems and benefits. A perspective from the United Kingdom. In: A. Dusa, eds. *Facing the future: european research infrastructures for the humanities and social sciences*. Berlin: SCIVERO, 49.
- Zylbersztejn, A., Gilbert, R., and Hardelid, P., 2020. Developing a national birth cohort for child health research using a hospital admissions database in England: the impact of changes to data collection practices. *PLoS one*, 15 (12), e0243843. doi:10.1371/journal.pone.0243843