emerald PUBLISHING

Journal of Children's Se

Considerations in the use of local and national data for evaluating innovation in children's social care

Journal:	*Journal of Children's Services*
Manuscript ID	JCS-12-2020-0081.R2
Manuscript Type:	Technical Paper
Keywords:	innovation, evaluation, children's social care, quantitative datasets, contextual safeguarding, recurrent care removals

Considerations in the use of local and national data for evaluating innovation in children's social care

Abstract

Design/methodology/approach

This paper examines the use of data routinely collected by local authorities as part of the evaluation of innovation. Issues entailed are discussed and illustrated through two case studies of evaluations conducted by the research team within the context of children's social care in England.

Purpose

We explore the possibilities in using such national, statutory datasets for evaluating change and the challenges of understanding service patterns and outcomes in complex cases when only a limited view can be gained using existing data. Our discussion also explores how methodologies can adapt to evaluation in these circumstances.

Findings

The quantitative analysis of local authority data can play an important role in evaluating innovation but researchers will need to address challenges related to: selection of a suitable methodology; identifying appropriate comparator data; accessing data and assessing its quality; and sustaining and increasing the value of analytic work beyond the end of the research. Examples are provided of how the two case studies experienced and addressed these challenges.

Originality/value

The paper discusses some common issues experienced in quasi-experimental approaches to the quantitative evaluation of children's services which have, until recently, been rarely used in the sector. There are important considerations which are of relevance to researchers, service leads in children's social care, data and performance leads, and funders of innovation.

Implications of the research for policy and practice

- Quasi-experimental methods can be beneficial tools for understanding the impact of innovation in children's services, but researchers should also consider the complexity of children's social care and the use of mixed and appropriate methods.
- Those funding innovative practice should consider the additional burden on those working with data and the related data infrastructure if wishing to document and analyse innovation in a robust way.

- Data which may be assumed to be uniform may in fact not be when considered at a multi-area or national level, and further study of the data recording practice of social care professionals is required.

Introduction

The term 'innovation' in the context of social care encompasses a range of forms of transformation in systems and practices to improve service experiences and outcomes (Mulgan, 2007). For the purpose of their Innovation Programme in Children's Social Care, the Department for Education (DfE) listed various areas where they felt innovative approaches were needed to meet their aims of improving life chances for children who receive help from the social care system and ensuring better value for money across the sector.[1] These included: enhancing understanding of the system conditions for improving practice; encouraging innovation through incentives; and creating a strong evidence base through experimenting with, and replicating, successful new approaches so that 'what works' could be disseminated across the sector.

In recent years there has been a gathering momentum for such evidence-based public policy, as illustrated by the development of a 'What Works Network' [2], and the What Works Centre for Children's Social Care (WWCCSC).[3] This emphasis on evidence-based social policy can also be understood to have its roots in a particular conceptualisation of methodological rigour, as exemplified by the WWCCSC terminology of 'robust standards of evidence'. Similarly, the second round of funding in the DfE Innovation Programme required evaluation designs that utilised the Maryland Scientific Methods Scale (Gottfredson *et al*., 2002).

This context places highest value on experimental and quasi-experimental methods, but may simultaneously devalue case studies and the qualitative methods that can provide insight into the mechanisms of change and the experiences of those implementing and using the service (Flyvbjerg, 2006). The selection of the most appropriate approach for the research question should always be the foremost consideration (Sebba et *al.*, 2017). Recent guidance on evaluation in clinical research highlights the need for attention to complexity (Craig *et al.*, 2019) which does not only lie within the intervention *or* the system, but in the interaction between the two:

> 'The gap between the evidence-based ideal and the political and material realities of the here-and-now may be wide. [...] The articulations, workarounds and muddling-through that keep the show on the road are not footnotes in the story, but its central plot. They should be carefully studied and represented in all their richness' (Greenhalgh and Papoutsi, 2018, p.2).

This depiction of complexity is highly relevant to understanding the efficacy of innovation in social care and demands a pluralistic approach to evaluation. Further complexity in understanding "what works" in children's social care is added through the ever-changing landscape of policy, legislation

and regulation in which children's social care departments operate. These changes include defining new forms of harm and requiring local responses to that harm, new time limits and performance measures, and thematic inspections focusing on one particular form of harm (ADCS, 2018). Nonetheless, the growing application of experimental methods within the social sciences, and in children's services in particular, offers opportunity to further develop an evidence base for services and to improve outcomes for children, young people, and families. Furthermore, the rigorous record keeping involved in child welfare decision-making offers a large quantity of data available for further study of practice and decision-making in ways that are yet to be fully utilised. The application of robust experimental methods requires data which is structured, uniform, reliable, and consistently collected over time. In social care, due to the local variation in service provision, the main source of standardised data is statutory returns to central government departments – primarily those datasets which are routinely submitted to DfE for national monitoring of statutory responsibilities.[4]

These data submissions are guided by annual documentation (which we go on to describe below) ~~which~~ that must be completed and submitted by local authorities, often through an annual process of converting their own local data into the accepted format. These standardised and longitudinal datasets open possibilities for researchers to apply more robust methodologies, particularly in comparing data before and after an intervention began, and comparing data in one local authority (LA) to others.

However, although these data offer researchers in social care a great possibility when seeking to understand and evaluate services, there are also a number of challenges to overcome in order to utilise them appropriately and responsibly (both to the people the data were collected on and the services which collected them). In this paper, we aim to describe some of these challenges to the use of local and national data within the evaluation of innovative social care services. We begin by describing two case studies of evaluations from the Department for Education's second round of Innovation Funding, and proceed to discuss the learnings from these evaluations in terms of the considerations for researchers utilising large data sets. We group these considerations under four headings:

1. Choice of methodology
2. Identifying comparators / counterfactual
3. Data availability, suitability, and quality
4. Sustainability

Importantly, it should be noted that this is not a methodological or statistical paper; and although we describe some of the statistical approaches taken in the two case studies, this is for the purpose

of understanding the nature of the data used. We do not discuss the merits of using one statistical test over another, nor do we provide in-depth details of the methodologies used. Instead, this paper describes the wider process of accessing data, discusses why researchers might take terms such as 'uniformity' in relation to national and local social care data with a pinch of salt, and ultimately makes the case for further exploration, utilisation and application of the vast volume of data being routinely collected by the sector. Readers interested in further methodological detail may refer to the full evaluation reports for each case study referenced below, and their technical appendices.

Two case studies of evaluation

In this paper we explore the challenges and possibilities of working with administrative data in complex evaluation through the lens of two case studies of evaluation conducted by the authors in England between 2017 and 2020, funded through the government in Round 2 of the DfE's Children's Social Care Innovation Programme:

i) The development and piloting of a Contextual Safeguarding system in the London Borough of Hackney ([redacted] *et al*., 2020); and

ii) The second wave of development of the Pause project, in multiple LAs across England ([redacted] *et al*., 2020).

Both were mixed methods evaluations, but we focus here on the quantitative aspects, which were rooted deeply in the legal, policy and administrative systems of children's services in England, making use of readily available data. The quantitative methodologies of each of these evaluations are briefly described here and the challenges they entailed are then considered in depth through this paper. Further detail of the full mixed methods employed can be found in the evaluation reports.

Contextual Safeguarding

The Contextual Safeguarding Project was funded by the DfE 2017-20 to redesign parts of the safeguarding system in Hackney in line with Contextual Safeguarding Theory, as developed by Firmin (2015; 2017). The new system aimed to target (through assessment and intervention) the extra-familial contexts in which young people faced risk or harm during adolescence – for example, peer groups, public spaces and school environments where sexual and criminal exploitation, peer-on-peer abuse, serious youth violence, and gang affiliation occurred.

The evaluation broadly followed a pre-post intervention and external comparator methodology, comparing service patterns, child-level outcomes, and the experiences of young people, families, communities and professionals in Hackney before and after the introduction of Contextual Safeguarding. Quantitative comparisons were also made with three LAs judged to be statistically

similar, to provide a counterfactual. These 'Statistical neighbours' were identified by comparing all English LAs on a number of factors relevant to EFRH, such as youth crime figures and wider care data. These factors are described in the following section. ~~(further details below).~~

The Child in Need (CiN) Census and SSDA 903 (Children Looked After Return) were used for these comparisons as they held information relevant to Contextual Safeguarding aims (explored further below) and there was assumed to be some degree of uniformity in data submissions across the four sites. Data could also be provided to the evaluators in anonymised form, ensuring that no personal information was included, and data were shared using a Secure File Transfer Protocol. The specific elements used of these datasets is shown in Table 1. As the concept of extra-familial risk and harm (EFRH) was emergent, with no distinctive factors agreed cross-nationally for its identification, the methodology for this quantitative evaluation was largely exploratory. A framework was developed which could enable understanding of whether and how cases of EFRH could be identified in pre-existing data, which indicators could be used to capture it, and whether changes could be seen over time. Due to the shortage of variables which directly signified EFRH, and limited granularity in the data, determining which factors could be used as proxy measures of EFRH formed a substantial area of focus (~~further details below).~~ Using this framework, changes over time in outcome variables for cases which might be considered to reflect a high degree of EFRH could be measured in each LA. The process then involved comparing the data across three years to identify any changes over time in Hackney before and after the introduction of Contextual Safeguarding, and between Hackney and the comparator LAs.

Insert Table 1 about here

Pause

Pause was funded in the first round of the DfE Innovation Programme (2015-17) to deliver trauma-informed, relationship-based practice over an 18-month period to support women at risk of recurrent child removals due to safeguarding concerns. Involvement in Pause is voluntary, and there is a requirement that women agree to use an effective reversible method of contraception during the period of intervention in order to focus on themselves during this time (Pause, 2017). A second round of funding (2017-20) enabled scale-up of the service in nine new areas in England as well as a pilot programme of support for care-experienced women who have had one or more children removed.

Pause is a large, multi-area project, which had begun at different time points across English LAs. It was determined that the ~~most appropriate~~ nationally-available dependent variable most appropriate for the evaluation to examine was ~~was~~ the number of children under 12 months old taken into care

within a LA. This had been set as a key outcome indicator for Pause, based on research by Broadhurst *et al.* (2015; 2017), most notably the findings that:

> "Approximately 1 in 4 birth mothers appearing as respondents in an index set of s.31 care proceedings are expected to re-appear in a subsequent set of proceedings within 7 years… approximately 70% of women who return to court do so in proceedings that concern an infant who is born subsequent to, or during previous proceedings" (Broadhurst *et al.,* 2017, p.5).

~~, based on review of existing research on recurrent removal (e.g., Broadhurst *et al*., 2015; 2017), the findings of Pause's Wave 1 evaluation (McCracken *et al*., 2017), and given that fewer children being taken into care is a key outcome within the Pause Theory of Change.~~

Due to the greater availability of longitudinal data, Pause sites which had been operating since the first round of DfE funding were selected as the 'experimental' group, and a matching process was undertaken in order to select a suitable 'control' group. These were sites where no other recurrent care service was in place, with additional matching based on IDACI (Income Deprivation Affecting Children Index) scores and recent Ofsted full inspection status. Following this refining process, 22 LAs were compared in the longitudinal trends on the dependent variable, and seven were matched which had sufficiently similar trends to the Pause LAs.

A difference-in-difference (DiD) methodology was selected so that changes in children entering care could be compared between Pause and non-Pause areas over time (DiD compares the changes in trends between two or more areas before and after an intervention is put in place and assumes that cases and comparators have parallel prior trends pre-intervention). In this case, we investigated whether the trend for rates of children under 12 months old entering care differed before and after Pause, and if this was different to a matched comparator without a Pause service. ~~Comparators were selected based on similarity to Pause authorities on a number of factors, including poverty, the absence of Pause or any other recurrent care service and prior parallel trends in the outcome of interest.~~

The data considerations identified across these two evaluations

Choice of evaluation methodology

The first task in the evaluation of innovative services is the selection of appropriate methodologies (Craig *et al*., 2019). While RCTs are often positioned as offering the highest standard of evidence, there are barriers to their effective use in the context of children's social care (Stewart-Brown *et al.,* 2011; Dixon *et al*., 2014; Mezey *et al*., 2015). These include concerns about the need to allow randomisation to select the support a child or family receives, rather than this being based on the

expert judgement of a trained professional foregrounding a child's best interests, and meeting the RCT requirement for a protocol to be in place before an intervention starts, which is not always possible in the fast-moving policy environment.

Quasi-experimental methods offer a statistical approach where randomisation is considered unfeasible or unethical, or where analyses are to be conducted on historical data (Harris *et al*., 2006). These approaches place the burden of valid comparison onto the purposeful selection of comparators, rather than randomisation. This is typically through identifying subjects who are as close to identical to those receiving the intervention as possible, or by 'controlling' for any identified differences statistically. However, in complex social contexts, and particularly in the case of innovation, it is not always clear what variables must be controlled for in order to create the 'perfect' comparators necessitated by experimental methods. Research teams require not only statistical and methodological acumen, but expertise in the relevant policy and practice contexts, including in-depth understanding of the nature of local authority data sets and quality assurance processes which inform the identification of comparators.

With both the Contextual Safeguarding and Pause evaluations, the selection of methods first meant getting a detailed understanding of each project, their desired outcomes and impact, and therefore the dependent variables which might indicate successful implementation of their programmes; this was initially through each project's theory of change. Both projects sought to innovate by developing practice in areas which are somewhat outside of traditional social care domains. For Contextual Safeguarding, this meant assessing and addressing EFRH outside of the family home. And for Pause, innovation inhered to the attention of the service to the *women* who have experienced child removal as the focus of support – rather than to their children – with the overarching aim of reducing risk of potential future children entering care. Both of these approaches therefore presented a challenge methodologically as multi-area statutory data related to these interventions did not exist; English case management systems for child welfare services typically work on the premise that a single child is the 'subject' of data, not the family, other groupings, or environments.

Contextual Safeguarding addresses EFRH that is specific to the contexts that young people inhabit and construct, including individual relationships, peer groups, physical locations, and online. However, context has previously been somewhat absent from case management systems, as have connections between young people involved in social networks, who frequent similar locations, or who are joined in a number of other complex ways through social or geographical factors. As such, the pre-existing data did not allow for the identification of cases linked through external or contextual factors and thus prevented measurement of change over time for these particular young

people. For instance, there is no means by which all recorded cases of harm to a child from alcohol use in a specific location can be identified.

In respect of Pause, children removed from the same mother cannot readily be detected in the statutory datasets as there are no identification numbers for mothers or families in statutory returns. Furthermore, such identifiers would be challenging to use due to the often complex and changing nature of family groups. However, without an indicator in the data which shows whether the same mother has been subject to multiple removals, further study of repeat removals of children can only be achieved through in-depth, rigorous and time-consuming investigation of public court proceedings data (Broadhurst *et al.*, 2015).

Both evaluations therefore highlighted limitations in the available data which affected the nature of methodologies selected to consider process and outcomes, including the need to identify proxy measures of impact. For Contextual Safeguarding, a longlist of data items from several statutory data returns was refined into those which were deemed relevant to the project aims. This process was supported by previous work by the Rees Centre, University of Oxford, in the first phase of Innovation Programme funding (Sebba et al., 2017) and subsequent work from DfE to identify features of practice and outcomes from innovation in children's social care (Spring Consortium, 2018). Through this, several factors were identified which could be compared longitudinally between Hackney and comparator LAs, including those which social workers categorise in assessments of children considered to be 'in need'. Due to the higher level of specificity in these factors (compared to the codes used to indicate why a child becomes 'looked after'), these offered a higher chance of pinpointing children where harm or risk of harm may derive from outside of the family. The factors which were identified as proxy measures for EFRH were: alcohol misuse by the child; drug misuse by the child; child going missing; socially unacceptable behaviour; gang-related; child sexual exploitation. All are readily identified in the literature as strongly associated with the most common forms of EFRH, i.e. child sexual exploitation, criminal exploitation/County Lines, serious youth violence, and gang-affiliation (Firmin, 2020). Two further secondary factors – sexual abuse and neglect – were also identified as relevant where they occurred in conjunction with one or more of the primary factors (Klatt *et al.,* 2014). There is limited guidance available to social workers or analysts as to how these assessment factors should be identified or recorded. There is no standardised test or threshold for any factor, and so whether or not a factor is recorded depends on the practitioner and the organisational culture of assessment and recording in any given authority. In some sense, it is this practice and recording culture that is being measured, as much as the underlying levels of need in the community.

For Pause, the ideal dependent variable for comparison would have been the number of children removed from women who had previously had one or more children removed. However, data on repeat removals are not part of routine data submissions in national statistics. The only records held about parents and children experiencing child removal are held by the Child and Family Court Advisory and Support Service (CAFCASS), which supports children involved in family court cases, including cases ending in child removal from the family through a court order (Broadhurst *et al*., 2017); however, early conversations with Pause indicated that work was carried out with women who have children placed through voluntary (Section 20) and informal arrangements as well as through formal proceedings. SSDA 903 data, therefore, record removals that would not appear in Cafcass data. Furthermore, extracting case-level information on recurrence in either individual LA or Cafcass data would have required greater resource than was available to the evaluation. Given that any removals of children born post-intervention to women who have worked with Pause are likely to take place early in children's lives (cf. Broadhurst *et al.,* 2017), it was judged that a focus on infant care entry should maximise the possibility of detecting effects. The variable used was the rate of children under the age of one entering care each year. As with the factors in assessment, this administrative data is not solely measuring underlying need in the community, but the practice and organisational culture around infant child removal. This is not inappropriate, given the aspiration to stimulate wider system change in attitudes and practice with women involved in the child protection system.

Identifying comparators / counterfactual

Judging the counterfactual – what would have happened in the absence of an innovation – is a central consideration in the development of an evaluation methodology. This presents a challenge in the evaluation of large-scale innovations in social care, which – especially if designed to produce systemic change – may have effects on all children and families living in the area where innovations occur, or on all practitioners who work in these areas. In such cases, there will be no suitable comparison group *within* the area. Equally, differences *between* areas are wide-ranging and complex: even if they are implementing similar interventions, they may have different populations, or other provision (e.g. accessibility of related services, such as specialist mental health provision) which has a confounding effect on the dependent variable. Such factors may not be readily visible to researchers without significant investment of time and resource.

Moreover, other LAs do not necessarily wait for a robust evaluation before adopting a promising idea: both Pause and Contextual Safeguarding began to be taken on as frameworks in other LAs for the development of services with similar objectives and target populations over the course of the

research period [6]. This has implications for selection of comparators in the research design, as these areas cannot be considered to represent 'treatment as usual' or 'control' conditions.

For both evaluations, outcomes and comparators were selected at LA level. This decision, to focus on area-level effects, arguably sets a high bar for assessing change, as evidence of change may be diluted by characteristics of those who have not received the intervention. However, area-level comparison is valuable for illuminating potential cost savings or population level outcomes at LA level, and so can inform decisions about investment in innovative services in a context of welfare austerity, when LAs have significant constraints on their budgets (Harris et al. 2019). An alternative approach of matching individuals receiving the intervention with those who were not, either in the same authority, or a different authority was not undertaken. Comparison with individuals within the authority was not appropriate for Contextual Safeguarding given the system level changes anticipated. The exploratory approach to developing an indicator of EFRH over the course of the study meant that this could not be used as a matching variable for young people in other authorities at an individual level. It could be used in such a way in subsequent studies. For the Pause evaluation, individual level matching was unfeasible because of the difficulty in getting data about individuals and families subject to recurrent care discussed above. Several tools and indexes support refinement of suitable LA comparators in children's social care. For instance, the IDACI provides a local measure of income deprivation weighted for the number of children in an area. Other local statistics related to social care outcomes and social (in)equality are also widely available on an annual basis. The Local Authority Interactive Tool (LAIT) can also be used to identify 'statistically similar' comparator LAs, based on a mixed basket of measures – although tools such as LAIT are limited in their capacity to identify comparators based on the specific context of an intervention. For instance, where a social care intervention is focussed on improving educational outcomes, similarities in education and schooling may be considered as holding greater weight for matching LAs than other variables such as the looked-after population.

The two evaluations undertook different approaches to identifying comparators, based on the considerations noted above. For Contextual Safeguarding, the priority was to find comparators which more closely reflected Hackney in terms of EFRH. For Pause, it was more important to consider whether potential comparator areas had alternative interventions to Pause but which also addressed recurrence. It was also necessary to think about what other factors might influence change in the dependent variable (very young children entering care) and to select comparators with similar levels of these other factors. For Contextual Safeguarding, the comparison was with similar contexts; for Pause, the comparison was with LAs with similar trajectories over time. This reflects the

theoretical underpinning of the interventions – one working with the relationship between harm and context and the other the relationship of harm and events spaced out over time.

The first step in finding comparators for Hackney was to compare the borough with all LAs in England on a number of standardized variables, selected from a list of available data. These variables were selected by a working group of professionals from Hackney, the University of Bedfordshire, and the evaluation team, based on their relevance to Contextual Safeguarding. The expert panel were asked to rank a longlist of 47 variables sourced from DfE and Ministry of Justice regional data and the LAIT: 21 variables appeared in the 'top 10 list' of two or more of the expert panel. Following this, statistical similarity was judged using Euclidean distance for all shortlisted variables, to mirror the approach taken in the LAIT, but using a subset of variables selected to better reflect Contextual Safeguarding. All LAs were compared and those which were most similar to Hackney on a compound measure of these 21 variables were short-listed. Of these, two London and one non-London comparator were selected based on a further selection exercise which made considerations about the comparators' ability and willingness to be involved in the wider evaluation. Consideration was also given to the location of acceptable comparators and one non-London site was selected on this basis. Further scoping was carried out by the research team to ascertain the status of Contextual Safeguarding work in each of these comparison areas.

For Pause, the selection of comparators was driven by the selection of the difference-in-difference (DiD) methodology. Crucially, DiD requires that the intervention and the control group have a similar (ideally parallel) trend in the dependent variable over time. Thus, for DiD analysis, each Pause LA required a matched LA which had a similar trend over the pre-Pause years for the rate of children under 12 months old entering care each year. Furthermore, none of these comparators could have a non-Pause recurrence service, and it was also decided that they needed to be similar in terms of IDACI scores and social care provision (as judged via Ofsted ratings). These factors were identified from the academic and grey literature as influential on the rate of care entry in a given authority (Bywaters *et al*., 2017; Bryant *et al.,* 2015). Subsequent research has strengthened the argument for including IDACI scores in selecting comparators for studies of infant care entry (Bilson and Bywaters, 2020). A stepwise process of refining LAs in England based on these criteria enabled the identification of seven LAs who had no known recurrence service, and matched trends for rates of removals over time to the Pause LAs before 2015 (when the Round 1 DfE investment in Pause began). Details of the matching methods used can be found in the technical appendix of the evaluation report.

Another important consideration worth noting from the Contextual Safeguarding evaluation, in particular, is the burden placed on comparator areas through involvement in any study. Even in the case of utilising statutory data returns which must be submitted annually, there is a requirement for data teams to work with researchers, and conduct all of the necessary data protection and due diligence associated with this. Furthermore, any qualitative elements or novel data collection of any kind will inevitably take practitioner time (e.g., identifying participants, securing preliminary consents). In the case of these evaluations, care was taken to provide feedback to the comparators with the intention of supporting practice and providing additional analytical capacity as a benefit of involvement with the research.

As highlighted by these two evaluations, the selection of comparators in children's services, and particularly in evaluating innovative practice, depends on the nature of the innovation and the availability of relevant data. Importantly, there is not a one-size-fits-all approach to identifying comparators, particularly at a LA-level. Although tools such as LAIT are useful in the early stages of comparing areas, the particular context of an innovation demands more rigorous exploration of similarities and differences relating to the intervention context. Locally available data may not hold the information required to make these judgements, and researchers need to ensure they have considered the resources necessary for this early preliminary investigation. For instance, it is rarely clear from the available data what other relevant interventions are being conduct in comparator LAs and, although methods may be designed to benefit from the uniformity of statutory data, these data may not be sufficient for the selection of comparators in the first instance.

Data availability, suitability, and quality

Once a method and sample population are selected, the next step is assessing the availability, suitability and quality of data. In the case of data related to children's services, there are additional considerations related to data protection due to the sensitivity of the content. Considerations of data type and data access and the type of data required is interlinked with the choice of comparators and the choice of methods.

The data most commonly used to evaluate children's social care interventions are collected and stored in case management systems within each LA, and annual data returns are submitted to various government departments on an annual basis. Summary statistics for each LA are published annually, with a six- to nine-month time lag. Statutory data collections and related guidance support consistency in the measures collected from one year to the next and across LAs. Despite the existence of detailed guidance for statutory data returns, there is often substantial variation in data recording within and across LAs. More significantly, thresholds for intervention (and therefore

recording and reporting) change over time in response to changes in budgets (Hood *et al.,* 2020), or increases in demand at the "front door" (Broadhurst *et al*., 2010). A child in need in one year in one authority might not be classed as one in another time or place.

In the statutory datasets, there are challenges in identifying individuals over time, across datasets and across LA boundaries, due to the use of various local personal identifiers that do not persist from year to year or place to place. In the context of complex social care innovations, such as Pause and Contextual Safeguarding, this is particularly pertinent as children and families may have lived complicated, complex and disrupted lives which see them interacting with statutory services repeatedly over a number of years and moving between LAs.

Innovative thinking may change our views on which data are important. Pause and Contextual Safeguarding, for instance, have highlighted the absence of national welfare data on mothers, families and siblings, extra-familial contexts, and peer groups. By working with parents after a child is removed and before any subsequent pregnancy, Pause works in the space between statutory social work interventions. Arguably, case management systems in England are not yet equipped to collect data on this type of innovative practice, nor do so in an inconsistent manner. The SSDA 903 data, for instance, takes the child as the individual unit of analysis but does not provide links between this child and other significant people in their lives, such as carers, parents, siblings, and peers. Women experiencing repeat removals of children from their care are not visible in the data; only the entry into care of individual children can be seen.

Similarly, statutory data returns provide limited information about the type and volume of support provided and whether this support is provided to individual children, whole families, or a wider group of young people. When innovative services are providing support that is atypical, for example to peer groups rather than individuals, or to parents without children currently in their care, this support is often recorded in bespoke databases and recording systems. These systems are often not linked by personal identifier to the statutory social care system. This poses a barrier for understanding the relationship between this support and the need for other forms of social care intervention – a key requirement for understanding the costs and benefits to the public purse. Furthermore, services may be funded by local or national government, but delivered by third sector or private providers, leading to a further disconnect between different data management systems.

To access statutory data, researchers must either make use of the limited summary statistics published each year, arrange direct access to the detailed statutory data returns via government departments, or work in partnership with individual LA data teams to arrange access to the raw outputs of the case management systems. This direct access allows for the extraction of more

detailed data about individual children and includes data not requested as part of the statutory return process, but it does place a burden on data teams to arrange and clean the data for use. Based on the data required, the level of access required can differ greatly. With Pause, the decision to use rates of infant care entry as the dependent variable in the DiD calculation meant the anonymised dataset was publicly available at a LA level through annual DfE statistics, and hence no personal information was shared. In contrast, the Contextual Safeguarding evaluation required Hackney and the three comparators to provide anonymised individual level data related to the EFRH factors identified for young people; this increased the task of data access significantly for both the research sites and evaluators.

Given the flexibility afforded to LAs in England to choose and customise their own case management systems in social care, data accessed from one area may look significantly different from another. Indeed, in the Contextual Safeguarding evaluation, many of the data tables returned were in different formats, missing certain variables, and sometimes missing entire tables. Given the selection of uniform, statutory data, this was somewhat surprising and highlighted that even datasets considered 'generic' may require a great deal of further processing and cleaning before analyses can take place. Establishing a relationship with LA data professionals enables subsequent requests and clarification questions, if needed. In the case of direct access to LA data, often a detailed understanding of local case management systems is required, which may be gained through direct or controlled access to the system under supervision of a LA representative and with necessary permissions from a data lead and director of service.

Where the data in statutory returns are insufficient and additional data are required, or where statutory data requires cleaning or manipulation to be suitable, additional costs may be incurred by the evaluation team and by the participating LAs. Innovation funding may not consider the additional (and often significant) costs of changes to data management systems as part of the innovations. The time of data managers may not have been factored in, nor the additional time burden on practitioners who might have been asked to undertake extra administrative work in recording their practice and new outcomes. Therefore, researchers must work alongside practice to identify what data is available, what additional work might be required to standardise it for research purposes, and the extent to which data are suitable for answering research questions.

When comparing data across LAs, it is also necessary to consider whether there is consistency in the 'sentiment' of the data, that is whether those entering the data are using the same mental models and definitions when considering which box to tick, or category to use. For instance, in the case of social workers' Section 17 (Child in Need) assessments, there are 41 factor codes which can be listed

as being identified at the end of an assessment (DfE, 2019). For each assessment, multiple factors can be identified, and each child may have multiple assessments. However, what is not clear from the data alone is how different practitioners might identify and record the presenting information when carrying out an assessment. Might it be common practice in one LA to record just the key factors identified and not all factors, but different elsewhere? Might there be variations in how codes are used based on the location or even the different case management systems? These are not easy questions to answer and, as yet, we are not aware of any research which investigates this further. However, from the Contextual Safeguarding evaluation, it was clear that there was some variability in how LAs recorded data, such as how 'No Further Action' codes were used and where in the timeline these were typically recorded for cases.

Ultimately, this uncertainty around the reliability of data presents some difficulties. Although clear and well-thought-out methods can be selected, and suitable comparators identified, there remains a possibility of ~~'type 1 error'~~statistical error or bias, where a positive effect of the intervention may in fact be a result of differences in data recording. This risk can be mitigated through a multi-method approach, building a mosaic of understanding, including an investigation of data usage as part of a wider evaluation programme. In the case of Contextual Safeguarding, the development of data systems aligned with Contextual Safeguarding Theory was recognised as an important next step in supporting LAs to adopt the approach.

In terms of suitability of data, despite the structured datasets available in children's services, when it comes to understanding innovative approaches that had not been envisaged at the point the dataset was designed, the data may in effect be 'flawed, uncertain, proximate and sparse' (Wolpert and Rutter, 2018, p.1). A particular barrier is the separation of data about children "looked-after" by the local authority and those who are "in need" but not looked-after into two separate datasets. This adds additional data cleaning and processing challenges for researchers wanting to understand a child's journey from the community into care, and back again (Bowyer et *al*., 2018).

A related issue is the focus of the innovation itself. Although the primary datasets available relate to statutory intervention, an innovation may be situated within an unrelated, or only partly related, service area – for example, seeking to prevent the escalation of need before the threshold for statutory intervention is reached. In such cases, and without a similar uniform dataset at the edges of the care system and across multiple service providers, researchers may have difficulty identifying meaningful data for the changes they hope to assess (Bowyer *et al*., 2018).

Sustainability

Finally, researchers should consider the possibilities of long-term benefits from their analytic work, as well as the implications for what happens when evaluations are completed and there is no longer additional resource from funders for collecting and analysing data. Indeed, the data produced for evaluation, and the systems put in place to collect those data, may have a wider benefit in supporting ongoing learning and improvement beyond the life of the evaluation. In this sense, research and evaluation might contribute to the sustainability of innovation itself, rather than just acting as an external arbiter of change. Given the length of time it takes for innovation in practice methods or service structures to embed (Sebba *et al*., 2017), this would also provide crucial insight into whether an innovation is sustainable and improves outcomes for children.

Funders may also wish to consider how to maximise the value of evaluations they have funded and the usage of the analytical resources developed. This might include analyses which can be rolled-out nationally, utilised across other projects, or incorporated into annual data analysis exercises. These contributions to wider learning are both methodological – such as the application of a new analytic approach – and instrumental – the provision of tools for data collection and analysis to enable ongoing examination of key datasets. Without a central repository for social care data and related tools, this process of scaling up analytic practices in the sector can only happen in a piecemeal fashion.

The Pause evaluation utilised a DiD methodology which has not received much usage within evaluations of children's services. The evaluation report documents the suitability of this methodology, important considerations, and caveats and ultimately demonstrates the effectiveness of the analytic approach, including its capacity to link with economic analysis to demonstrate projected cost savings associated with the intervention ([redacted] *et al*., 2020). This offers an early case study in the use of DiD in social care for evaluation, and there may be merit in exploring its use in evaluating other innovations in the future, as well as offering the opportunity for ongoing evaluation of Pause (including, potentially, in comparison with other services targeting risk of recurrence). The use of statutory data from published datasets makes this method very accessible to a wide range of researchers.

Quantitative methods commonly entail substantial early work in developing analysis models and securing data access; however, many analytic processes can be automated (for instance through the development of scripts in programming languages such as R and Python). As such, researchers are in a position to offer additional future benefits to LAs and the wider sector by sharing those resources and additional outputs. For the evaluation of Contextual Safeguarding, a large amount of code was

developed in order to standardise and analyse the CiN Census. A central component of the analyses was exploration of the factors identified at assessment and, in particular, how these could be used to look for correlations in different types of risk (see [redacted] *et al.*, 2020). Individual data reports were provided for each of the comparator sites in order to support further thinking about EFRH in their areas at the end of the evaluation. This approach to better understanding different 'types' of cases and further identifying those with higher identified risk of harm from extra-familial sources might provide a footprint for future usage of these statutory datasets by researchers.

Importantly, these considerations of sustaining analytic capacity should not be a final thought at the conclusion of evaluations. Researchers have the opportunity to provide additional analytic capacity to LAs and service providers throughout their involvement, providing support and resource to teams which may not have the flexibility to explore new methodologies themselves.

Conclusion

The evaluations of the Contextual Safeguarding and Pause innovation projects required careful considerations of how best to utilise existing data within quasi-experimental multi-method designs to assess relative change associated with the interventions. As set out in this paper, these considerations can largely be grouped into four areas:

- Selecting a suitable methodology;
- Identifying appropriate comparator data;
- Accessing data and assessing its quality;
- Sustaining and increasing the value of analytic work beyond the end of the research.

These four components are not a suggestion of a stepwise method for quantitative evaluation in children's services, nor are they an exhaustive list of the many considerations of which researchers must be aware. Rather, they represent important interlinked considerations when conducting quantitative evaluation in social care. For instance, the availability of data might influence the selection of method, and vice versa. Both might influence how comparators are selected, and a difficulty in selecting a comparator might limit the use of particular methods and require new approaches to be introduced.

These components also intersect with other critical methodological considerations for evaluation of social care innovation. We have noted the increasing emphasis on rigorous experimental and quasi-experimental approaches prevalent in the UK and elsewhere, but not all LAs and organisations have the resources to conduct RCTs and nor is the method always the most appropriate (or ethical) for a

particular innovation approach (cf. Stewart-Brown *et al.,* 2011; Craig *et al.*, 2019; Greenhalgh and Papoutsi, 2018). The evaluations discussed here highlight the potential of quasi-experimental approaches which utilise pre-existing data, as part of a mosaic of understanding within a complex multi-method design. The variations in the nature of data and how it is recorded, even in so-called standardised returns, mean such analyses should be approached with a healthy level of scepticism. This is particularly pertinent in the context of innovation, which may be having effects, as yet not fully understood, on the use of the data systems themselves.

Although not the subject of this paper, our two evaluations also highlighted the necessity of combining quantitative with qualitative methods if the process and outcomes of innovation were to be fully understood – including, crucially, from the perspectives of those implementing the new service or approach (practitioners) and those children, families and communities involved with the service or affected by its outcomes.

Across LAs and in research teams around England, the current interest in innovation as a way of addressing social problems and perceived service deficits has factored in the substantial funding for innovation projects and their evaluations. However, the resulting large volume of analysis sits in siloes and, at the time of writing, there is no repository for analytical approaches in children's social care research, nor is there a means by which researchers developing analytical models can contribute their work for the benefit of LAs more widely. Given the public and statutory nature of this work, this might be considered a missed opportunity to develop sustainable approaches to understanding service provision, and so to build an evidence-based understanding of how social care innovation can ensure the rights and best interests of children. Admittedly, this type of unified approach would require additional infrastructure and care to safely navigate ethics processes and data protection; however, the benefits across the sector would be significant. As more LAs go on to experiment with innovation, and both Contextual Safeguarding and Pause are increasingly being explored in other LAs, collating standard, re-useable approaches to understanding impact would be a significant contribution to the sector.

The methods outlined in this paper, and described more fully in the technical appendices of the relevant evaluations, offer some applicable approaches to evaluation of innovations, which could be useful to local areas experimenting in this way. To make use of the methods, local areas would need access to either administrative individual-level data from other authorities or information about service design (such as the presence of a recurrent care service and the design of that service), as well as the analytic capability and capacity to gather, clean and access the required datasets. As noted, such arrangements are time-consuming and resource intensive. One approach to addressing

these challenges might be longer-term collaboration between groups of local authorities to run experiments and act as a comparison group, justifying the costs of data sharing agreements and analytic capacity for longer term gain. Another approach might be for the DfE to fund and support improved access by LAs to the national repository of individual-level data for the purpose of comparison and analysis. The methods discussed in this paper are a compromise. They seek to explore innovative practice using pre-existing service data. Ultimately, 'innovation' challenges previous conceptualisations of social problems and how they have been addressed by the social care sector, and seeks to improve service experiences and outcomes through (sometimes radical) transformation of practice methods and systems (Mulgan, 2007). We argue that this challenge is mirrored in the value attributed to pre-existing service data. When innovators try to generate new data to reflect the new practice methods, practitioners may suddenly find themselves needing to record things that their current case management systems are unequipped to process, and data leads may suddenly find themselves in ownership of data which doesn't fit into routine data submissions. It may be discovered that local data no longer reflects services or desired outcomes as innovations become embedded, and changes to methods of collection and analysis may be necessary. Applying these methods to pre-existing service data provides an interim solution, to help decision-makers understand changes in the patterns of service use and emerging impact on budgets, but are no substitute for investment in the recording and reporting infrastructure to support new ways of thinking and practicing with children and families. As such, researchers are an important piece in the innovation puzzle and have a duty to support ongoing conversations in which data might be useful going forward, if innovations are judged to be effective.

In conclusion, the quantitative analysis of LA data can play an important role in evaluating innovation. However, the complexity of children's social care and the very nature of innovation means that the use of experimental methodologies is rarely straightforward. Researchers and evaluators must carefully consider their usage of data and, further still, should see themselves as part of the innovation process through the development of methodologies which may be of value to the sector beyond their involvement.

[1] https://www.gov.uk/guidance/childrens-social-care-innovation-programme-insights-and-evaluation
[2] https://www.gov.uk/guidance/what-works-network
[3] https://whatworks-csc.org.uk/
[4] https://www.gov.uk/government/collections/childrens-social-care-statistics
[5] https://www.gov.uk/government/publications/local-authority-interactive-tool-lait
[6] https://csnetwork.org.uk/
[7] https://contextualsafeguarding.org.uk/portfolio-items/national-and-london-scale-up-project/
[8] https://www.pause.org.uk/what-we-do/where-we-work/

References

ADCS (2018) *Safeguarding Pressures: Phase 6*, The Association of Directors of Children's Services Ltd, available at https://adcs.org.uk/assets/documentation/ADCS_SAFEGUARDING_PRESSURES_PHASE_6_FINAL.pdf

Bilson, A., & Bywaters, P. W. B. (2020), "Born into care: evidence of a failed state", *Children and Youth Services Review*, Vol 116, 105164, pp.1-7.

Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). "Secondary use of EHR: data quality issues and informatics opportunities", *Summit on Translational Bioinformatics*, pp.1–5, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041534/.

Bowyer, S., Gillson, D., Holmes, L., Preston, O. and Trivedi, H. (2018), *Edge of Care Cost Calculator: Change Project Report,* Research in Practice, Dartington.

Broadhurst, K., Wastell, D., White, S., Hall, C., Peckover, S., Thompson, K., Pithouse, A., Davey, D. (2010), "Performing 'Initial Assessment': Identifying the Latent Conditions for Error at the Front-Door of Local Authority Children's Services", *The British Journal of Social Work*, Vol. 40, No.2, pp.352–370.

Broadhurst, K., Alrouh, B., Yeend, E., Harwin, J., Shaw, M., Pilling, M., Mason, C. and Kershaw, S. (2015), "Connecting events in time to identify a hidden population: birth mothers and their children in recurrent care proceedings in England", *The British Journal of Social Work*, Vol. 45, No. 8, pp.2241–2260.

Broadhurst, K., Mason, C., Bedston ,S., Alrouh, B., Morriss, L., McQuarrie,T., Palmer, M., Shaw, M., Harwin, J., and Kershaw, S. (2017), *Vulnerable Birth Mothers and Recurrent Care Proceedings: Final Summary Report,* Centre for Child and Family Justice Research, Lancaster.

Bryant, B., Parish, N., and Rea, S. (2016), *Action Research Into Improvement In Local Children's Services: Final research report*. Local Government Association, available at

https://www.local.gov.uk/sites/default/files/documents/160621_LGA%20children%27s%20services%20improvement%20action%20research_final%20report.pdf.

Bywaters, P., Brady, G., Bunting, L., Daniel, B., Featherstone, B., Jones, C., Morris, K., Scourfield, J., Sparks, T. and Webb, C. (2018), "Inequalities in English child protection practice under austerity: A universal challenge?", *Child & Family Social Work*., Vol. 23, No.1, pp.53– 61.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M. (2019), *Developing and Evaluating Complex Interventions: Updated Guidance*, Medical Research Council, available at https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/.

DfE (Department for Education) (2019), *Children in Need Census 2020 to 2021: Guide*, Department for Education, London.

Dixon, J. Biehal, N., Green, J., Sinclair, I., Kay, C. and Parry, E., (2014), "Trials and Tribulations: Challenges and Prospects for Randomised Controlled Trials of Social Work with Children", *The British Journal of Social Work*, Vol. 44, No. 6, pp.1563–1581.

Firmin, C.E. (2015), *Peer on Peer Abuse: Safeguarding Implications of Contextualising Abuse between Young People within Social Fields.* Professional Doctorate Thesis. University of Bedfordshire, available at https://core.ac.uk/download/pdf/29822361.pdf.

Firmin, C. (2017), "Contextual Risk, Individualised Responses: An Assessment of Safeguarding Responses to Nine Cases of Peer-on-Peer Abuse", *Child Abuse Review*, Vol. 27, No. 1, pp. 42-57.

Firmin, C. (2020) *Contextual Safeguarding and Child Protection: Rewriting the Rules*, Routledge, Abingdon, Oxon.

Flyvbjerg B. (2006), "Five misunderstandings about case-study research", *Qualitative Inquiry*, Vol. 12, No. 2, pp.219-245.

Gottfredson, D.C., Sherman, L.W., and (2002), "Maryland scientific methods scale", in Sherman, L.W., Farrington; D.P., Welsh, B.C. and Mackenzie, D.L. (Eds.) *Evidence-Based Crime Prevention*, Routledge, London, pp.13-21.

Greenhalgh, T. and Papoutsi, C., (2018), "Studying complexity in health services research: desperately seeking an overdue paradigm shift", *BMC Medicine*, Vol. 16, No. 95, pp. 1-6.

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E. and Finkelstein, J. (2006), "The use and interpretation of quasi-experimental studies in medical informatics", *Journal of the American Medical Informatics Association*, Vol. 13, No. 1, pp.16–23.

Harris, T., Hodge, L. and Phillips, D. (2019), *English Local Government Funding: Trends and Challenges in 2019 and Beyond*, The Institute for Fiscal Studies, London.

Klatt, T., Cavner, D. and Egan, V. (2014) 'Rationalising predictors of child sexual exploitation and sex-trading', *Child Abuse and Neglect*, 38(2), pp. 252–260.

McCracken, K., Priest, S., FitzSimons, A., Bracewell, K., Torchia, K., Parry, W. and Stanley, N. (2017), *Evaluation of Pause: Research Report*, Department for Education, London.

Mezey, G., Robinson, F., Campbell, R., Gillard, S., Macdonald, G., Meyer, D., Bonell, C., & White, S. (2015), "Challenges to undertaking randomised trials with looked after children in social care settings", *Trials*, Vol. 16, No. 206, pp. 1-5.

Mulgan, G. (2007), *Ready or Not? Taking Innovation in the Public Sector Seriously*, Nesta, London.

Pause (2017), *Pause Framework,* Edition 1, Pause, London.

Bowyer, S., Gillson D, Holmes, L., Preston, O. & Trivedi, H. (2018), *Edge of Care Cost Calculator: Change Project Report*, Research in Practice, Dartington.

Sebba, J., Luke, N., McNeish, D., and Rees, A. (2017), *The Children's Social Care Innovation Programme: Final evaluation Report*, Department for Education, London.

Spring Consortium (2018), *Seven Features for Practice and Seven Outcomes: Learning from the Children's Social Care Innovation Programme*, Department for Education, London.

Stewart-Brown, S., Anthony, R., Wilson, L., Winstanley, S., Stallard, N., Snooks, H. and Simkiss, D. (2011), "Should randomised controlled trials be the 'gold standard' for research on preventive interventions for children?", *Journal of Children's Services*, Vol. 6 No. 4, pp. 228-35.

Wolpert, M. and Rutter, H. (2018), "Using flawed, uncertain, proximate and sparse (FUPS) data in the context of complexity: learning from the case of child mental health", *BMC Medicine*, 16, 82, pp.1-11.

Table 1 Data taken from the Child in Need Census and SSDA 903 dataset

Dataset	Data used in evaluation
Child in Need Census	Age
	Gender
	Ethnicity
	Source of referral
	Whether cases were progressed or No Further Action
	Primary need
	Number of referrals per child
	Factors identified at assessment
	Case duration
	Reason for case closure
SSDA 903	Placement type
	Placement location
	Out of borough placements
	Convictions of Looked After Children
	Placement / status change