International Journal of Population Data Science

Trust, security and public interest: striking the balance
A narrative review of previous literature on public attitudes towards the sharing, linking and use of administrative data for research

Elizabeth Waind[1,*]

Abstract

This narrative literature review explores previous findings in relation to the UK public's attitudes towards the sharing, linking and use of public sector administrative data for research. A total of 16 papers are included in the review, for which data was collected between the years 2006-2018.

The review finds, on the basis of previous literature on the topic, that the public is broadly supportive of administrative data research if three core conditions are met: public interest, privacy and security, and trust and transparency. None of these conditions is sufficient in isolation; the literature shows public support is underpinned by fulfillment of all three. However, it also shows that in certain cases where the standard of one condition is very high – particularly public interest – this could mean the standard of another may, if necessary, be lower. An appropriate balance must be struck, and the proposed benefits of sharing and using data for research must outweigh the potential risks. Broad, conditional support for the use of administrative data in research has not only been found consistently, but has also been held over time.

Most studies identified by this review have focused on exploring the views of the general public towards the acceptability of administrative data use in broad terms. However, with the exception of that related to healthcare data, the review identified little work focused on gaining input from relevant demographics and communities in relation to specific data types or areas of research. In addition to fulfilling the core conditions of public support identified by broader work, initiatives making use of administrative data should aim to seek the views of relevant sub-sectors of the public in the development of research in relation to specific issues.

Keywords

literature review; review; public attitudes; public views; administrative data; administrative data research; data sharing; public engagement; public involvement; public opinion; public sector data; data linking; data use

Introduction

Public sector administrative data – information originally created for operational purposes when people interact with public services such as schools and hospitals – is a valuable resource for research. When analysed, this existing wealth of data has the potential to provide valuable insights into society and highlight where change is needed to improve policy and service provision.

However, in the UK, administrative data is a largely untapped resource, with government departments and public bodies not routinely sharing their data with one another or academic researchers [1]. This is a missed opportunity, as linking data from across different areas of the public sector and making it available for research can provide valuable insights into how different services interact with one another, and how a person's experiences in one area of life may influence outcomes in another. This is important for a thorough understanding of how policy can work best to support people and enable them to thrive.

The UK government's 2017 Digital Economy Act [2] provides the legal framework for public authorities to share administrative data for research under Section 64 – 'Disclosure of information research for purposes'. This allows investments such as ADR UK (Administrative Data Research UK) – a programme funded by the Economic & Social Research Council (ESRC) with a mission to enable secure access to linked UK public sector administrative data for approved researchers working on projects in the public interest – to operate [3].

However, in addition to operating in line with this legal framework, it is essential those handling and using data do so openly and ethically, and in the knowledge that the public is supportive of how and why their data is being used. Administrative data includes all those who interact with public services and therefore most of the population; that's what

*Corresponding Author:
Email Address: elizabeth.waind@ukri.org (Elizabeth Waind)

makes it so valuable to a more thorough understanding of what works in public policy. If we are to use data about the public, this cannot be done without the public's support and, where possible, their input.

The UK Centre for Data Ethics & Innovation (CDEI), in its July 2020 report on 'Addressing trust in public sector data use' [4], stresses: "*The sharing of personal data must be conducted in a way that is trustworthy, aligned with society's values and people's expectations. Public consent is crucial to the long-term sustainability of data sharing activity*". When the public is not sufficiently consulted and informed about the use of their data, initiatives which hope to make better use of data for the benefit of society cannot hope to succeed.

A well-known example of a programme which failed to engage effectively with the public and thus experienced a detrimental loss of public trust was National Health Service (NHS) England's care.data initiative. Care.data, launched in 2013, aimed to link information from across different NHS providers in community, general practice and hospital settings to give a fuller picture of the different services and enable the improvement of patient outcomes [5]. In 2016, the programme was closed following criticism of its public communications campaign, with the public not having been sufficiently informed of how and why their data would be used [6]. Care.data is a prime example of the necessity of public support for initiatives which aim to make use of public sector administrative data.

Objectives

This review is primarily intended as a source of information for those handling and using public sector data for research, to enable them to operate in a way in which the public find acceptable. It has the following main objectives:

1. To explore and summarise attitudes of the UK public, as found by previous research, towards the sharing, linking and use of public sector administrative data for research, and the conditions under which it is perceived it should and should not happen;

2. In addition, to be a source of advice on approaches to public engagement for organisations and researchers working with administrative data. This objective will be addressed in the final section of the paper, 'Approaches to public engagement'.

Definitions

For the purpose of this review, the following key terms are defined:

Administrative data is information created when people interact with public services, such as schools, the NHS, the courts or the benefits system. It is not originally created for research, but as a by-product of government services [1].

Anonymised data, as defined by the UK Information Commissioner's Office (ICO) [7], refers to "*data in a form that does not identify individuals and where identification through its combination with other data is not likely to take place*" (p.48). Safeguards such as those set out under the 'Five Safes' [8] – Safe people, Safe projects, Safe settings, Safe outputs and Safe data – provide the conditions under which identification of de-identified data is not likely to take place, therefore making data anonymous.

De-identified data refers to data which has had all personal identifying elements such as names and addresses removed, meaning individuals are no longer directly identifiable. The UK Digital Economy Act Research Code of Practice and Accreditation Criteria [9] states: "*Data must be de-identified before they can be made available so that the data do not directly identify individuals and are not reasonably likely to lead to an individual's identity being ascertained*".

It is important to note both 'de-identified' and 'anonymised' data are referred to across the literature reviewed, with definitions not consistently provided in all cases. This does not detract from the fundamental findings of the studies included in this review; however, when the terms are used in the context of a previous study's findings, their precise definitions should be considered with some caution.

Personal data, as defined by the ICO [10] in the context of the European Union General Data Protection Regulation (GDPR), is "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.*"

Public engagement, as defined by Rowe and Frewer in their 2005 'A typology of public engagement mechanisms' [11], is a combination of three concepts: 'public communication' (a one-way flow of information to the public); 'public consultation' (in which the opinions of the public are sought but no dialogue is involved) and 'public participation' (an exchange of information between the public and those leading the initiative in question). The literature reviewed for the purpose of this paper falls under 'public consultation' and 'public participation', with a mixture of consultation (for example, surveys) and participatory (for example, focus groups) methods adopted across the included studies.

Transparency, as defined by the ICO [12] in the context of GDPR, "*is about being clear, open and honest with people from the start about who you are, and how and why you use their personal data*".

Methods

To meet the above objectives, a review of published public consultations and attitudinal studies on the topic has been completed, with a focus on work conducted in the UK. This is not a systematic review, nor a review of the quality of previous work; it is a narrative review summarising the main trends identified across relevant previous literature.

The studies were identified via an unsystematic online search without fixed search terms. This involved broad searches using internet search engines, and browsing the websites of relevant data infrastructures and social research organisations. Examples of terms used in online searches (for reference and not fixed or exhaustive) include: 'public' *and* 'views' *or* 'attitudes' *or* 'consultation' *and* 'administrative data' *or* 'public sector data' *and* 'research' *or* 'sharing' *or* 'linking' *or* 'use'.

As this is an unsystematic review and many of the relevant papers are independently published and therefore not found in academic publications or databases, an unsystematic approach was a more effective way of identifying literature than being confined to particular search terms within specific databases and potentially omitting relevant non-academic papers. As described by Grant and Booth [13], a literature review may or may not include comprehensive searching or quality assessment.

Only papers relevant to the attitudes of the UK public towards the sharing, linking and use of public sector administrative data for research were selected. Some of the studies reviewed cover attitudes to data use more generally, not only in relation to research, but are nevertheless relevant to the aims of this review. Literature in relation to any type of public sector administrative data and any type of research topic was considered. Literature not considered relevant and therefore not reviewed included: papers focused solely on commercial access to public sector data; papers concerned with the linking of public sector data to private sector data; and those focused more broadly on the public's knowledge of, but not attitudes towards, the collection and storage (and not necessarily use) of data.

In total, 16 papers were identified as relevant for inclusion in the review, for which data was collected between the years 2006-2018, therefore covering over a decade of recent work. This includes mostly independent papers published by data infrastructures, research institutions or public bodies, as well as academic research papers and existing reviews of previous research on the topic. Table 1 lists the 16 studies included in the review and their main characteristics.

Overview of existing literature

Overall, this review finds that previous relevant public attitudes work in the UK has largely focused on the acceptability of the sharing, linking and use of data amongst the general public, rather than amongst specific sub-sectors of the public. It has mainly explored the conditions under which data sharing, linking and use is broadly acceptable, therefore offering insight into appropriate approaches for data infrastructures and researchers in general. The literature reviewed broadly identifies three core conditions of public support: public interest; privacy and security; and trust and transparency. These will be discussed in turn below.

Aside from studies focused around the use of healthcare data, this review identified little previous work dedicated to seeking the views of relevant demographics or communities in relation to specific data types or areas of research. The emphasis on healthcare data as opposed to other types of data may reflect a recent finding of analysis by the ESRC that, in the UK, healthcare data has been used for research to a greater extent than other types of public sector administrative data [14]. Nevertheless, in general, this review finds that the primary focus of previous literature has been to explore whether research using administrative data is, on the whole, acceptable, and under what conditions, with less input sought on how research in relation to specific issues can best hope to meet the interests and concerns of relevant data subjects.

Existing public knowledge of administrative data research

In general, the literature reviewed has found that existing public knowledge of how public sector data is currently shared, linked and used for research is low.

Most participants of a public consultation titled 'Dialogue on Data: Exploring the public's views on using administrative data for research purposes' [15] – the intention of which was to inform the approach of the Administrative Data Research Network (ADRN), the predecessor to ADR UK – for example, attached some value to social research more broadly, though some questioned its value to begin with and compared research findings to common sense.

A systematic review of studies investigating public responses to the sharing and linking of health data for research by Aitken et al. [16] found participants of several studies were reported as being surprised that data is not more widely used, and considered not using data for research to be wasteful. Similarly, the findings of research undertaken by the Office for National Statistics (ONS) over the period 2009-2013 [17] indicated that nearly half of the public assume government already links data about the population on a routine basis and holds it in a central data store. Study participants have also been found to express confusion between the use of data for research as opposed to for the everyday operation or activities of a public body or service [15, 17].

The studies reviewed suggest that, as knowledge increases over the course of public consultation, so does support for administrative data research [13, 16, 18–20]. This suggests that when the public has a better understanding of the value of research and the safeguards in place to protect data, they are more supportive of the use of administrative data for that purpose. These findings demonstrate the need for greater transparency and more effective communication of the use of administrative data for research and its benefits.

Public interest

The literature reviewed has widely found public interest (also termed 'public good', 'public benefit' or 'social value') to be the primary driver of support for the sharing and use of administrative data.

Participants of the 'Dialogue on Data' [15], for example, argued that there should always be social value associated with social research. Aitken et al., in a 2018 discrete choice experiment examining public preferences regarding the linking of health data for research [21], found the most common preference (amongst 57% of respondents) regarding the purpose of data linking was that it should only be done for public benefit. A systematic review and ethical enquiry into public views on the use of patient data for research by Stockdale et al. [22] identified a similarly widespread willingness to share this data for research for the 'common good'.

Of respondents to the 2015 Northern Ireland Life and Times (NILT) Survey [23], 85% agreed that data should be used when there is a benefit to society, as long as the data can be anonymised and privacy is maintained. A public consultation by Ipsos MORI [24] on behalf of the UK

Table 1: Study characteristics

	Reference	Study aim	Key message(s)
1	Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. Science and Public Policy. 2016 May 11;43(5):713-723.	To explore perceptions of the role, relevance and functions of trust (or trustworthiness) in relation to research practices.	The public's relationships of trust and/or mistrust in science and research are not straightforward; public trust is highly conditional and variable.
2	Aitken M, de St. Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Medical Ethics. 2016;17(73).	To explore current evidence on the public acceptability of data sharing and data linkage practices.	There is widespread (conditional) public support for data sharing and linkage for research purposes, though a range of concerns exist.
3	Aitken M, McAteer G, Davidson S, Frostick C, Cunningham-Burley S. Public Preferences regarding Data Linkage for Health Research: A Discrete Choice Experiment. International Journal of Population Data Science. 2018;3(11).	To examine the relative importance of several conditions upon which public support for research conducted through data linkage or sharing is contingent.	There is public support for the linking of health data and use by university and health service researchers without private sector involvement and with independent oversight. The type of data being linked and how profits are managed and shared are the two most important factors shaping preferences.
4	Cameron D, Pope S, Clemence M. Dialogue on Data: Exploring the public's views on using administrative data for research purposes. Ipsos MORI Social Research Institute. 2014.	To explore public understanding and views of administrative data and data linking.	The public would be broadly happy with administrative data linking for research projects provided (i) those projects have social value, broadly defined (ii) data is deidentified, (iii) data is kept secure, and (iv) businesses are not able to access the data for profit.
5	Davidson S, McLean C, Cunningham-Burley S, Pagliari C. Public acceptability of cross sectoral data linkage: Deliberative research findings. Scottish Government Social Research. 2012 Aug.	To explore views of the public on the acceptability of linking personal data for statistical and research purposes.	The public is broadly supportive of data linkage, particularly for health research. However, support is conditional and ambivalences and concerns exist, including unease about private sector access to public sector data.
6	Davidson S, McLean C, Treanor S, Aitken M, Cunningham-Burley S, Laurie G, Sethi N, Pagliari C. Public acceptability of data sharing between the public, private and third sectors for research purposes. Scottish Government Social Research. 2013 Oct.	To enhance understanding of sensitivities around data sharing between the public, private and third sectors for statistical and research purposes.	Concerns and sensitivities around data sharing between the public, private and third sectors cluster around: security and privacy; data uses and the public interest; labelling; statistical disclosure; and transparency. There is a strong case for ongoing public engagement in the development of policy and strategy.
7	Davies M, Jones H, Conolly A. Public Attitudes to Data Linkage: A report prepared for University College London by NatCen Social Research. NatCen Social Research. 2018 March.	To explore understanding and perceptions of data linkage, particularly between health examination survey data and administrative records.	Individuals could see the benefit of providing personal data if there was personal or societal benefit. Several factors underpinned views and concerns, including: trust and legitimacy of organisations; timeframe for consent to data linkage; and transparency.
8	Ipsos MORI. The Use of Personal Health Information in Medical Research: General Public Consultation. Medical Research Council. 2007.	To identify public concerns and misconceptions surrounding use of personal health information for medical research.	If the public is informed about what medical research entails, they are generally positive towards it. However, confidentiality and consent feature highly in the debate over data use.

Table 1: Continued

	Reference	Study aim	Key message(s)
9	Office for National Statistics. The Census and Future Provision of Population Statistics in England and Wales: Public attitudes to the use of personal data for official statistics. 2014 March.	To explore public attitudes towards the collection and use of data for production of official statistics and research.	The public are supportive of data sharing when personal or public benefit can be demonstrated; and public views differ according to who is using the data and for what purpose.
10	Oswald M. Share and share alike? An examination of trust, anonymisation and data sharing with particular reference to an exploratory research project investigating attitudes to sharing personal data with the public sector. SCRIPTed. 2014 Dec;11(3):245-272.	To explore attitudes to sharing personal data with the public sector.	The benefits-versus-costs problem in relation to the sharing of personal data is significant: the more tangible and/or immediate the benefit, the stronger the correlation to (and possibly the cause of) comfort in data sharing.
11	Rempela ES, Barnett J, Durrant H. Public engagement with UK government data science: Propositions from a literature review of public engagement on new technologies. Government Information Quarterly. 2018 Oct;35(4):569-578.	To examine the potential for public engagement with data science.	Government data science public engagement should: consider the varied and many 'publics'; not assume providing information will lead to acceptance; determine the contingencies of trust through trustworthy practice; incorporate robust, critical, and ongoing deliberation; and be holistic, moving beyond privacy and consent.
12	Robinson G, Dolk H, Dowds L, Given J, Kane F, Nelson E. Public Attitudes to Data Sharing in Northern Ireland: Findings from the Northern Ireland Life and Times Survey 2015. Ulster University. 2018 Feb.	To explore attitudes to data sharing amongst the Northern Ireland public.	Public support for data sharing sits on three pillars: trust in organisations, data protection measures, and public benefit. If any of these are reduced or taken away, public support falls.
13	Royal Statistical Society. Royal Statistical Society research on trust in data and attitudes toward data use/data sharing. 2014.	To get a snapshot of public trust in institutions handling their data, and attitudes towards data linkage and privacy.	There is a 'data trust deficit': trust in institutions to use data appropriately is lower than trust in them in general. When there are safeguards and a case for public benefit, more take a positive view in favour of data use and sharing.
14	Stockdale J, Cassell J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland (Revised). Wellcome Open Research. 2019 Jan;3(6).	To explore patient and public views on patients' medical data being used for research; to understand and map these views onto established biomedical ethical principles.	The public generally support the use of patient data for research, but demand that projects: are conducted in a secure way to prioritise privacy and minimise harm; set research objectives primarily concerned with the common good; and do this in a spirit of transparency and inclusivity of stakeholder views.
15	Tully MP, Hassan T, Oswald M, Ainsworth J. Commercial use of health data – A public "trial" by citizens' jury. Wiley: Learning Health Systems. 2019;3.	To investigate what informed citizens consider to be appropriate uses of health data in a learning health system.	Uses of anonymised patient data were considered appropriate by most when they could deliver public benefit. Positive health outcomes for patients were more acceptable than improved efficiency of NHS services.
16	Wellcome Trust. Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data. London: Wellcome Trust. 2013 July.	To understand the general public's attitudes to different types of personal data and data linking.	There is fear of personal data falling into the 'wrong hands' and widespread wariness about being 'watched'. Main benefits associated with storing personal data are convenience, advantageous offers and efficient customer service. Anonymity and consent issues are paramount.

Medical Research Council regarding the use of personal health information in research found that 70% of participants agreed the advantages of medical research – which the authors say are mainly considered 'societal' – outweigh the disadvantages (mainly seen as the disclosure of personal health information). Study participants have been found to consider financial profit an unacceptable motive for the use of administrative data [25–27]. A 2014 study by Marion Oswald investigating attitudes towards sharing personal data with the public sector [27] found most participants would be comfortable with their data being used to improve public services, but only 23% were comfortable with it being used by the NHS, and 27% by central government, to make profits to *fund* services.

Concerns have also been identified that some research using administrative data could inadvertently work against the public interest, for instance by causing certain demographic groups or local populations to be profiled or labelled [15, 16]. Aitken *et al.*'s review [16] identified concerns that policy based on analysis of large datasets may not sufficiently account for individual needs.

However, although previous literature has given indications as to what members of the public consider to be in the 'public interest', no widely understood definition of the concept appears to have been identified, and perhaps what matters more than defining the term is that the public perceives benefits of some sort. Understanding what the specific communities that administrative data research aims to impact perceive as the potential benefits of the work therefore remains an important goal of public engagement.

Data types

The literature reviewed has identified differences in the perceived sensitivity of, and potential benefit of, using different types of data for research. During the 'Dialogue on Data' [15], some participants expressed that some types of data – for example, records of domestic violence and data relating to HIV status – were too sensitive and personal to be shared outside of the agency that collected it. Nevertheless, by the end of the Dialogue the researchers found that participants were comfortable with data linkage using all types of data, as long as approval and security processes were in place.

Aitken *et al.*'s discrete choice experiment [21] found the type of data being linked to be the most influential factor shaping preferences regarding linking health data for research. How profits are managed and shared was found to be the second most influential factor, with the purpose of the research coming out third.

Meanwhile, 2013 research by the Wellcome Trust [28] found many regarded personal – as opposed to de-identified – health data differently to other types of data. Namely, they perceived an *"unquestionable benefit to people"* of experts having access to this type of information, especially in relation to illness [28, p.11].

Demographic differences

The literature reviewed also found demographic differences in levels of public support for administrative data research.

Younger age groups, for example, have in some instances been found to be more supportive of data sharing for research than older age groups [21, 28], though Stockdale *et al.*'s review [22] found evidence of both younger and older age groups being in favour of data sharing.

Aitken *et al.* [21] found participants not in full-time employment were more concerned with regulation measures and the type of data being linked than those in full-time employment. Those working full-time were more concerned with the purpose of data linking, who the researchers were and profit management. The Wellcome Trust [28] reported participants from socio-economic group C2DE (those in skilled, semi-skilled and unskilled manual jobs or on low or no income[1] felt more powerless to deal with the consequences of a data breach than those from socio-economic group ABC1 (those in managerial, administrative and professional, and supervisory and clerical jobs). Participants from group ABC1 were found to be more likely to view health data research as socially beneficial.

These demographic differences suggest some areas of research may be more acceptable than others to the specific groups whose lives they aim to benefit, on the basis of their demographic characteristics. It is therefore important to involve the communities most relevant to specific areas of research in public engagement activities, so the views of those most affected by the work are sufficiently understood.

Privacy and security

Safeguards to protect the privacy of data subjects and prevent data from being misused have also been identified by previous literature as key to public support for the sharing, linking and use of administrative data. The concerns identified can be broken down into three main areas: de-identification and anonymisation; data access and security; and governance and regulation.

De-identification and anonymisation

De-identification or anonymisation appears to be the minimum standard expected for the use of administrative data in research to be acceptable. Across the literature reviewed, study participants were found to be significantly more comfortable with their data being collected, stored and used when anonymised [15–17, 26, 27, 29]. Most participants of the 'Dialogue on Data' [15] – though not all – no longer considered de-identified data as 'personal' and had no concerns around the use of such data.

For the 85% of respondents of the 2015 NILT Survey who agreed data should be used where there is a benefit to society, this was based on an assurance that data would be anonymised [23]. Meanwhile, Ipsos MORI's consultation [24] found 62% of respondents would be 'certain or more likely' to provide their health information if there were assurances of confidentiality. Oswald [27] found less than 40% of respondents were comfortable with data sharing, even when anonymised, though this was specific to medical and locational data.

---

[1]Based on the National Readership Survey Social Grade: http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/.

Participants of Wellcome's attitudinal work [28, p.3.] had a strong sense of personal health data as *"confidential, private and sensitive"*, and not to be shared outside of *"secure, authorised bodies such as the NHS"*. Population-level (de-identified) data, however, was regarded as anonymous, and to be collected for the common good.

Participants of several studies raised concerns about whether it may be possible to re-identify individuals if linked data, for example, included information that was unusual and might only apply to a small number of people [15, 19, 20, 22]. Nevertheless, in most of the studies reviewed, respondents were largely supportive of data sharing when de-identification or anonymisation was guaranteed.

Data access and security

Study participants have expressed concern about data being leaked, lost, stolen or subject to unauthorised access and used against the public interest – whether de-identified or not – with additional safeguards to protect data therefore being considered critical [15, 22, 24, 28, 29].

In the context of ADRN, participants of the 'Dialogue on Data' [15] were reassured on learning of the restrictions on access to data, and were strongly in favour of secure physical settings and concerned about remote access to a secure environment. The authors stress, however, that the concept of remote access may not have been consistently explained across the workshops. They found most participants did not fully understand that data would not leave the physical setting when made accessible to researchers via a remote connection, and stress further work on how best to explain the concept is needed. Meanwhile, those who generally thought de-identified data is very low risk were more comfortable with remote access if protections were in place.

Participants of the Dialogue also felt reassured there were no plans for a so-called 'super database' under ADRN, containing multiple linked datasets. However, this appears to have been a spontaneous consideration of participants, and the authors do not explain what such a database was understood to be. The response of one participant suggests it was conceived as a service offering open access to data, rather than to only approved researchers: *"Everyone's information is going to be centralised. How can they guarantee everyone's motives?"* [15, p.30].

Stockdale *et al.* [22] found participants were concerned that sharing their electronic health records (EHRs) may lead to them being subject to unauthorised access and used to their disadvantage, while Wellcome [28] found the same for the sharing of personal data more generally. Amongst participants of Davidson *et al.*'s 2012 [19] and 2013 [26] consultations exploring attitudes towards cross-sectoral data sharing, these concerns were echoed, with participants of the 2012 study being concerned that data linking would increase the likelihood of security breaches as a large amount of information could be obtained at once.

Governance and regulation

In addition to the physical security of data, the literature reviewed identified a preference for protections in the form of governance and ethical frameworks to regulate data use.

In their reviews of previous literature, both Aitken *et al.* [16] and Stockdale *et al.* [22] identified an increase in public acceptance after study participants were informed about governance mechanisms. Davidson *et al.*'s 2012 work [19] identified concerns about who would oversee the operation of data sharing frameworks and where accountability would lie if linked data were lost or stolen.

Ipsos MORI [24] and Davidson *et al.* [26] both identified a preference for an independent organisation to act as a 'buffer' between researchers and the public to prevent the misuse of information. Participants of the 'Dialogue on Data' [15] felt reassured ADRN would provide a systematic way to regulate administrative data linking. In a series of focus groups conducted by Aitken *et al.* to explore public attitudes towards the use of health data [18], however, participants expressed concern that committees of oversight bodies would by default operate in favour of data sharing.

Ultimately, for administrative data research to be acceptable to the public, a myriad of safeguards are needed to protect the confidentiality of data subjects and limit the potential for data misuse.

Trust and transparency

The literature suggests individuals and institutions accessing data must be trusted to keep it secure and use it appropriately. Meanwhile, it also indicates that the specifics of projects using administrative data affect the level of public support, and transparency is therefore key to allowing the public to remain informed about how their data is used in any given context.

Trust

The literature reviewed found clear differences in the levels of trust attributed to different types of organisation, with the reasons given providing indications as to how an institution or individual might build trust.

Commercial organisations have been found to receive lower levels of trust than public bodies in general. Work by NatCen exploring Health Survey for England (HSE) participants' attitudes to data linkage [20] found government collecting data in the form of the Census, and health data collected by the NHS, was considered important for future planning. However, it was felt commercial companies would only want to access data for commercial gain. Participants of Davidson *et al.*'s 2013 workshops [26] expressed that who was accessing data and for what purpose is of greater concern than the type of data being accessed. They demonstrated widespread acceptance of public bodies accessing anonymised data from other public sector organisations for research, driven by a perception that these organisations are dedicated to delivering public benefits and safeguarding data.

Participants of the 'Dialogue on Data' [15], however, were worried about government data getting into the hands of commercial companies due to low trust in government in general. For those who were more trusting of government, its use of data was considered benign and in the public interest. The Wellcome Trust's attitudinal work [28] found some cynicism in relation to the government linking data and fears about government 'taking something away' from people.

Furthermore, not all public bodies receive the same levels of trust, with the literature identifying greater public trust in the NHS to keep information secure than in other public bodies [19, 23, 24, 27, 29]. Davidson *et al.*'s 2012 study [19] found this to rest upon a perception that health professionals serve to help the public and abide by a moral code of conduct, supposedly more so than other public workers.

Some study participants have also identified public benefits of private companies having access to data in certain circumstances; in such cases, there has been a preference for greater safeguards and controls than might be expected for public bodies [18, 25]. Aitken *et al.* [18] found that, although some organisations are trusted more than others, this does not mean access to data by these groups is automatically supported, and vice versa.

Participants of the 'Dialogue on Data' [15] felt researchers who gain access to linked data should be unbiased and qualified, while those working for private companies should not have access. Participants of Aitken *et al.*'s discrete choice experiment [21] felt most comfortable with university or government researchers and NHS staff accessing data, while Aitken *et al.*'s focus groups [18] revealed a feeling that academic researchers were less likely to be motivated by profit than other researchers.

These findings suggest that, to develop and maintain trust, an individual or organisation must demonstrate dedication to the public interest and safeguarding data. However, the findings also show that trust is not straightforward; in some cases, lower trust may simply mean a need for greater data protections and stronger assurances of public interest.

Transparency

Study participants have expressed a desire for greater transparency in general around how administrative data is held and used, with efficient communications around data use being seen to have a direct impact upon public acceptability [16, 17, 20, 22, 24, 29].

Participants of research by Tully *et al.* investigating public perceptions of appropriate uses of health data [25] felt the public benefit of data use by commercial organisations must be made explicit. Participants of Aitken *et al.*'s focus groups [18] expressed concern that a lack of openness may be a deliberate effort to withhold information from the public, with the authors finding that transparency plays an important role in levels of trust. Participants of NatCen's study [20] who had previously disagreed to have their HSE data linked to other forms of data recalled the main reason being a lack of comprehensive information about how data might be linked.

These findings suggest transparency has a direct impact on public support for data sharing initiatives. Furthermore, they suggest that if those handling and using data are transparent about its use, they are likely to receive greater levels of trust.

Striking the balance

None of the studies reviewed have identified any sole assurance that is enough alone to secure support amongst the UK public for research using public sector administrative data. Rather, all have highlighted that support cannot be guaranteed without fulfilment of all three core conditions: public interest, privacy and security, and trust and transparency.

In the 'Dialogue on Data' [15], tangible social value did not sit in isolation as a condition of public support; de-identification, data security and denying access to businesses wanting to use the data for profit were also considered necessary. The 2015 NILT Survey [23] found public support for data sharing to rest upon three pillars: trust in organisations, data protection measures and public benefit. If any are reduced or removed, public support falls. Stockdale et al.'s review of previous literature [22] found that, while there was a general willingness to share patient data for research in the public interest, this seldom led to unconditional support and rested upon data security and the motivations for using the data.

However, public support is not straightforward, and the literature shows the specifics of any given project have an impact on public expectations of the required standard of each core condition [15–17, 24, 27, 29]. Study participants have expressed they do not expect even the highest levels of data protection to be entirely foolproof, but are comfortable if the risk to privacy and potential for misuse are outweighed by the potential benefits in each case [15–17].

Work by the Royal Statistical Society exploring public attitudes towards data linking and privacy [29, p.3.] found 35% of respondents disagreed that: *"Once my data has been anonymised and there is no way I can be identified, I'm not really bothered how it is used"*, showing a notable proportion continued to care about how their data was used even when anonymised, and that the specifics of a programme of work may affect their views. Research by ONS between 2009 and 2013 [17] similarly found public views towards the use of data for research to differ according to who is using the data and for what purpose.

Ipsos MORI [24] identified a tension between the greater good and individual privacy. 69% of participants said they were likely to allow personal health information to be used for research – suggesting the specifics of the research are important – compared to only 14% who were certain to. Oswald *et al.* [27, p.270] found the *"benefits-versus-costs problem"* to be significant, and the more tangible the benefit, the greater the comfort in data sharing.

These findings show public support for research using administrative data is complex. They indicate that some projects may require the assurance of greater safeguards than others if aspects of their approach are considered less robust, even once a certain standard of each necessary condition is achieved. The potential public benefits must ultimately outweigh the risks to privacy and the possibility of misuse, and an appropriate balance of all three core conditions outlined above must be struck to achieve this.

Approaches to public engagement

The second objective of this paper is to be a source of advice on approaches to public engagement for organisations and researchers working with administrative data. In addition to exploring the attitudes and sentiments of the public towards administrative data research as discussed above, the literature has given important indications as to the type of engagement the public expect to have with it.

Aitken *et al.*'s focus groups [18, p.719] found preference for an *"open exchange of information and greater equity in the science-public relationship"*, in which public engagement is an indicator of the trustworthiness of data users rather than a way in which to build trust. The authors argue transparency must involve open communication of uncensored information, but also that trustworthiness is more likely to be achieved if engagement involves open dialogue in which public concerns can be responded to.

Davidson *et al.* [26] found unanimous agreement that the public should be involved in decision-making in relation to data sharing amongst participants of their 2013 workshops. Meanwhile, the Dialogue on Data [15] found mixed views amongst participants, with some expressing the public should be actively engaged in the operation of the ADRN, and others feeling this was unnecessary. Nevertheless, most participants of the Dialogue did feel there should be some place for members of the public to be involved in the running of the ADRN centres.

Rempela *et al.* [30], in light of the findings of a literature review of public engagement in new technologies, stress that transparency alone is not enough, and argue that data science initiatives should involve the public in technological development. This participatory approach (reflecting the 'public participation' element of public engagement) – is more effective at having a meaningful impact than a one-way communications-based approach (the 'public communication' element), say the authors. They stress that, while previous widescale consultations such as the 'Dialogue on Data' represent a step towards better understanding public views, they *"do not equate nor reflect public influence"* [30, p.575]. They also suggest identifying subsets of the public with whom it is more relevant to engage is important.

These findings suggest public engagement should move beyond public communication alone and seek to actively involve the public in decision-making processes associated with data use via public participation. Public engagement work should focus on gaining public input – via open and meaningful dialogue with relevant publics – in the development of administrative data research.

Conclusions and limitations

This review finds the UK public are broadly in favour of research using public sector administrative data, as long as three core conditions are fulfilled: public interest, privacy and security, and trust and transparency. Ultimately, an appropriate balance must be struck to ensure the proposed benefit outweighs the potential risk, and this is dependent upon the specifics of any given project, including: the data being used; the questions being asked; the protections in place; and the institutions or individuals accessing data. These attitudes have been held over time, with the studies reviewed covering over a decade of recent research.

It is important to note that the nature of research is such that it is not always possible to know if it will ultimately prove beneficial. The findings are not known at the start, and all that can be aimed for is intended benefit. Nevertheless, initiatives making use of public sector administrative data should aim to meet these three core conditions to ensure that their work is acceptable to the public.

In addition, this review finds that most relevant previous work within the UK – with the exception of that related to the use of healthcare data – has focused on capturing the attitudes of the general public towards the sharing, linking and use of data as a general principle, rather than in relation to particular data types or areas of research. The review also identified little work focused on capturing the views of specific demographics and communities in relation to research relevant to them specifically. A thorough understanding of the interests and concerns of those with lived experience of an issue – whether this is in relation to crime and justice, inequality and social inclusion, or another subject area – would be extremely valuable to ensuring research has the greatest positive impact possible.

Due to the volume of existing literature on general attitudes towards the use of data more broadly, and the consistent findings it has had, it is appropriate to move beyond general consultation on the acceptability of using public sector administrative data for research. Initiatives making use of administrative data should aim to focus on the specific issues being investigated by seeking the input of relevant demographics and communities. This is to ensure both that their interests are sufficiently considered in the development of research which affects them, and that the favoured participatory model of engagement – as indicated by the literature – is fulfilled. Nevertheless, it is important to continue to monitor any changes to broader public attitudes and adapt approaches if necessary.

Limitations

Although the most relevant method for this review, the unsystematic search method used to identify relevant literature is limited, as it leaves the findings more vulnerable to biases in relation to the literature selected, as well as potentially making the findings more difficult to replicate.

Acknowledgments

Conflicts of interest

The author is Communications & Engagement Manager for the ADR UK (Administrative Data Research UK) Strategic Hub, based within the Economic & Social Research Council (ESRC). ADR UK is UK-wide partnership funded by the ESRC with a mission is to enable secure access to UK administrative data for research in the public interest.

Ethics statement

Due to the nature of this work as a review of existing published studies, ethical approval was not required for this article.

References

1. Administrative Data Research UK. What is administrative data? (accessed 14/08/2020).

2. Digital Economy Act 2017. Disclosure of information for research purposes. Digital Economy Act 2017 part 5: Chapter 5, Section 64 (accessed 27/04/2020).

3. Administrative Data Research UK. Our Mission (accessed 18/11/2020).

4. Centre for Data Ethics and Innovation. Addressing trust in public sector data use. 2020 July (accessed 31/10/2020).

5. NHS England. NHS England announces new technical guidance to improve patient care. 2013 May (accessed 31/10/2020).

6. Mundasad S. NHS data-sharing project scrapped. BBC News. 2016 July (accessed 31/10/2020).

7. Information Commissioner's Office. Anonymisation: Managing data protection risk code of practice. 2012.

8. Stokes P. The 'Five Safes' – Data Privacy at ONS. Office for National Statistics. 2017 Jan (accessed 17/04/20).

9. Digital Economy Act 2017. Research Codes of Practice and Accreditation Criteria. Digital Economy Act 2017 part 5: Codes of Practice. Updated 2020 Feb (accessed 27/04/20).

10. Information Commissioner's Office. What is personal data? Guide to the General Data Protection Regulation (GDPR) (accessed 31/07/20).

11. Rowe G, Frewer LJ. A typology of public engagement mechanisms. Science, Technology & Human Values. 2005 April;30(2):251-290. https://doi.org/10.1177/0162243904271724.

12. Information Commissioner's Office. Principle (a): Lawfulness, fairness and transparency. Guide to the General Data Protection Regulation (GDPR) (accessed 17/04/20).

13. Grant M J, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information and Libraries Journal. 2009;26:91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x.

14. Economic and Social Research Council. The use of UK administrative data. 2020 September.

15. Cameron D, Pope S, Clemence M. Dialogue on Data: Exploring the public's views on using administrative data for research purposes. Ipsos MORI Social Research Institute. 2014.

16. Aitken M, de St. Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Medical Ethics. 2016;17(73). https://doi.org/10.1186/s12910-016-0153-x.

17. Office for National Statistics. The Census and Future Provision of Population Statistics in England and Wales: Public attitudes to the use of personal data for official statistics. 2014 March.

18. Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. Science and Public Policy. 2016 May 11;43(5):713-723. https://doi.org/10.1093/scipol/scv075.

19. Davidson S, McLean C, Cunningham-Burley S, Pagliari C. Public acceptability of cross sectoral data linkage: Deliberative research findings. Scottish Government Social Research. 2012 Aug.

20. Davies M, Jones H, Conolly A. Public Attitudes to Data Linkage: A report prepared for University College London by NatCen Social Research. NatCen Social Research. 2018 March.

21. Aitken M, McAteer G, Davidson S, Frostick C, Cunningham-Burley S. Public Preferences regarding Data Linkage for Health Research: A Discrete Choice Experiment. International Journal of Population Data Science. 2018;3(11). https://doi.org/10.23889/ijpds.v3i1.429.

22. Stockdale J, Cassell J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland (Revised). Wellcome Open Research. 2019 Jan;3(6). https://doi.org/10.12688/wellcomeopenres.13531.2.

23. Robinson G, Dolk H, Dowds L, Given J, Kane F, Nelson E. Public Attitudes to Data Sharing in Northern Ireland: Findings from the Northern Ireland Life and Times Survey 2015. Ulster University. 2018 Feb.

24. Ipsos MORI. The Use of Personal Health Information in Medical Research: General Public Consultation. Medical Research Council. 2007.

25. Tully MP, Hassan T, Oswald M, Ainsworth J. Commercial use of health data – A public "trial" by citizens' jury. Wiley: Learning Health Systems. 2019;3. https://doi.org/10.1002/lrh2.10200.

26. Davidson S, McLean C, Treanor S, Aitken M, Cunningham-Burley S, Laurie G, Sethi N, Pagliari C. Public acceptability of data sharing between the public, private and third sectors for research purposes. Scottish Government Social Research. 2013 Oct.

27. Oswald M. Share and share alike? An examination of trust, anonymisation and data sharing with particular reference to an exploratory research project investigating attitudes to sharing personal data with the public sector. SCRIPTed. 2014 Dec;11(3):245-272. https://doi.org/10.2966/scrip.110314.245.

28. Wellcome Trust. Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data. London: Wellcome Trust. 2013 July.

29. Royal Statistical Society. Royal Statistical Society research on trust in data and attitudes toward data use/data sharing. 2014.

30. Rempela ES, Barnett J, Durrant H. Public engagement with UK government data science: Propositions from a literature review of public engagement on new technologies. Government Information Quarterly. 2018 Oct;35(4):569-578. https://doi.org/10.1016/j.giq.2018.08.002.