

## Advancing cross-sectoral data linkage to understand and address the health impacts of social exclusion: Challenges and potential solutions

Lindsay A. Pearce<sup>1,2,\*</sup>, Rohan Borschmann<sup>1,2,3,4,5</sup>, Jesse T. Young<sup>3,6,7</sup>, and Stuart A. Kinner<sup>1,2,3,8</sup>

### Submission History

Submitted:	26/09/2022
Accepted:	17/04/2023
Published:	07/06/2023

<sup>1</sup>School of Population Health, Curtin University, Perth, Western Australia, Australia

<sup>2</sup>Justice Health Group, Centre for Adolescent Health, Murdoch Children's Research Institute, Melbourne, Victoria, Australia

<sup>3</sup>Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia

<sup>4</sup>Department of Psychiatry, University of Oxford, Oxford, UK

<sup>5</sup>Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Victoria, Australia

<sup>6</sup>School of Population and Global Health, The University of Western Australia, Perth, Western Australia, Australia

<sup>7</sup>National Drug Research Institute, Curtin University, Perth, Western Australia, Australia

<sup>8</sup>Griffith Criminology Institute, Griffith University, Brisbane, Queensland, Australia

### Abstract

The use of administrative health data for research, monitoring, and quality improvement has proliferated in recent decades, leading to improvements in health across many disease areas and across the life course. However, not all populations are equally visible in administrative health data, and those that are less visible may be excluded from the benefits of associated research.

Socially excluded populations – including the homeless, people with substance dependence, people involved in sex work, migrants or asylum seekers, and people with a history of incarceration – are typically characterised by health inequity. Yet people who experience social exclusion are often invisible within routinely collected administrative health data because information on their markers of social exclusion are not routinely recorded by healthcare providers. These circumstances make it difficult to understand the often complex health needs of socially excluded populations, evaluate and improve the quality of health services that they interact with, provide more accessible and appropriate health services, and develop effective and integrated responses to reduce health inequity.

In this commentary we discuss how linking data from multiple sectors with administrative health data, often called *cross-sectoral data linkage*, is a key method for systematically identifying socially excluded populations in administrative health data and addressing other issues related to data quality and representativeness. We discuss how cross-sectoral data linkage can improve the representation of socially excluded populations in research, monitoring, and quality improvement initiatives, which can in turn inform coordinated responses across multiple sectors of service delivery. Finally, we articulate key challenges and potential solutions for advancing the use of cross-sectoral data linkage to improve the health of socially excluded populations, using international examples.

### Keywords

data linkage; cross-sectoral; social exclusion; marginalised populations; health equity; social determinants of health

\*Corresponding Author:

Email Address: [Lindsay.Pearce@curtin.edu.au](mailto:Lindsay.Pearce@curtin.edu.au) (Lindsay A. Pearce)



## Article

Administrative health data contain routinely collected information on individual health service encounters such as hospitalisations, emergency department presentations, primary care contacts, pharmaceutical dispensing, and psychiatric services [1]. These data provide a powerful basis to generate information on the health of populations, permitting a detailed understanding of health disparities, the resources required to improve them, and the consequences of inaction [2–4]. This information can facilitate evidence-informed health system change and can be used to hold governments accountable to addressing the health needs of populations [2, 5]. Advances in electronic data collection, storage, and privacy have led to increases in research, monitoring, and quality improvement using administrative health data, driving improvements in health across many disease areas and across the life course [6, 7].

### What gets measured gets done

However, not all populations are equally visible in administrative health data, and those that are less visible may not benefit equally from research using these data. Populations that experience social exclusion – such as the homeless, people with substance dependence, people involved in sex work, migrants or asylum seekers, and people with a history of incarceration – are often invisible within routinely collected administrative health data [8, 9]. This is because these individuals' markers of social exclusion (i.e., housing status, substance use disorder, engagement in sex work, legal or immigration status, or incarceration) are typically not routinely recorded in administrative health data, or are recorded inaccurately or inconsistently [9, 10]. Administrative data that capture service contact that is a marker of disadvantage (e.g., housing or homelessness services, criminal justice system) are rarely routinely linked with health data. At a population level, this means that it is usually impossible to reliably identify individuals who may be experiencing social exclusion using administrative health data alone.

To date, poor availability and quality of information on disadvantage associated with social exclusion in administrative health data have limited our understanding of the healthcare needs of people who experience social exclusion – and consequently, what needs to be done to address those needs [5, 11, 12]. While administrative health data are good at assessing health equity in broad terms, such as by using area-based measures of deprivation (e.g., postcode to approximate socioeconomic status), these measures are unable to identify specific socially excluded populations with distinct needs and characteristics [9]. This unfortunate reality illustrates the maxim that 'what gets measured gets done'.

### Social exclusion and health inequity

People who are socially excluded often experience profound health and social inequities relative to other members of the population [9, 13–15]. They are often characterised by multiple and overlapping experiences of marginalisation including adverse childhood experiences, other adverse life events, trauma, poverty, racism, and discrimination [10, 13].

These and other barriers to social *inclusion*, such as language differences, restrictive requirements for service use (e.g., abstinence from drug use or absence of a criminal record), legal or immigration status, difficulties maintaining personal hygiene, and lack of identification documents, compound health inequities by impeding engagement with primary and preventive health and social services [8, 13]. Socially excluded populations tend to experience higher levels of morbidity, which impose a disproportionate burden on emergency and hospital care [8, 9, 13, 16–19].

*Inclusion health* is the broad term used to describe the service, research, and policy agenda aimed at addressing health and social inequities experienced by populations vulnerable to social exclusion [16]. The 2018 Lancet Inclusion Health Series [9, 13, 16] raised awareness of the health inequity associated with social exclusion, and the need for research to understand the often complex health needs and trajectories of socially excluded populations through typically poorly integrated health, welfare, and criminal justice systems. In that Series, Aldridge and colleagues [9] conducted a systematic review and meta-analysis of studies examining morbidity and mortality among four populations experiencing considerable social exclusion: the homeless; people with substance dependence; people involved in sex work; and people with a history of incarceration. Extreme inequities in the prevalence of disease were observed across a wide range of health conditions including infections, mental illness, cardiovascular disease, and respiratory disease. Meta-analyses revealed substantially elevated all-cause mortality compared to the general population for both men (standardised mortality ratio (SMR)=7.88; 95% confidence interval (CI) 7.03–8.74) and women (SMR=11.86; 95% CI 10.42–13.30). A subsequent systematic review and meta-analysis [15] found that all-cause mortality was highest among people with multiple markers of social exclusion compared to those with only one marker (hazard ratio (HR)=1.57, 95% CI 1.38–1.77). These findings support the long-held understanding that social exclusion is an important social determinant of health [20].

### Advancing evidence and action through cross-sectoral data linkage

Despite compelling evidence that social exclusion is health-depleting, efforts to reduce these health inequities have been both inadequate and ineffective. Experiences of social exclusion and their health impacts are deeply rooted in social disadvantage, which can only be addressed through long-term systemic change at multiple levels of society. However, we are permitting inaction by not leveraging the full potential of administrative data to better understand and address health inequity experienced by socially excluded populations. This has resulted in missed opportunities to generate evidence on the complex interplay between social exclusion and health, and to use this information as a basis for evidence-informed, targeted responses.

There is growing recognition that cross-sectoral data linkage is a key method to overcome the limitations of using administrative health data alone to understand and improve the health inequities experienced by socially excluded

populations [9, 10, 13, 21]. Cross-sectoral data linkage integrates data from health and 'non-health' sectors, such as housing, education, criminal justice systems, employment, and social services. Cross-sectoral data linkage also refers to linking data from mainstream health services (e.g., primary care, emergency department, and hospital data), specialist services that work with socially excluded populations (e.g., substance use treatment, assertive community outreach, harm reduction, or homeless services), and vital statistics [9].

## Benefits of cross-sectoral data linkage for inclusion health

There are several benefits of cross-sectoral data linkage for understanding and improving the health of socially excluded populations, some of which are described below.

### Benefit 1. Systematic identification of socially excluded populations

By linking health records with records of contact with other sectors and services that indicate probable disadvantage (e.g., housing or homelessness services, criminal justice system) we can identify a sampling frame of individuals at risk of social exclusion to apply to administrative health data. This offers an improvement over the status quo, in which socially excluded populations are often unidentifiable in administrative health data, precluding targeted investigation into their health needs. Cross-sectoral data linkage has been used to systematically identify populations experiencing incarceration [17, 22–24] and homelessness [18, 19, 25] in health administrative data. Similar applications of cross-sectoral data linkage would be helpful to identify other difficult-to-reach populations, such as migrants, asylum seekers, and sex workers, who remain underrepresented in health research [15, 26, 27].

Data linkage between multiple sectors or services can also allow for the systematic identification of people experiencing multiple forms of social exclusion, which is likely to have an additive or synergistic effect on morbidity, mortality, and quality of life [15]. For example, in Scotland, Tweed et al [10] linked national health service records with individual-level data from homelessness services, opioid treatment records, and a psychosis care register to examine the health of people affected by the intersection of homelessness, drug use, and serious mental illness. Two Danish studies [18, 19] linked national homelessness and substance use treatment registers, respectively, with psychiatric and vital statistics data to systematically identify these populations in administrative health datasets and examine their health and mortality outcomes.

### Benefit 2. Improved ascertainment of social exclusion status

When information regarding markers of disadvantage or social exclusion is routinely collected in health records, it is often of poor quality (i.e., inaccurately or inconsistently reported). This can lead to inaccurate conclusions regarding the burden of morbidity and mortality experienced by populations characterised by disadvantage or social exclusion; biased estimates of prevalence, incidence, and association; and (to the

extent that they are evidence-informed) potentially misguided policy and practice responses.

Cross-sectoral data linkage can markedly improve ascertainment of markers of disadvantage and social exclusion status by permitting the triangulation of data from multiple data sources. One example is ascertainment of Indigenous identification in administrative data, which is often inconsistently reported [5, 11]. While Indigeneity does not in itself indicate social exclusion, many Indigenous people worldwide experience profound health and social inequity due to the ongoing impacts of colonisation and systemic racism [28]. A qualitative synthesis of Australian studies [5] concluded that ascertainment of Aboriginal and Torres Strait Islander status improved when using multiple datasets, with measurable implications for morbidity and mortality estimates. One study found that under-ascertainment of Aboriginal and Torres Strait Islander status led to over-estimation of improvements in life expectancy, and in the prevalence of stigma-associated conditions such as sexually transmitted and blood-borne infections [5]. Another recent Australian study [12] using national death data identified overestimations in life expectancy of 2.3 and 2.1 years for Aboriginal and Torres Strait Islander males and females, respectively, in the absence of additional data linkage to enhance identification of Aboriginal and Torres Strait Islander status. Triangulation of data from multiple sources could be particularly useful for identifying populations that are less likely to disclose information on disadvantage or social exclusion due to stigma, discrimination, or legal consequences [29].

### Benefit 3. Improved estimation of service use and health outcomes

Cross-sectoral data linkage can improve ascertainment of health outcomes experienced by socially excluded populations. Due to the barriers to inclusion described above, people experiencing social exclusion often have low rates of engagement with primary and preventive healthcare services and rely heavily on acute and emergency care [8, 13]. This means that their health needs are often systematically underrepresented in primary and preventive healthcare records. Health and social services that are outside the scope of mainstream healthcare (e.g., government assistance, child protection, migrant services, assertive community outreach, harm reduction, homelessness services) often act as a safety net for individuals experiencing barriers or reluctance towards accessing primary and preventive health services. Inclusion of these broader data sources – often characterised as 'non-health' data – can capture other service use and population characteristics, thereby broadening our understanding of the health and social needs of socially excluded populations.

### Benefit 4. Informing cross-sectoral responses

The inclusion of 'non-health' data makes it possible to examine the association between social exclusion and health outcomes through a social-ecological framework that more broadly considers individual, social, and structural factors that affect health [30]. By integrating data from health and 'non-health' sectors, cross-sectoral data linkage can provide insights into the social and structural determinants of health.

This is important because socially excluded populations often experience disadvantage across numerous social and structural determinants that span multiple governmental sectors including housing, education, employment, criminal justice, physical environment, and social support [8]. Such complex health needs necessitate a diversity of data to optimally inform coordinated, person-centred responses across multiple sectors [15]. Research, monitoring, and quality improvement will therefore have greater impact if they can address both the health and non-health contributors to health inequity [13].

For example, in a Canadian study of 13,318 individuals who had experienced opioid overdose between 2014 and 2016, surveillance data were linked with national data on immigration, employment, social assistance, and police contact [31]. Considering these broader determinants of health revealed that markers of vulnerability to social exclusion were common in the year prior to first overdose, with 66% experiencing unemployment, 50% receiving social assistance, and 39% having at least one contact with police. These findings provide useful evidence to inform targeted preventive interventions across multiple sectors. Vertical approaches that consider health as distinct from the broader social issues faced by socially excluded populations will be less effective because they overlook other important contributing factors that influence health outcomes [9].

## Challenges and potential solutions for cross-sectoral data linkage

Despite widespread consensus on the need for cross-sectoral data linkage [9, 10, 13, 21, 32] and its demonstrated utility [10, 17–19, 31], cross-sectoral data linkage for inclusion health research, monitoring, and quality improvement remains underutilised globally. Several challenges have curtailed the use of cross-sectoral data linkage, but corresponding opportunities exist to advance this methodology in the context of inclusion health (see Table 1). We highlight several international examples.

### Challenge 1. Data availability

Much of the existing data linkage research on social exclusion centers on populations for which data tend to be more readily available and reliable. Systematic data collection typically hinges on the existence of sectors or services dedicated to supporting a given population (e.g., criminal justice systems, housing services) or collection of information on markers of disadvantage or social exclusion at service contact (e.g., substance use disorder recorded using standard coding conventions such as the International Classification of Diseases (ICD) [33]). Both reviews in the Lancet Inclusion Health Series primarily identified research on people with substance dependence, with fewer studies on homeless populations or people with a history of incarceration [9, 13]. Migrants, asylum seekers, and sex workers are examples of difficult-to-reach populations with extremely limited data capture, which contributes to their exclusion from health research [26]. Almost no studies were found on these groups in either Lancet review.

One way to improve data availability for hard-to-reach populations is by supporting the sectors and services

working with these populations to initiate or improve data collection and cross-sectoral linkage for research and quality improvement purposes. For example, low-barrier sexual health clinics often collect confidential data on sex worker status to inform clinical care and risk mitigation. These data have been used for epidemiological research [34] and have contributed to our understanding of the health burden associated with sex work. At a clinic or systems level, linkage of these de-identified data to other health and non-health datasets would provide critical insights into the causes and contributors of documented health inequities [27, 34, 35]. Similarly, records of migrants and asylum seeking populations are sometimes collected by governments [26] or through community-based health and 'social' support services [36].

Coalitions formed between researchers and service providers, premised on a shared interest in improving health outcomes and service delivery, can help support ad-hoc or repeated linkages of these data to other health and non-health datasets. Although resource intensive, the novelty of this work and importance for health equity may be attractive to funding agencies or governments. Where sufficient data are not yet collected, such coalitions could also work towards the establishment and quality assurance of a minimum required data collection within sectors and services targeted at difficult-to-reach populations that would allow for linkage of these identifiers to additional administrative data.

### Challenge 2. Service contact is an imperfect indicator of need

Health service use is an imperfect indicator of health need, particularly for socially excluded populations that often have low rates of access to primary and preventive healthcare, and relatively high rates of acute and emergency healthcare contact. Administrative health data therefore tend to under-ascertain the upstream health needs of socially excluded populations, capturing only symptomatic reasons for presentation. Furthermore, stigmatised behaviours and associated health outcomes that are more common among socially excluded populations, such as self-harm, suicidality, and non-fatal overdose, often do not lead to health service contact. They are therefore often under-reported in administrative health data. Socially excluded populations may also withhold information on these health experiences due to fear of stigmatisation, differential treatment, or legal consequences [37–41], an adaptive behavior documented among people with substance dependence [42–44] and people living with HIV [42–46]. Consequently, under-ascertainment of health service use and health outcomes can lead to inaccurate conclusions regarding health burden, missed opportunities for early intervention, and poorly informed policy and practice change.

Linking data from multiple sources can enable more accurate ascertainment of the health needs of socially excluded populations (e.g., linking data from low-barrier assertive community outreach services with other types of health and non-health data [47]). Triangulation of data collected from multiple sources, including combining administrative records with patient self-report, can also enhance ascertainment of health outcomes that are not well captured in administrative health data. One study

Table 1: Key challenges, consequences, and potential solutions for cross-sectoral data linkage

Challenges	Consequences	Potential solutions
Markers of SE are not routinely collected in administrative health data, or are of poor quality	SE populations are often invisible in administrative health data; existing health research largely focusses on populations defined by health diagnoses (e.g., substance dependence)	Linking administrative health data with data from 'non-health' sectors and/or specialist services that work with SE populations can systematically identify individuals experiencing SE in health data
Administrative health data do not contain information on the social determinants of health	Responses that are informed only by administrative health data may overlook important contributing factors that influence health inequity	Linking administrative health data to 'non-health' data (e.g., housing, education, employment, criminal justice system, social supports) to understand the role of the social determinants of health
SE status and associated health outcomes are often under-ascertained in administrative health data	Under-ascertainment can lead to inaccurate conclusions regarding health burden, and poorly informed policy and practice responses	Data triangulation and validation using multiple data sources can improve ascertainment of SE status and associated health outcomes
Data on hard-to-reach SE populations (e.g., migrants, asylum seekers, sex workers) are rarely collected or made available	These populations are under-represented in health research, resulting in extremely limited knowledge on their health needs and missed opportunities for evidence-informed responses to reduce health inequity	Supporting the sectors and services working with these populations to initiate or improve data collection and sharing for research and quality improvement purposes can improve their representation in health research
Service sectors and providers that work with SE populations often lack capacity for routine data collection and reporting	There is insufficient quantity and quality of data collected from these entities to include in cross-sectoral data linkage research	Forming coalitions between SE service providers, researchers, data custodians, and linkage administrators can address gaps by establishing basic data collection systems to permit linkage to administrative health data
Decentralised health and social service delivery results in piecemeal data collection	Datasets do not use consistent identifiers, requiring imperfect probabilistic linkage methods that can introduce bias into research and increase administrative burden during linkage. Data gaps in populations or service coverage may be present.	Careful consideration of the methods and quality of data linkage on research outputs, including the publication of validation studies, will expand the knowledge base on data linkage methods applied with SE populations. Working closely with data custodians to understand populations and service represented, and to identify additional datasets that can fill gaps in populations and services of interest, will build understanding of the impacts on internal and external validity.
Data custodians are reluctant to participate in data linkage and have research findings made publicly available irrespective of findings ("epistemophobia")	Cross-sectoral data linkage is unable to progress without the support of data custodians	Developing partnerships between researchers, data custodians, and linkage administrators that prioritise research benefit for SE populations, focus on the value of data linkage from a policy and practice perspective, and place decision making with data governance bodies that are pre-authorised to release data for agreed purposes can help facilitate cross-sectoral data linkage
Data custodians are reluctant to participate in data linkage research out of concern for privacy and confidentiality of data in their possession, particularly with sensitive or heavily anonymised data	Cross-sectoral data linkage is unable to progress without the support of data custodians	Reassurance of recent advancements around data safe haven (DSH) environments, Five Safes Framework, secure file transfer systems, and the separation principle can help reassure data custodians of the high degree of security and anonymity in cross-sectoral data linkage

Continued

Table 1: Continued

Challenges	Consequences	Potential solutions
Extensive ethical requirements and legislative barriers for data sharing and linkage across sectors and service providers are common	Cross-sectoral data linkage is administratively burdensome, time-consuming, and resource-intensive; these factors restrict research capacity and timely and evidence-informed responses to reduce health inequity	Forming centralised bodies to help coordinate complex data linkage processes, streamline protocols, organise connections between researchers and data custodians, and reduce administrative burden on researchers could improve linkage efficiency (e.g., PHRN, HDR UK)
There is insufficient workforce capacity and technical expertise to (a) conduct cross-sectoral data linkage and (b) analyse complex linked data	If possible to proceed at all, data linkage projects risk becoming highly inefficient and resource intensive, are prone to linkage error, and research outputs are prone to bias	Expanding workforce training opportunities and access to hands-on experience working with linked data in postgraduate education, financial investment in data linkage infrastructure to meet demand for services, and focusing on interdisciplinary teams will improve efficiency, accuracy, and capacity for cross-sectoral data linkage
Many countries do not have the infrastructure for high quality, publicly accessible, and centralised data collection and linkage	Issues of data quality, completeness, and availability restrict health research involving SE populations, particularly in LMICs	Continuing to invest in routine health information systems (e.g., DHIMS-2) can improve data quality and accessibility for health research
SE populations are frequently excluded from meaningful involvement in data collection and cross-sectoral data linkage research	Cross-sectoral data linkage risks reinforcing exclusion, and not benefitting from lived experience perspectives on data collection, evidence generation, and associated responses	Engaging with SE populations as early as possible can help to improve ascertainment of SE status, address possible concerns around misrepresentation, improve relevance of research findings, and inform policy responses

*Abbreviations:* SE: social exclusion or socially excluded; LMICs: low- and middle-income countries; PHRN: Population Health Research Network (Australia); HDR: Health Data Research (UK); DHIMS-2: District Health Information Management System 2.

of adults recently released from prisons in Australia found moderate to good agreement between administrative health records and patient self-report across primary care ( $\kappa = 0.69$ , positive predictive value (PPV) = 80.2%), emergency department ( $\kappa = 0.41$ , PPV = 64.0%) and hospital services ( $\kappa = 0.62$ , PPV = 76.0%), and for the use of prescribed medications ( $\kappa = 0.62$ , PPV = 68.3%) [48]. These findings complement existing research examining the validity of combining administrative and self-reported data to verify experiences of self-harm [49] and non-fatal drug overdose [50] among people released from prison.

### Challenge 3. Decentralised data collection

The siloed nature of service delivery provides additional challenges for cross-sectoral data linkage. The delivery of services by different levels of government or non-governmental sectors means that data collection is often decentralised. This makes it challenging to link data across sectors [3, 51]. Deterministic linkage using unique personal identifiers (e.g., personal health number) that are available across all datasets is simpler and more accurate. This is most common in centralised service delivery and data collection systems, such as in Scandinavia [52–54]. Decentralised data collection results in imperfect linkage because data must be linked based on the probability that two or more records match, using non-unique identifiers (e.g., name, sex, date of birth) [55]. Notably, some datasets will not have sufficient information to conduct

probabilistic linkages, precluding data linkage altogether [55]. While probabilistic linkage methods are widely used and can produce high quality linkages [56, 57], the impact on linkage error and bias in study findings can vary across datasets and can disproportionately impact socially excluded populations who may be less likely to have identification and whose identifiers may be more prone to misspelling or inaccuracy [57, 58]. It is therefore important that analysts carefully consider the quality of data linkage on research outputs, and seek out opportunities for validation studies to inform future research [59].

It is largely for this reason that much of the research using cross-sectoral data linkage to examine the health consequences of social exclusion has taken place in high-income countries with more advanced, centralised data collection systems and more liberal approaches to privacy legislation that permit the appropriate use of these data for research, monitoring, and quality improvement. For example, Sweden, Denmark, and Norway maintain numerous nation-wide registries that can be linked for research purposes using unique personal identifiers [52–54, 60, 61]. This has contributed to a favourable environment for routine linkage of both existing datasets and primary (e.g., survey) data [18, 19]. In Australia, the Australian Institute of Health and Welfare maintains over 150 national databases that include housing, homelessness, criminal justice, disability, alcohol and other drugs, mental health, healthcare, education, and vital statistics [61], many of which are available for research purposes and can be linked

using a National Linkage Map. In the UK, Administrative Data Research (ADR UK) facilitates access to public sector data from England, Northern Ireland, Scotland, Wales, and the Office for National Statistics for use in public interest projects led by accredited researchers. For example, the Welsh Secure Anonymised Information Linkage (SAIL) Databank (one component of ADR UK) includes over 80 datasets across health, justice, education, and child protection sectors [62]. In British Columbia (BC), Canada, Population Health BC provides researchers with linkage services and access to over 90 datasets across health, justice, education, housing, social care benefits, child welfare, employment, and immigration sectors [63].

Cross-sectoral data linkage is more complex in countries with marked decentralisation of service delivery, including countries with two-tier service delivery systems. In such settings, data are collected from numerous public and private health and social service systems that often lack consistent personal identifiers, resulting in piecemeal data collection and complex linkage processes that introduce avoidable linkage error [55]. In these environments, researchers should work closely with data custodians to understand populations and services represented within individual datasets, identify additional datasets that may help fill gaps in population or service coverage of interest, carefully consider impact on internal and external validity, and conduct validation studies to assess the impact of linkage methods and quality on research outputs.

The routine collection of electronic health and non-health data remains limited in many low- and middle-income countries (LMICs), and the use of these data for research, monitoring, and quality improvement purposes has been restricted due to concerns over their quality, completeness, and public availability [64]. A 2020 review of studies using routine health information in LMICs [64] provided evidence of increased uptake and investment in routine health information systems in recent decades. For example, the implementation of the District Health Information Management System 2 (DHIMS-2) in sub-Saharan Africa has allowed more efficient, integrated, and timely health data collection from local to national levels [65]. DHIMS-2 has permitted data integration from community-based health registers, population surveys, and meteorological departments for health system performance measurement and research [64–66]. It is conceivable that such systems could be used to integrate data from other non-health sectors as they evolve in LMICs. However, of a mere 132 studies identified in the LMIC review [64], none focussed on socially excluded populations as defined by the Lancet Inclusion Health Series [9]. These findings may reflect both the infancy of routine data collection in LMICs for inclusion health research and differing research priorities in LMICs, which were overwhelmingly focussed on malaria and maternal health. Nonetheless, over time, these routine health information systems may provide the foundation for future cross-sectoral data linkage involving socially excluded populations in these settings.

#### Challenge 4. Reluctance to share data

Despite exceptionally high standards of ethical clearance required for health research on vulnerable populations [67],

many data custodians remain reluctant to share data for research, monitoring, and quality improvement purposes. This is particularly common in sectors not typically involved in data linkage research [32]. Data custodians are responsible for maintaining the privacy and confidentiality of data in their possession. Thus, sharing these data require a high degree of trust that privacy and confidentiality will be maintained – particularly when they are focussed on vulnerable and socially excluded populations. Concerns around data security result in a heightened sense of data protection, particularly for those new to data linkage. The emergence of data safe haven (DSH) environments that monitor secure storage and use, establish robust data governance protocols, and prohibit the export of individual-level data are premised on the principles of data safety and trust [68]. Additional advancements including the Five Safes Framework for planning, assessing, and managing risks associated with data sharing and release [69]; secure transfer of encrypted data across data custodians and linkage teams [70]; linkage methods that preclude the need to share full identifiers across sectors [71, 72]; and application of the separation principle [73] – such that no one can view both personal identifying information (used for data linkage) and content data (e.g., clinical or service use information) at any point during linkage and analysis – should in theory address data custodian concerns around data security. Importantly, these advancements can build public trust in the management and protection of sensitive administrative data, which can in turn support data custodian decisions to participate in data linkage research [74–76]. However, it is currently unknown to what degree these advancements have impacted the willingness of data custodians to share data for health research purposes. This is an important area for future research.

Socially excluded populations are often over-represented in particularly sensitive data, such as notifiable disease registries (e.g., HIV, HCV) and other heavily anonymised health data (e.g., abortion records, police contacts). The highly sensitive nature of these data can result in an even greater sense of reluctance to share these data for health record linkage, additional administrative requirements that can reduce research capacity, and more conservative approaches to interpreting legislation that protects (but rarely precludes) their use [77]. This can inhibit research on issues disproportionately impacting socially excluded populations, despite the potential benefits for health equity. Nonetheless, several jurisdictions maintain population-level disease registries (for example, on HIV in Australia and Scotland) that can be linked to additional administrative datasets by data custodians or researchers [78, 79]. The advancements listed above may be of particular value when planning and undertaking data linkage using sensitive and anonymised data. However, linkage to sensitive data remains challenging. Researchers must not undervalue the importance of establishing strong partnerships with data custodians to understand data sharing concerns and legislative requirements, and to jointly identify linkage and analytic arrangements that address them [80, 81]. One option is storing linked data with the data custodian and arranging secure data access for external analysts.

A perennial challenge for researchers interested in conducting cross-sectoral data linkage is the reluctance of

many data custodians to participate in rigorous, independent research or quality improvement, and to have research findings made publicly available, whether or not they seem favourable to the custodian [82]. Furthermore, some 'non-health' sectors may form the view that health is not part of their remit, impeding cooperation for health data linkage because it is perceived to be irrelevant. This is often manifest in legislative and/or administrative barriers that explicitly or implicitly limit data access, linkage, or public reporting. Data custodians may also actively try to block data access because they do not want independent scrutiny, which in turn can compound health inequities among socially excluded populations. Overcoming this kind of "epistemophobia" requires the development of meaningful partnerships between researchers and data custodians that prioritise research benefit for the population(s) under study, communicate the value of cross-sectoral data linkage from a policy and practice perspective, and shift focus away from concerns over scrutiny and towards opportunities for evidence-informed service and health improvement [32, 82]. Decisions about data access may also be better placed in the hands of centralised data governance bodies that are pre-authorised to release data for agreed purposes based on public benefit, an established and transparent decision-making framework, and high ethical and data security standards. An example of this is the Confidential Advisory Group of the Health Research Authority in the UK [83].

Finally, engaging decision makers to understand their perspectives on the value of cross-sectoral data linkage for informing policy and practice responses within inclusion health [84], and sharing this information with data custodians, can help close the loop on data sharing hesitancy by focussing on real-world, cross-sectoral benefits.

### Challenge 5. Prohibitive administrative burden

Cross-sectoral data linkage is complex in theory and practice. It involves sharing and linking data from multiple sectors and jurisdictions, all of which may have their own unique requirements and protocols. Cross-sectoral data linkage is therefore prone to extensive requirements for ethical, data custodian, and legislative approval, contributing to its time-consuming and resource-intensive nature [3]. These factors alone substantially reduce capacity for cross-sectoral data linkage and inhibit evidence-informed action on emergent health concerns that disproportionately impact socially excluded populations (e.g., COVID-19 [85]). Centralised bodies dedicated to the coordination of cross-sectoral data linkage can help researchers navigate complex linkage processes, communicate research requirements across multiple data custodians and jurisdictions, and streamline duplicative administrative protocols. Importantly, a key purpose for such bodies must be to reduce often prohibitive administrative burden, rather than introducing an additional layer of bureaucracy.

For example, the Australian Population Health Research Network (PHRN) was established in 2009 as a national coordinating network to support researchers to access and link more than 300 available health and health-related data collections [86]. Australia has a federated system of government comprising six states, two territories, and a Commonwealth government, each with its own data

collections and governance. Thus, data linkage projects often rely on several jurisdictional data centres and require multiple approvals and data extractions depending on the number of data collections and jurisdictions involved [86]. A benefit of the PHRN is its national coordinating function: it can coordinate meetings between researchers and all relevant linkage centres and data custodians to jointly establish feasibility of research proposals, identify data access requirements, prepare linkage applications, and organise approval processes. The PHRN also provides publicly available education and training to researchers, data custodians, and regional data linkage units. However, cross-sectoral data linkage in Australia remains challenging and underutilised, due in part to its resource intensity [3, 87]. Despite substantial investment of AUD\$55 million in this national data linkage infrastructure between 2009 and 2019, and a 3.5-fold increase in the number of peer-reviewed publications involving 'health-only' data linkage during that period, only a small increase in the number of studies involving cross-sectoral data linkage has been observed [6]. Entities such as the PHRN, that have strong links to data custodians and linkage administrators, are well placed to advocate for streamlined administrative processes that improve efficiency of linkage processes. Simultaneously, advances such as the Australian National Mutual Acceptance (NMA) Scheme [88] that aim to streamline scientific and ethical review across data custodians and jurisdictions can help to reduce duplicative processes whilst maintaining integrity in ethical and data security protocols [32]. Similar work is being done in the UK via the Health Data Research (HDR UK) institute [89].

### Challenge 6. Limited workforce capacity

Insufficient workforce capacity, skill, and technical expertise contribute to the limited use of cross-sectoral data linkage [3, 32]. Cross-sectoral data linkage is complicated and requires a multi-disciplinary data linkage workforce that 1) is well-rounded in the fundamentals of health research including (at a minimum) basic statistics and epidemiology, 2) is familiar with the intricacies of administrative data collection and linkage across sectors, 3) has commensurate understanding of the social and structural determinants of health, and 4) is prepared to navigate the complex interplay of political sensitivities, epistemophobia, and siloing of within-sector mandates that can impede cross-sectoral data linkage [3, 7]. Furthermore, the complexity of linked administrative data demands skilled analysts that are proficient in both the theoretical and hands-on applications of social science, public health, health data science, biostatistics, and epidemiology. In recent years, there has been an increase in the number of university programs in these domains; in particular, programs relating to health data science. However, it remains unclear whether the supply of graduates with broad skills in these combined fields has kept pace with increased demand for data linkage and associated research. Workforce shortages in both linkage and analytic domains need to be addressed by continuing to expand workforce training opportunities and hands-on experience with cross-sectoral data in postgraduate education, increasing investment in data linkage infrastructure to meet demand for these services, and forming interdisciplinary teams to achieve the required diversity in knowledge



and skill [90]. For example, in Ontario, Canada, an innovative partnership was established between ICES (a health service research institute), computer scientists, and machine learning specialists to apply more complex analytic methods to the analysis of province-wide administrative data [91].

### Challenge 7. Including socially excluded populations

Cross-sectoral data linkage provides a unique opportunity for the inclusion of socially excluded populations as experts in their own experiences [92]. Yet, socially excluded populations are only sporadically engaged in cross-sectoral data linkage and associated research. Cross-sectoral data linkage, including the pursuit of improved data collection within sectors and services that predominantly serve socially excluded groups, is currently an underused opportunity to meaningfully engage socially excluded populations and reinforce their right to self-determination in health research. It is imperative that populations who are observed through cross-sectoral data linkage are meaningfully involved as early as possible, from advising on what data to collect and how to collect it from socially excluded groups at the point of service access, to planning research, identifying research questions and the data sources needed to answer them, data analysis, interpretation, knowledge dissemination, and decision making about how to respond to knowledge generated [93–96]. This is essential to ensure that information generated from linked data are contextualised by lived experience, and that socially excluded populations are not further excluded from evidence generation and decision making regarding their health and wellbeing. Engagement of socially excluded populations to inform and participate in data linkage research has been demonstrated [25] but is most often focused on engagement of First Nations communities [97]. In Scotland, the Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) is a patient advocacy panel that scrutinises applications requesting access to NHS data with a view to ensuring that linkage projects lead to tangible benefit for populations of interest and that risks are adequately addressed [98]. HDR UK involves patients and the public through a Public Advisory Board, workshops, opinion surveys, and public events to engage and involve underserved communities [89]. Similar applications may benefit inclusion health research and should not preclude the early engagement of populations of interest during development of research proposals. Ongoing evaluation and feedback is essential to ensure that such bodies are capable of understanding the basics of data linkage research, efficient, pragmatic, and do not lead to undue administrative barriers [99]. Better engagement with people experiencing social exclusion can also inform efforts to improve ascertainment of social exclusion status (e.g., by matching data collection with identities held by populations [100]); address concerns regarding (mis)representation in administrative data; educate patient populations on what data are collected and how it will be used to inform policy and practice; build inclusive coalitions between socially excluded populations, data custodians, and researchers; and develop a shared understanding of research benefit [98] – both within existing data collection systems and when developing new ones.

## Conclusion

In the absence of routine and reliable data collection on markers of social exclusion in administrative health data, innovative ways to measure the causes and consequences of social exclusion using data linkage are needed. Cross-sectoral data linkage is an effective methodology to generate the information required to understand and address the health inequities experienced by socially excluded populations. Presently, this potential is not being fully realised. However, there is growing global momentum to make cross-sectoral data linkage more accessible, with promising future impacts on inclusion health. Jurisdictions such as Australia, Canada, Scandinavia, and the UK continue to lead the way by making population-level, cross-sectoral data linkage more accessible to researchers, and identifying innovative ways to improve data availability, security, linkage efficiency, researcher-data custodian connections, workforce capacity, and lived experience engagement. Ongoing advancements in data collection and integration in LMICs will help bridge the digital divide and set the foundation for future non-health data collection and cross-sectoral data linkage research. Greater public discussion of challenges encountered, and solutions implemented, will be essential for advancing cross-sectoral data linkage and its possible benefits for the health of socially excluded populations.

## Funding declaration

This work was funded by the Australian National Health and Medical Research Council (NHMRC) award no. GNT1163689. JY receives salary and research support from a National Health and Medical Research Council Investigator Grant (GNT1178027). RB receives salary and research support from a National Health and Medical Research Council Emerging Leadership Investigator Grant (EL2; GNT2008073). The funding body had no additional role in the development of this work.

## Ethics statement

Ethics approval was not required.

## Conflicts of interest

None declared.

## References

1. Mazzali C, Duca P. Use of administrative data in healthcare research. *Intern Emerg Med*. 2015;10(4):517–24. <https://doi.org/10.1007/s11739-015-1213-9>
2. Young JT, Bellgrove MA, Arunogiri S. Assessment of attention-deficit hyperactivity disorder in people with substance use disorder: Another case of what gets measured gets done. *Australian & New Zealand Journal of Psychiatry*. 2021;00(0). <https://doi.org/10.1177/00048674211009607>

3. Palamuthusingam D, Johnson DW, Hawley C, Pascoe E, Fahim M. Health data linkage research in Australia remains challenging. *Internal Medicine Journal*. 2019;49(4):539–44. <https://doi.org/10.1111/imj.14244>
4. USAID, World Bank, World Health Organization. The roadmap for health measurement and accountability. 2015.
5. Thompson SC, Woods JA, Katzenellenbogen JM. The quality of indigenous identification in administrative health data in Australia: insights from studies using data linkage. *BMC Med Inform Decis Mak*. 2012;12:133. <https://doi.org/10.1186/1472-6947-12-133>
6. Young A, Flack F. Recent trends in the use of linked data in Australia. *Aust Health Rev*. 2018;42(5):584–90. <https://doi.org/10.1071/ah18014>
7. McGrail K, Moran R, O’Keefe C, Preen D, Quan H, Sanmartin C, et al. A Position Statement on Population Data Science: The science of data about people. *International Journal of Population Data Science*. 2018;3(1). <https://doi.org/10.23889/ijpds.v3i1.415>
8. United Nations. Leaving no one behind: The imperative of inclusive development. 2016.
9. Aldridge RW, Story A, Hwang SW, Nordentoft M, Luchenski SA, Hartwell G, et al. Morbidity and mortality in homeless individuals, prisoners, sex workers, and individuals with substance use disorders in high-income countries: a systematic review and meta-analysis. *The Lancet*. 2018;391(10117):241–50. [https://doi.org/10.1016/S0140-6736\(17\)31869-X](https://doi.org/10.1016/S0140-6736(17)31869-X)
10. Tweed E, Leyland A, Morrison D, Katikireddi SV. Using cross-sectoral data linkage to understand the health of people experiencing multiple exclusion. *European Journal of Public Health*. 2020;30(Supplement\_5) <https://doi.org/10.1093/eurpub/ckaa165.052>
11. Randall DA, Lujic S, Leyland AH, Jorm LR. Statistical methods to enhance reporting of Aboriginal Australians in routine hospital records using data linkage affect estimates of health disparities. *Aust N Z J Public Health*. 2013;37(5):442–9. <https://doi.org/10.1111/1753-6405.12114>
12. Australian Institute of Health and Welfare. Improving Indigenous identification in mortality estimates. 2019.
13. Luchenski S, Maguire N, Aldridge RW, Hayward A, Story A, Perri P, et al. What works in inclusion health: overview of effective interventions for marginalised and excluded populations. *Lancet*. 2018;391(10117):266–80. [https://doi.org/10.1016/s0140-6736\(17\)31959-1](https://doi.org/10.1016/s0140-6736(17)31959-1)
14. Fitzpatrick S, Bramley G, Johnsen S. Pathways into Multiple Exclusion Homelessness in Seven UK Cities. *Urban Studies*. 2013;50(1):148–68. <https://doi.org/10.1177/0042098012452329>
15. Tweed EJ, Thomson RM, Lewer D, Sumpter C, Kirolos A, Southworth PM, et al. Health of people experiencing co-occurring homelessness, imprisonment, substance use, sex work and/or severe mental illness in high-income countries: a systematic review and meta-analysis. *Journal of Epidemiology and Community Health*. 2021;75(10):1010. <https://doi.org/10.1136/jech-2020-215975>
16. Marmot M. Inclusion health: addressing the causes of the causes. *The Lancet*. 2018;391(10117):186–8. [https://doi.org/10.1016/S0140-6736\(17\)32848-9](https://doi.org/10.1016/S0140-6736(17)32848-9)
17. Graham L, Fischbacher CM, Stockton D, Fraser A, Fleming M, Greig K. Understanding extreme mortality among prisoners: a national cohort study in Scotland using data linkage. *Eur J Public Health*. 2015;25(5):879–85. <https://doi.org/10.1093/eurpub/cku252>
18. Arendt M, Munk-Jørgensen P, Sher L, Jensen SO. Mortality among individuals with cannabis, cocaine, amphetamine, MDMA, and opioid use disorders: a nationwide follow-up study of Danish substance users in treatment. *Drug Alcohol Depend*. 2011;114(2-3):134–9. <https://doi.org/10.1016/j.drugalcdep.2010.09.013>
19. Nielsen SF, Hjorthøj CR, Erlangsen A, Nordentoft M. Psychiatric disorders and mortality among people in homeless shelters in Denmark: a nationwide register-based cohort study. *Lancet*. 2011;377(9784):2205–14. [https://doi.org/10.1016/s0140-6736\(11\)60747-2](https://doi.org/10.1016/s0140-6736(11)60747-2)
20. World Health Organization Europe. Social determinants of health: The solid facts (2nd Edition). 2003. Available from: [https://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0005/98438/e81384.pdf](https://www.euro.who.int/__data/assets/pdf_file/0005/98438/e81384.pdf).
21. World Health Organization. Multisectoral action in developing information systems for health. 2019.
22. Kinner S. Health service utilisation and preventable mortality in justice-involved young people: A national, retrospective data linkage study. Proposal to NHMRC 2018 Project Grants Competition. 2018.
23. Kinner SA, Degenhardt L, Coffey C, Hearps S, Spittal M, Sawyer S, et al. Substance use and risk of death in young offenders: A prospective data linkage study. *Drug and Alcohol Review*. 2015;34:46–50. <https://doi.org/10.1111/dar.12179>
24. Gan WQ, Kinner SA, Nicholls TL, Xavier CG, Urbanoski K, Greiner L, et al. Risk of overdose-related death for people with a history of incarceration. *Addiction*. 2021;116(6):1460–71. <https://doi.org/10.1111/add.15293>
25. Luchenski S, Clint S, Aldridge R, Hayward A, Macquire N, Story S, et al. Involving people with lived experience of homelessness in electronic health records research. *International Journal of Population Data Science*. 2017;295 Proceedings of the IPDLN Conference (August 2016)(1) <https://doi.org/10.23889/ijpds.v1i1.315>

26. Hedrick K, Armstrong G, Coffey G, Borschmann R. Self-harm among asylum seekers in Australian onshore immigration detention: How incidence rates vary by held detention type. *BMC Public Health*. 2020;20(1):592. <https://doi.org/10.1186/s12889-020-08717-2>
27. Platt L, Grenfell P, Meiksin R, Elmes J, Sherman SG, Sanders T, et al. Associations between sex work laws and sex workers' health: A systematic review and meta-analysis of quantitative and qualitative studies. *PLoS Med*. 2018;15(12):e1002680. <https://doi.org/10.1371/journal.pmed.1002680>
28. Anderson I, Robson B, Connolly M, Al-Yaman F, Bjertness E, King A, et al. Indigenous and tribal peoples' health (The Lancet–Lowitja Institute Global Collaboration): A population study. *The Lancet*. 2016;388(10040):131–57. [https://doi.org/10.1016/S0140-6736\(16\)00345-7](https://doi.org/10.1016/S0140-6736(16)00345-7)
29. Lanau A, Matolcsi A. Prostitution and Sex Work, Who Counts? Mapping Local Data to Inform Policy and Service Provision. *Social Policy & Society*. 2022:1–15. <https://doi.org/10.1017/S1474746422000136>
30. Centres for Disease Control and Prevention. The social-ecological model: A framework for prevention 2022 [Available from: <https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html>]
31. Carrière G, Sanmartin C, Garner R. Understanding the socioeconomic profile of people who experienced opioid overdoses in British Columbia, 2014 to 2016. *Statistics Canada Health Reports*. 2021;32(2). <https://doi.org/10.25318/82-003-x202100200003-eng>
32. Canaway R, Boyle D, Manski-Nankervis JE, Bell J, Hocking J, Clarke K, et al. Gathering data for decisions: Best practice use of primary care electronic records for research. *Med J Aust*. 2019;210(6):S12–S6. <https://doi.org/10.5694/mja2.50026>
33. World Health Organization. International Classification of Diseases and Related Health Problems (ICD) 2023 [Available from: <https://www.who.int/standards/classifications/classification-of-diseases>]
34. McCann J, Crawford G, Hallett J. Sex Worker Health Outcomes in High-Income Countries of Varied Regulatory Environments: A Systematic Review. *International Journal of Environmental Research and Public Health*. 2021;18(8):3956. <https://doi.org/10.3390/ijerph18083956>
35. Muldoon KA. A systematic review of the clinical and social epidemiological research among sex workers in Uganda. *BMC Public Health*. 2015;15(1):1226. <https://doi.org/10.1186/s12889-015-2553-0>
36. Feldman R. Primary health care for refugees and asylum seekers: A review of the literature and a framework for services. *Public Health*. 2006;120(9):809–16. <https://doi.org/10.1016/j.puhe.2006.05.014>
37. Chaudoir SR, Quinn DM. Revealing concealable stigmatized identities: The impact of disclosure motivations and positive first disclosure experiences on fear of disclosure and well-being. *J Soc Issues*. 2010;66(3):570–84. <https://doi.org/10.1111/j.1540-4560.2010.01663.x>
38. Von Hippel C, Brener L. Specificity of Discrimination: Does It Matter From Whence It Comes? *Journal of Applied Social Psychology*. 2012;42(4):1029–42. <https://doi.org/10.1111/j.1559-1816.2011.00851.x>
39. Biancarelli DL, Biello KB, Childs E, Drainoni M, Salhaney P, Edeza A, et al. Strategies used by people who inject drugs to avoid stigma in healthcare settings. *Drug Alcohol Depend*. 2019;198:80–6. <https://doi.org/10.1016/j.drugalcdep.2019.01.037>
40. Neale J, Tompkins C, Sheard L. Barriers to accessing generic health and social care services: A qualitative study of injecting drug users. *Health & Social Care in the Community*. 2008;16(2):147–54. <https://doi.org/10.1111/j.1365-2524.2007.00739.x>
41. Sankar P, Jones NL. To tell or not to tell: Primary care patients' disclosure deliberations. *Arch Intern Med*. 2005;165(20):2378–83. <https://doi.org/10.1001/archinte.165.20.2378>
42. Pearce LA, Hodayra F, Dale LM, Moallem S, Barker B, Norton A, et al. Non-disclosure of drug use in outpatient health care settings: Findings from a prospective cohort study in Vancouver, Canada. *Int J Drug Policy*. 2020;84:102873. <https://doi.org/10.1016/j.drugpo.2020.102873>
43. Islam MM, Topp L, Iversen J, Day C, Conigrave KM, Maher L. Healthcare utilisation and disclosure of injecting drug use among clients of Australia's needle and syringe programs. *Aust N Z J Public Health*. 2013;37(2):148–54. <https://doi.org/10.1111/1753-6405.12032>
44. Rockett IRH, Putnam SL, Jia H, Smith GS. Declared and undeclared substance use among emergency department patients: A population-based study. *Addiction*. 2006;101(5):706–12. <https://doi.org/10.1111/j.1360-0443.2006.01397.x>
45. Hormes JM, Gerhardstein KR, Griffin PT. Under-reporting of alcohol and substance use versus other psychiatric symptoms in individuals living with HIV. *AIDS Care*. 2012;24(4):420–3. <https://doi.org/10.1080/09540121.2011.608795>
46. Johansson A, Vorobjov S, Heimer R, Dovidio JF, Uusküla A. The Role of Internalized Stigma in the Disclosure of Injecting Drug Use Among People Who Inject Drugs and Self-Report as HIV-Positive in Kohtla-Järve, Estonia. *AIDS Behav*. 2017;21(4):1034–43. <https://doi.org/10.1007/s10461-016-1647-8>
47. Nielsen CM, Hjorthøj C, Killaspy H, Nordentoft M. The effect of flexible assertive community

- treatment in Denmark: A quasi-experimental controlled study. *The Lancet Psychiatry*. 2021;8(1):27–35. [https://doi.org/10.1016/S2215-0366\(20\)30424-7](https://doi.org/10.1016/S2215-0366(20)30424-7)
48. Carroll M, Sutherland G, Kemp-Casey A, Kinner SA. Agreement between self-reported healthcare service use and administrative records in a longitudinal study of adults recently released from prison. *Health & Justice*. 2016;4(1):11. <https://doi.org/10.1186/s40352-016-0042-x>
  49. Borschmann R, Young JT, Moran P, Spittal MJ, Snow K, Mok K, et al. Accuracy and predictive value of incarcerated adults' accounts of their self-harm histories: Findings from an Australian prospective data linkage study. *CMAJ Open*. 2017;5(3):E694–e701. <https://doi.org/10.9778/cmajo.20170058>
  50. Keen C, Kinner SA, Borschmann R, Young JT. Comparing the predictive capability of self-report and medically-verified non-fatal overdose in adults released from prison: A prospective data linkage study. *Drug Alcohol Depend*. 2020;206:107742. <https://doi.org/10.1016/j.drugalcdep.2019.107742>
  51. Youens D, Moorin R, Harrison A, Varhol R, Robinson S, Brooks C, et al. Using general practice clinical information system data for research: The case in Australia. *Int J Population Data Science*. 2020;5(1):01. <https://doi.org/10.23889/ijpds.v5i1.1099>
  52. Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, et al. The Danish health care system and epidemiological research: From health care contacts to database records. *Clin Epidemiol*. 2019;11:563–91. <https://doi.org/10.2147/clep.S179083>
  53. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekblom A. The Swedish personal identity number: Possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol*. 2009;24(11):659–67. <https://doi.org/10.1007/s10654-009-9350-y>
  54. United Nations. Register-based statistics in the Nordic countries: Review of best practices with focus on population and social statistics. 2007.
  55. Connelly R, Playford CJ, Gayle V, Dibben C. The role of administrative data in the big data revolution in social science research. *Social Science Research*. 2016;59:1–12. <https://doi.org/10.1016/j.ssresearch.2016.04.015>
  56. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. *Int J Epidemiol*. 2016;45(3):954–64. <https://doi.org/10.1093/ije/dyv322>
  57. Taylor LK, Irvine K, Iannotti R, Harchak T, Lim K. Optimal strategy for linkage of datasets containing a statistical linkage key and datasets with full personal identifiers. *BMC Med Inform Decis Mak*. 2014;14(85). <https://doi.org/10.1186/1472-6947-14-85>
  58. Doidge JC, Harron KL. Reflections on modern methods: Linkage error bias. *International Journal of Epidemiology*. 2019;48(6):2050–60. <https://doi.org/10.1093/ije/dyz203>
  59. Harron KL, Doidge JC, Knight HE, Gilbert RE, Goldstein H, Cromwell DA, et al. A guide to evaluating linkage quality for the analysis of linked data. *International Journal of Epidemiology*. 2017;46(5):1699–710. <https://doi.org/10.1093/ije/dyx177>
  60. Swansea University. Secure Anonymised Information Linkage Databank (SAIL Databank) 2019 [Available from: <https://popdatasci.swan.ac.uk/centres-of-excellence/sail/#>].
  61. Australian Institute of Health and Welfare. Our data collections 2023 [Available from: <https://www.aihw.gov.au/about-our-data/our-data-collections>]
  62. SAIL Databank 2021 [Available from: <https://saildatabank.com/>]
  63. Ark TK, Kesselring S, Hills B, McGrail K. Population Data BC: Supporting population data science in British Columbia. *Int J Population Data Science*. 2020;4(2) <https://doi.org/10.23889/ijpds.v4i2.1133>
  64. Hung YW, Hoxha K, Irwin BR, Law MR, Grépin KA. Using routine health information data for research in low- and middle-income countries: A systematic review. *BMC Health Services Research*. 2020;20(1):790. <https://doi.org/10.1186/s12913-020-05660-1>
  65. Mutale W, Chintu N, Amoroso C, Awoonor-Williams K, Phillips J, Baynes C, et al. Improving health information systems for decision making across five sub-Saharan African countries: Implementation strategies from the African Health Initiative. *BMC Health Services Research*. 2013;13(2):S9. <https://doi.org/10.1186/1472-6963-13-S2-S9>
  66. Okullo AE, Matovu JKB, Ario AR, Opigo J, Wanzira H, Oguttu DW, et al. Malaria incidence among children less than 5 years during and after cessation of indoor residual spraying in Northern Uganda. *Malaria Journal*. 2017;16(1):319. <https://doi.org/10.1186/s12936-017-1966-x>
  67. Trutwein B, Holman CD, Rosman DL. Health data linkage conserves privacy in a research-rich environment. *Ann Epidemiol*. 2006;16(4):279–80. <https://doi.org/10.1016/j.annepidem.2005.05.003>
  68. Lea NC, Nicholls J, Dobbs C, Sethi N, Cunningham J, Ainsworth J, et al. Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research. *JMIR Med Inform*. 2016;4(2):e22. <https://doi.org/10.2196/medinform.5571>
  69. Australian Institute of Health and Welfare. The Five Safes framework. 2023.

70. Schneider M. Secure Automated File Exchange (SAFE) – Enabling More Efficient Transfers of Sensitive Data. *Int J Population Data Science*. 2020;5(5). <https://doi.org/10.23889/ijpds.v5i5.1599>
71. Coulson TG, Bailey M, Reid C, Shardey G, Williams-Spence J, Huckson S, et al. Linkage of Australian national registry data using a statistical linkage key. *BMC Med Inform Decis Mak*. 2021;21(1):37. <https://doi.org/10.1186/s12911-021-01393-1>
72. von Sanden N. Improving Inter-Agency Data Sharing Through Linkage Spine Interoperability. *Int J Population Data Science*. 2020;5(5). <https://doi.org/10.23889/ijpds.v5i5.1577>
73. Australian Bureau of Statistics. A guide for data integration projects involving Commonwealth data for statistical and research purposes: The separation principle 2023 [Available from: <https://statisticaldataintegration.abs.gov.au/topics/applying-the-separation-principle>]
74. Belfrage S, Helgesson G, Lynøe N. Trust and digital privacy in healthcare: a cross-sectional descriptive study of trust and attitudes towards uses of electronic health data among the general public in Sweden. *BMC Medical Ethics*. 2022;23(1):19. <https://doi.org/10.1186/s12910-022-00758-z>
75. Institute of Medicine (US) Roundtable on Value & Science-Driven Health Care. *Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary*. National Academies Press (US). 2010;5, *Healthcare Data as a Public Good: Privacy and Security*.
76. Harkness F, Rijneveld C, Liu Y, Kashef S, Cowan M. A UK-wide public dialogue exploring what the public perceive as 'public good' use of data for research and statistics. *Administrative Data Research UK (ADR UK) Office for Statistics Regulation*; 2022. Available from: [https://osr.statisticsauthority.gov.uk/wp-content/uploads/2022/10/Public\\_perceptions\\_of\\_public\\_good.pdf](https://osr.statisticsauthority.gov.uk/wp-content/uploads/2022/10/Public_perceptions_of_public_good.pdf).
77. Sethi N, Laurie GT. Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. *Medical Law International*. 2013;13(2-3):168–204. <https://doi.org/10.1177/0968533213508974>
78. McDonald SA, Hutchinson SJ, Bird SM, Mills PR, Dillon J, Bloor M, et al. A population-based record linkage study of mortality in hepatitis C-diagnosed persons with or without HIV coinfection in Scotland. *Stat Methods Med Res*. 2009;18(3):271–83. <https://doi.org/10.1177/0962280208094690>
79. Mallitt K-A, Wilson DP, Jansson J, McDonald A, Wand H, Post JJ. Identifying missed clinical opportunities for the earlier diagnosis of HIV in Australia: A retrospective cohort data linkage study. *PLOS ONE*. 2018;13(12):e0208323. <https://doi.org/10.1371/journal.pone.0208323>
80. Walker DM, Hefner JL, DePuccio MJ, Garner JA, Headings A, Joseph JJ, et al. Approaches for overcoming barriers to cross-sector data sharing. *Am J Manag Care*. 2022;28(1):11–6. <https://doi.org/10.37765/ajmc.2022.88811>
81. Susha I, Rukanova B, Zuiderwijk A, Gil-Garcia JR, Gasco Hernandez M. Achieving voluntary data sharing in cross sector partnerships: Three partnership models. *Information and Organization*. 2023;33(1):100448. <https://doi.org/10.1016/j.infoandorg.2023.100448>
82. Kinner SA, Young JT. Understanding and improving the health of people who experience incarceration: An overview and synthesis. *Epidemiologic Reviews*. 2018;40(1):4–11. <https://doi.org/10.1093/epirev/mxx018>
83. NHS Health Research Authority. Confidentiality advisory group 2023 [Available from: [https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/#:~:text=The%20Confidentiality%20Advisory%20Group%20\(CAG,Health%20for%20non%2Dresearch%20uses\]](https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/#:~:text=The%20Confidentiality%20Advisory%20Group%20(CAG,Health%20for%20non%2Dresearch%20uses))
84. Tweed E. *Unlocking data to inform public health policy and practice: Decision-maker perspectives on the use of cross-sectoral data as part of a whole-systems approach*. University of Glasgow: National Institute for Health Research; 2021.
85. Manuel DG, van Walraven C, Forster AJ. A commentary on the value of hospital data for covid-19 pandemic surveillance and planning. *Int J Population Data Science*. 2020;5(4) <https://doi.org/10.23889/ijpds.v5i4.1393>
86. Flack F, Smith M. The Population Health Research Network - Population data centre profile. *Int J Population Data Science*. 2019;4(2) <https://doi.org/10.23889/ijpds.v4i2.1130>
87. Andrew NE, Sundararajan V, Thrift AG, Kilkenny MF, Katzenellenbogen J, Flack F, et al. Addressing the challenges of cross-jurisdictional data linkage between a national clinical quality registry and government-held health data. *Aust N Z J Public Health*. 2016;40(5):436–42. <https://doi.org/10.1111/1753-6405.12576>
88. Victoria State Government. *National mutual acceptance: National system for mutual acceptance of scientific and ethical review of multi-centre human research projects*. 2023.
89. Health Data Research UK. *Main webpage 2023* [Available from: <https://www.hdruk.ac.uk/about-us/>]
90. Jorm L. Routinely collected data as a strategic resource for research: Priorities for methods and workforce. *Public Health Research and Practice*. 2015;25(4):e2541540. <http://doi.org/10.17061/phrp2541540>
91. Schull M, Brundo M, Ghassemi M, Gibson G, Goldenberg A, Paprica P, et al. *Building a research*

- partnership between computer scientists and health service researchers for access and analysis of population-level health datasets. *Int J Population Data Science*. 2020;5(5) <https://doi.org/10.23889/ijpds.v5i5.1529>
92. Bridges D. 'Nothing about us without us': the ethics of outsider research. *Fiction written under Oath? Essays in Philosophy and Educational Research*. Dordrecht: Springer Netherlands; 2003. p. 133-51.
93. National Health and Medical Research Council. *Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and Communities*. 2018.
94. Open Society Foundations. *Nothing about us without us: Greater, meaningful involvement of people who use illicit drugs*. 2008.
95. Williams M. Ngaa-bi-nya Aboriginal and Torres Strait Islander program evaluation framework. *Evaluation Journal of Australia*. 2018;18(1):6–20. <https://doi.org/10.1177/1035719X18760141>
96. Walter M, Lovett R, Maher B, Williamson B, Prehn J, Bodkin-Andrews G, et al. Indigenous data sovereignty in the era of big data and open data. *Aust J Soc Issues*. 2020;1–14. <https://doi.org/10.1002/ajs4.141>
97. Mecredy G, Naponse-Corbiere P, Walker J. Collaboration with First Nations Communities to Produce Tailored Community-Driven Results. *Int J Population Data Science*. 2020;5(5) <https://doi.org/10.23889/ijpds.v5i5.1515>
98. National Health Service Scotland. *Public Benefit and Privacy Panel for Health and Social Care*. 2023.
99. Morris C, Gray A, Scott J. NHS Scotland Public Benefit and Privacy Panel (PBPP) – Much Ado about Governance 3 years on. *Int J Population Data Science*. 2018;3(4). <https://doi.org/10.23889/ijpds.v3i4.808>
100. Viano S, Baker D. How administrative data collection and analysis can better reflect racial and ethnic identities. *Review of Research in Education*. 2020;44:301–31. <https://doi.org/10.3102/0091732X20903321>

