

A scientometric review of health data sharing for secondary use: Insights, frontiers and the path ahead

Health Information Management Journal

1–12

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/18333583251393431

journals.sagepub.com/home/himj



Michelle A Krahe, *BAppSci(Hons), PhD¹* , **Rebekah Eden**, *BIT(hons), BAppSc, PhD²* , **Jason D Pole**, *BSc(Hons), MSc, PhD²* , **Bernadette Richards**, *BA, Dip Ed, LLB(Hons), PhD²* , **Quita Olsen**, *MEpi²* , **Amalie Dyda**, *BHLthSc(Hons), MAE, PhD²* , **Elton Lobo**, *BEng, MEng, PhD²*, **Nalini Pather**, *BMedSc(Hons), MMedSci., GCULT, AdvDip(Ethics), PhD^{2,3}* , **Clair Sullivan**, *MD²* and **Leanna Woods**, *BN(Hons), PhD²* 

Abstract

Background: As digital technologies advance, vast amounts of routinely collected health data are increasingly available for quality improvement and research. However, concerns persist around the reuse of personal health information. Understanding public attitudes and knowledge is essential to building social licence and enabling ethical, large-scale data use. **Objective:** This study explores key research themes in sharing health data for secondary use since 2020, highlighting major topics, emerging research frontiers and future directions for practice. **Method:** An analysis of 95 publications from Web of Science, PubMed and Scopus was conducted using scientometric methods. Citation, co-citation and keyword co-occurrence analyses, along with strategic diagrams, were performed using VOSviewer to identify thematic clusters. **Results:** Research has shifted from early exploratory studies to more multidisciplinary and technology-focused approaches. Key themes include digital tool adoption, integrated data systems and ethical data sharing solutions. The concept of consent has seen the most theoretical development, while public attitudes – particularly around ethical and sociocultural issues – remain underexplored but crucial. **Conclusion:** Ethical governance, transparency and community engagement are central to advancing health data sharing. Building public trust and securing a social licence are foundational to success, especially as challenges around consent, data linkage and public perception remain. **Implications for health information management practice:** This analysis provides insight into public willingness to share health data for secondary data use and offers guidance for fostering a strong social licence while building public trust. Strengthening these trust and engagement frameworks is vital to achieving ethical data use and maximising the potential health system benefits of secondary data use.

Keywords (MeSH)

data sharing; public opinion; social responsibility; informed consent; trust; digital health; health information management

Introduction

In an era of rapid digitalisation and the advancement of sophisticated health technologies (e.g. artificial intelligence, wearable devices, augmented reality), the management of health data, the dynamics of data sharing and public perceptions regarding the distribution and repurposing of data have become increasingly critical (Alam et al., 2024). Balancing patients' right to personal privacy with the societal benefits of health data reuse presents a persistent and nuanced challenge. This tension became pronounced during the COVID-19 pandemic, when efforts to leverage information technology (e.g. contact tracing and telemedicine), relied heavily on the collection and dissemination of health data. These efforts aimed to enhance patient care and improve healthcare system efficiency (Hvalič-Touzery et al., 2024; Sullivan et al., 2021; Tosoni et al., 2022). These initiatives raised a range of privacy

(both regulatory and legal aspects), ethical, political, technical and social concerns (Chan and Saqib, 2021; Gerke et al., 2020), while also highlighting their significant role in saving millions of lives. This underscores the importance of measuring and securing public trust and support for the use of health data, not only within direct healthcare delivery but also in broader public health contexts (Kerasidou and Kerasidou, 2023).

¹James Cook University, Australia

²The University of Queensland, Australia

³UNSW Sydney, Australia

Accepted for publication April 22, 2025

Corresponding author:

Michelle A Krahe, College of Medicine and Dentistry, James Cook University, 1/14-88 McGregor Rd, Smithfield, QLD 4878, Australia.
Email: michelle.krahe@jcu.edu.au

The concept of *social licence* – an intangible yet critical form of public, community or stakeholder approval – plays a pivotal role in activities that directly affect individuals, particularly when sensitive information such as health data are involved (Muller et al., 2021). Health data encompass any information related to individuals or populations' health and may be structured, unstructured, identified, identifiable or de-identified. These data serve to advance medical research, improve the quality and efficiency of care, inform public health initiatives and ultimately improve healthcare outcomes (Murdoch and Detsky, 2013). In the context of secondary uses of health data, obtaining and maintaining a social licence requires more than legal compliance or technical safeguards; it depends on aligning data repurposing with public values, interests and expectations (Muller et al., 2021). A recent systematic review by Benevento et al. (2023) identified the type of data use as the most significant determinant of individuals' willingness to share their health data. This willingness was shaped by trust and confidence in the responsible, ethical and transparent use and handling of data, which, in turn, depended on how effectively concerns regarding privacy, consent and potential misuse were addressed (Benevento et al., 2023). As technical capacity for data sharing and linkage grows, traditional mechanisms such as informed consent and de-identification may become insufficient to sustain public confidence (Adams et al., 2022). Social licence is therefore developed and maintained through ongoing engagement, transparent governance, ethical oversight and responsiveness to community concerns. Establishing such a foundation involves iterative dialogue with stakeholders, showing accountability and ensuring that data practices remain aligned with evolving societal values. Understanding these factors is essential for fostering social licence and developing strategies to support the ethical and effective use of health data at scale.

While research indicates that people are generally willing to share their health data for research purposes, this willingness is contingent on factors such as the type of data recipient, the nature of the data, consent, sociodemographic characteristics (e.g. race, education, religion) and health literacy (Brall et al., 2021; Cascini et al., 2024; Hutchings et al., 2020; Kim et al., 2019; Kirkham et al., 2022; Seltzer et al., 2019; Soni et al., 2019). Trust, emerging as a dominant factor, operates on multiple levels and is essential for establishing information exchange partnerships; it is a key component of social licence and results from transparency, mutual understanding and accountability (Kerasidou and Kerasidou, 2023; Naeem et al., 2022). Findings from Braunack-Mayer et al. (2024) reinforce these patterns: participants were highly supportive of sharing general practice data with their clinicians and for direct patient benefit, but showed lower willingness to share data for secondary purposes such as research or health service planning. These patterns are consistent with Australian citizen jury deliberations, which found that informed community members generally supported sharing government-held health data with private industry for research and development, provided the intended purpose was clearly in the public interest,

responsible governance frameworks were in place and the data were securely managed (Street et al., 2021). Similarly, a national survey of Australians found that just over half of participants were willing to share government health data with private companies, with strong support for opt-in consent and conditions on data sharing (Braunack-Mayer et al., 2021). Participants expressed concerns about private sector corporate interests, profit motives. They also questioned the government's ability to manage data safely, indicating that public confidence is conditional on transparency, ethical oversight and accountability. Collectively, these studies, along with earlier reviews provide a nuanced understanding of public preferences and concerns regarding health data sharing, particularly with private sector actors. They offer essential guidance for strengthening trust, fostering social licence and promoting responsible and ethical secondary use of health data.

Building on the existing literature, an identified gap lay in the tendency to generalise the nature of the social licence to share, and the challenges associated with it. This review aimed to address this gap by synthesising current trends, offering a comprehensive analysis and contextualising these findings within broader frameworks. By doing so, it sought to provide insights into how these challenges could be navigated, with a particular emphasis on building trust and securing social licence for secondary data sharing. Our inquiry was guided by the following research questions:

- What were the primary conceptual themes that influenced the public's willingness to share health data for secondary use, since 2020?
- Which specific topics within health data sharing for secondary use attracted the most scholarly attention, and what were the research frontiers?
- What contemporary trends are emerging from the literature that could shape future research priorities and practice approaches to health information management?

Method

This review employed the scientometric methods of document co-citation and keyword co-occurrence analyses to examine the contemporary knowledge base and trends in a defined body of literature representing the public's (e.g. patients, health consumers, citizens) perceptions of the secondary use of health data (Du et al., 2024). Scientometrics is a branch of bibliometrics, characterised by documenting and visualising the structural and relational features of the accumulated knowledge base within a specific discipline or topic (van Eck and Waltman, 2014). Given its ability to capture the evolution and focus of research, this approach is well-suited for analysing the conceptual and topical trends in a body of literature.

Identification of documents

The dataset used in this review is drawn from our systematic review and meta-analysis, which examined public

perceptions of health data repurposed for secondary use (Olsen et al., 2025). The review was scoped to include peer-reviewed, full-text primary research articles (qualitative, quantitative or mixed-methods) published in English between January 2020 and December 2023 (to map the contemporary research front). Eligible studies explored the perceptions of the public or health consumers across all demographic groups. Studies were excluded if they focused on health care professionals, representatives from commercial health organisations, or data generated and stored outside health organisations (e.g. wearable devices, social media). Clinical trials (where consent for data sharing had already been obtained), reviews, editorials, commentaries, grey literature, protocols and conference abstracts were also excluded. All records identified and screened in the systematic review formed the source dataset for this scientometric review.

Data analysis and visualisation

Bibliometric data were sourced from the core collection database of Web of Science, PubMed or Scopus on 17 December 2024, and metadata were collated into a CSV file for analysis. Data analysis and visualisation were conducted using VOSviewer software (version 1.6.7; van Eck and Waltman, 2010).

Document citation and co-citation analysis. The first research question was addressed through document citation and co-citation analyses. Citations are frequently used in bibliometric studies as a metric of scholarly influence, with highly cited documents often reflecting key research foci within a field. Document citation analysis was used to identify the most frequently cited documents and examine distinguishing conceptual themes. Co-citation analysis was used to assess the frequency with which documents are cited together in the reference lists (Saxena et al., 2024).

Keyword co-occurrence analysis. The second research question was addressed using keyword co-occurrence analyses to explore the relationships between keywords. Co-word analysis is a text-mining technique that analyses the co-occurrence of word pairs, where keywords frequently appearing together in the same documents, are likely related. In this analysis, keywords (i.e. author-defined and indexed terms) – key terms or phrases in the titles and abstracts frequently associated with a specific topic or research area – were extracted. Keywords with three or more co-occurrences were retained and manually reviewed for ambiguous or insignificant words, such as function words and irrelevant verbs, which subsequently were excluded. The synonyms of keywords were merged and standardised (e.g. secondary use and secondary data use).

Data visualisation. A network map was generated using VOSviewer. To identify thematic clusters, the force-directed layout algorithm with linlog/modularity normalisation method was applied to adjust for potential bias (van Eck and Waltman, 2014). Each keyword is represented by a

node, with the size reflecting the frequency of the keyword's occurrence. The edges (i.e. connections) between nodes represent co-occurrence relationships, indicating that two terms appeared together in a document; the thickness of the edge reflects the frequency of co-occurrence, with the maximum number of lines set to 500. Closely related nodes are grouped into clusters distinguished by unique colours, which represent sets of words that frequently co-occur and form distinct thematic areas. The clusters are described using the metrics:

- **Keywords:** A set of keywords that constitute a particular cluster (i.e. research theme).
- **Size:** The number of keywords in the cluster.
- **Frequency:** The average number of keyword occurrences for all keywords in the cluster.
- **Total Link Strength (TLS):** The total strength of the links between a keyword and other keywords, for all keywords in the cluster.
- **Average Citation Score (ACS):** The average citation impact of all documents associated with the keywords in that cluster.

A density map was generated to visualise the evolution of research topics based on the relative frequency of keyword occurrences over time. Colour graduations illustrate the frequency with which keywords appear on average across different time periods. Darker-coloured (i.e. purple and blue) nodes are associated with topics studied in earlier literature, while lighter-coloured nodes correspond to topics in more recent documents.

Strategic diagram. The third research question was addressed using strategic diagram analysis, a bibliometric method that visually maps thematic research clusters. This approach positions clusters within a four-quadrant layout based on two key dimensions: centrality (their degree of external connectivity) and density (the level of internal cohesion). This analytic technique, widely adopted in recent scientometric studies (Cobo et al., 2011), provides insights into both the maturity and relevance of research themes, helping to reveal how topics are structured and interlinked across the field. For this study, centrality and density values were calculated using Gephi (version 0.10; Bastian et al., 2009), following standard co-word network analysis protocols.

- The *x*-axis represents betweenness centrality, which reflects how strongly a thematic cluster connects to other clusters in the network. Higher centrality values indicate a theme's influence and its role in bridging different areas of research, highlighting interdisciplinary significance.
- The *y*-axis represents density, capturing the internal strength and cohesion of the cluster. High-density clusters typically indicate a well-developed theme with substantial conceptual and methodological consistency.

The diagram is divided into four quadrants:

- **Quadrant one (Q1):** Core themes – clusters that are both highly central and dense, representing core, well-developed topics that are pivotal to the structure and evolution of the research field.
- **Quadrant two (Q2):** Specialised themes – clusters with high density but low centrality, indicating specialised, mature topics that are internally coherent but less connected to other areas. These often represent focused subfields or methodologies.
- **Quadrant three (Q3):** Emerging or declining themes – clusters with low density and low centrality, often reflecting early-stage topics yet to mature, or themes that are losing relevance and traction over time.
- **Quadrant four (Q4):** Foundational themes – clusters with high centrality but low density, representing foundational topics that are widely relevant across the field but remain underdeveloped or conceptually diffuse. These clusters often indicate opportunities for future research development and integration.

Results

From an initial 4085 documents, 95 met the inclusion criteria and formed the primary dataset for this study. Overall, a gradual increase in publication frequency over time was observed – from 17 documents (17.9%) published in 2020, to 32 (33.67%) published in 2023. In terms of increasing annual scientific production, the median yearly growth rate

was 18.5%, with a maximum of 42.1% in 2022. Research included contributions by 33 countries or regions (63.6% developed), 293 organisations or institutions and 515 authors. The 95 documents were published in 68 journals, with 38 (55.89%) appearing in 10 journals. The three most frequently occurring journals in the dataset are *BMC Medical Ethics* ($n=10$), *Journal of Medical Internet Research* ($n=6$), and the *International Journal of Medical Informatics* ($n=5$).

Document citation and co-citation analysis

Table 1 presents the 10 most cited documents. Of these, survey designs ($n=6$), focus groups ($n=2$), interviews ($n=1$) or a combination of surveys with workshops ($n=1$) were used. Study cohorts varied in size, ranging from 30 to nearly 37,000 people, and the types of health data examined included general health data, and more specifically, personal medical records, and genomic datasets. The distribution of research focus included: how individuals perceive artificial intelligence (AI) and the secondary use of health data ($n=3$), ethical and trust considerations in data reuse ($n=3$) and factors that shape perceptions and willingness to share data ($n=4$).

Document co-citation analysis identified 2670 citations from the reference lists of the primary documents. Table 2 shows the 10 most co-citations, representing the field's most influential works. Of these, five were reviews, four were empirical, and one was a commentary. The research focus observed in the citation analysis (social licence and public acceptability, data sharing and governance, trust and

Table 1. Top 10 cited documents on health data sharing for secondary use (2020–2023).

Rank	Document	Focus	Citations
1	Middleton et al. (2020). Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data?	Perceptions and willingness to share	75
2	McCradden et al. (2020a). Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study.	AI and secondary use of health data	46
3	Aggarwal et al. (2021). Patient perceptions on data sharing and applying artificial intelligence to health care data: cross-sectional survey.	AI and secondary use of health data	47
4	McCradden et al. (2020b). Conditionally positive: a qualitative study of public perceptions about using health data for artificial intelligence research.	AI and secondary use of health data	37
5	Milne et al. (2021). Demonstrating trustworthiness when collecting and sharing genomic data: public views across 22 countries.	Ethical and trust considerations	33
6	Richter et al. (2021). Secondary research use of personal medical data: attitudes from patient and population surveys in The Netherlands and Germany.	Perceptions and willingness to share	29
7	Trinidad et al. (2020). The public's comfort with sharing health data with third-party commercial companies.	Perceptions and willingness to share	26
8	Belfrage et al. (2022). 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.	Ethical and trust considerations	24
9	Atkin et al. (2021). Perceptions of anonymised data use and awareness of the NHS data opt-out amongst patients, carers and healthcare staff.	Ethical and trust considerations	20
10	Hassan et al. (2020). A deliberative study of public attitudes towards sharing genomic data within NHS genomic medicine services in England.	Perceptions and willingness to share	20

AI: Artificial intelligence; DNA: Deoxyribonucleic acid; NHS: National Health Service.

Table 2. Top 10 co-cited documents in the health data sharing for secondary use literature (2020–2023).

Rank	Document	Focus	Document type	Co-citations
1	Aitken et al. (2016). Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies.	Social licence and public acceptability	Review	26
2	Kalkman et al. (2022). Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence.	Data sharing and governance	Review	24
3	Carter et al. (2015). The social licence for research: why care. data ran into trouble.	Social licence and public acceptability	Commentary	13
4	Garrison et al. (2016). A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States.	Data sharing and governance	Review	13
5	Hill et al. (2013). "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study.	Social licence and public acceptability	Review	13
6	Spencer et al. (2016). Patient perspectives on sharing anonymised personal health data using a digital system for dynamic consent and research feedback: A qualitative study.	Consent mechanisms	Empirical	13
7	Damschroder et al. (2007). Patients, privacy and trust: Patients' willingness to allow researchers to access their medical records.	Trust and privacy	Empirical	12
8	Stockdale et al. (2018). "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.	Social licence and public acceptability	Review	12
9	Sanderson et al. (2017). Public attitudes towards consent and data sharing in biobank research: A large multi-site experimental survey in the United States.	Consent mechanisms	Empirical	11
10	Ghafur et al. (2020). Public perceptions on data sharing: key insights from the UK and the USA.	Trust and privacy	Empirical	10

privacy, and consent mechanisms) represents the organising concepts in this key literature.

Keyword co-occurrence analysis

A final set of author-defined and indexed keywords was analysed ($n=67$). The most frequently occurring keywords were *privacy* ($n=32$; TLS=196), *data sharing* ($n=29$; TLS=169), *attitude* ($n=26$; TLS=162), *consent* ($n=25$; TLS=161), and *trust* ($n=24$; TLS=157). Figure 1 is a co-word network map representing 67 nodes, 762 edges and a TLS of 1439. Keywords are organised into six coloured clusters, each representing conceptual similarity among the keywords. As shown in Table 3, the largest thematic cluster (C01:red) comprises 17 keywords, with a moderate ACS of 5.52 and the highest TLS of 685. The smallest clusters (C05:purple and C06:orange) each consist of seven keywords. Cluster 5 has the highest ACS (12.28) and the highest frequency (10.85), emphasising the prominence of these keywords within the dataset but also has the lowest TLS (206), indicating fewer connections with other clusters.

Figure 2 depicts the temporal development of co-words in health data sharing for secondary use research, grouped into three periods. The earliest period, represented by dark purple nodes, focused on public perceptions (perceptions, attitude, perspectives), ethical constructs (ethics, consent, public trust), and the potential of health data to advance medical research and support precision healthcare (data linkage, precision medicine, participation, big data, AI).

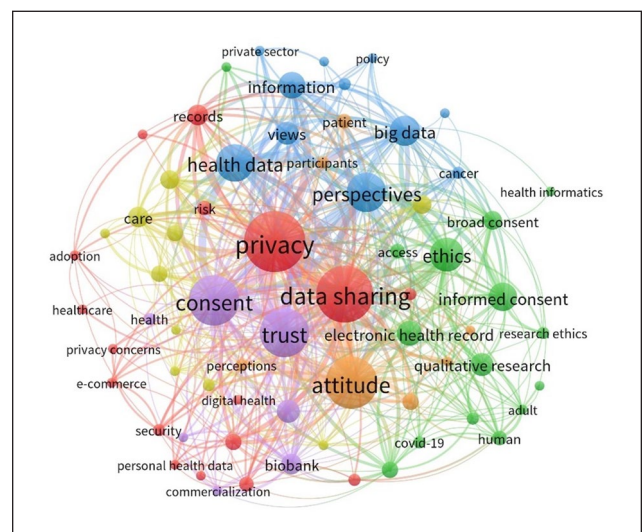


Figure 1. Co-word network visualisation of the health data sharing literature (2020–2023).

Note: Map generated in VOSviewer.

The middle period, indicated by dark green nodes, marks a transition to a more multidisciplinary focus with an emphasis on the ethical, social and technical dimensions of data sharing. This phase also reflects the integration of advanced technologies (health informatics, DNA, biobank, data sharing), greater engagement with governance and public attitudes (data governance, public engagement, attitudes and deliberation, views), and a heightened urgency driven by

Table 3. Thematic clusters in health data sharing research for the period of 2020–2023.

ID	Colour	Keywords ^a	Size	Frequency	TLS	ACS
C01	Red	privacy, data sharing, records, willingness, risk	17	7.11	685	5.52
C02	Green	ethics, informed consent, electronic health record, qualitative research, broad consent	15	6.20	502	9.59
C03	Blue	perspectives, health data, big data, information, views	11	8.00	567	9.09
C04	Yellow	care, governance, public attitudes, secondary data use, preferences	10	5.00	331	8.43
C05	Purple	consent, trust, participation, biobank, health	7	10.85	206	12.28
C06	Orange	attitude, data governance, patient, perceptions, participants	7	7.28	326	5.70

TLS: Total link strength; ACS: Average citation score.
^aTop 5 keywords, listed in order of frequency.

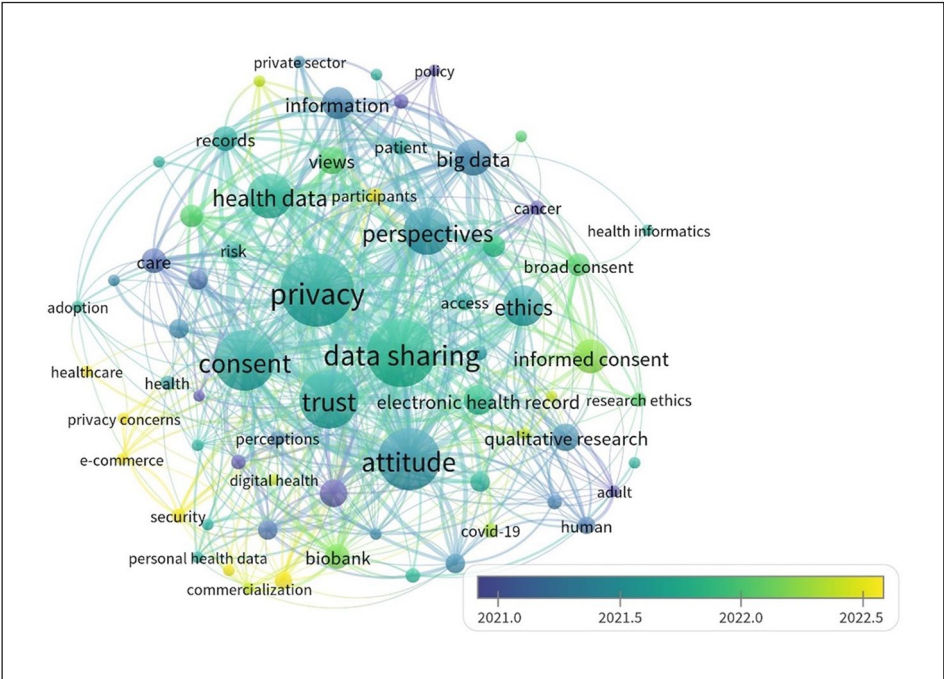


Figure 2. Co-word network map with density overlay according to average publication year.
Note: Map generated in VOSviewer.

global events like the COVID-19 pandemic (patient, research ethics, trust, privacy, risk, COVID-19). The research front appears as light green and yellow nodes. This research focuses on integrating health data (digital health, healthcare, technology), the adoption of digital tools (commercialisation, e-commerce) and data sharing solutions (security, privacy concerns).

Strategic diagram

Figure 3 visually represents the relationships between each of the thematic keyword clusters using a quadrant-based layout. These are described below and in more detail in the discussion section.

- Q1 – Core Themes: Cluster 5 keywords are characterised by high centrality and density, indicating that it represents the most central and well-developed research theme and acts as a bridge across different research areas.

- Q2 – Specialised Themes: Cluster 6 keywords have a high density but low centrality, suggesting it is a well-developed but narrowly focused area. While it may not strongly influence other themes, it represents a specialised field that complements broader research directions.
- Q3 – Emerging or Declining Themes: Clusters 2, 3 and 4 are positioned in the lower-left quadrant, indicating low density and low centrality. Their proximity to one another and partial overlap suggests interconnected subfields that might be evolving in parallel. These keyword themes may either be gaining relevance and integrating into the field or losing prominence as research priorities shift.
- Q4 – Foundational Themes: Cluster 1 keywords have a low density but higher centrality, representing foundational knowledge or broad conceptual frameworks to the field. Given the central positioning, this cluster could act as a bridge between more specialised themes and the broader research landscape.

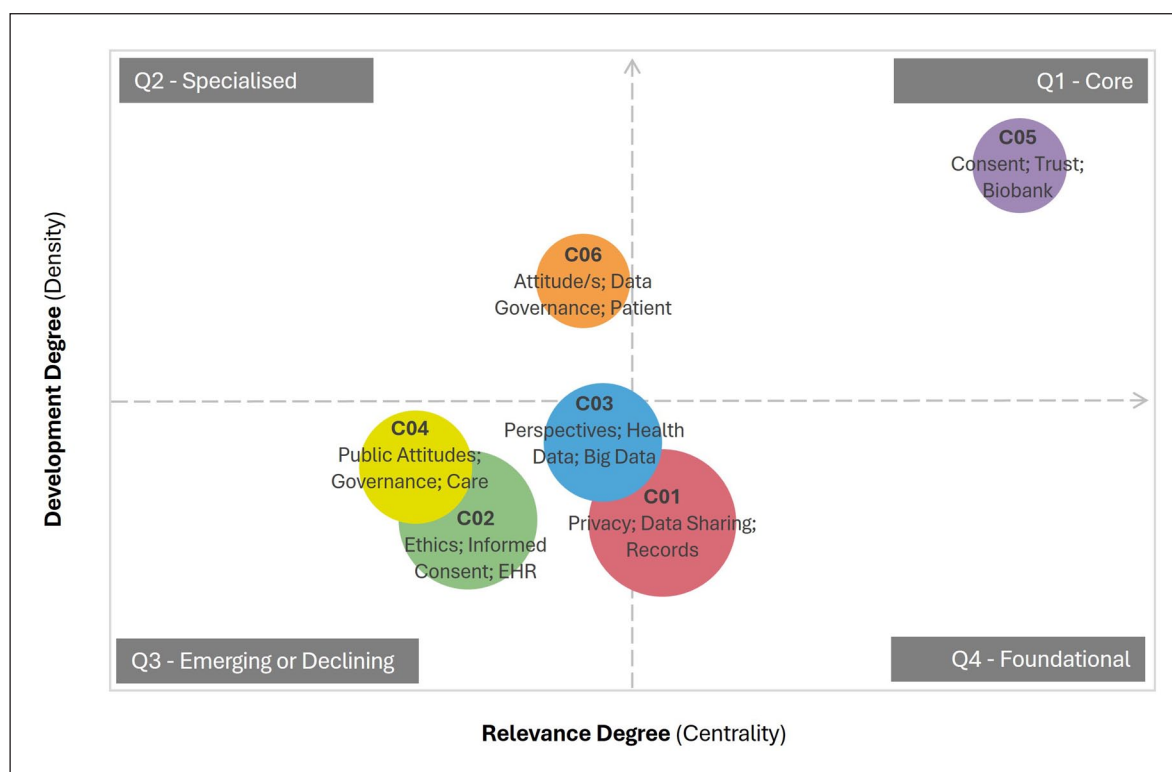


Figure 3. Strategic diagram of thematic research clusters associated with health data sharing for secondary use (2020–2023).

Box 1. Key focus areas for future research and practice priorities to health data sharing for secondary use.

Key focus area	Research priorities	Practice priorities	Expected outcomes
Regulatory and governance frameworks	Investigating ethical standards for data sharing for secondary use, focusing on dynamic consent and participatory governance.	Developing clear guidelines for policymakers and data custodians to ensure transparency, accountability and ethical compliance in data-sharing practices.	Strengthens regulatory clarity and public confidence in health data use.
Public trust and attitudes	Exploring co-designed governance models that prioritise cultural sensitivity and trust-building, particularly in underrepresented communities.	Implementing participatory processes where communities actively contribute to shaping data-sharing practices.	Enhances trust, increases willingness to share health data.
Technology integration and AI in health data analytics	Evaluating the role of AI in health data analytics while ensuring transparency, fairness and privacy through the use of privacy-preserving techniques.	Integrating explainable AI models into healthcare analytics, to improve stakeholder understanding and trust in AI-driven decisions.	Ensures responsible AI adoption and minimises algorithmic bias in healthcare.
Refining consent mechanisms	Developing dynamic, granular consent models.	Designing intuitive consent tools that balance personal autonomy with ethical and legal compliance.	Empowers individuals while maintaining data integrity and research value.
Equity and representation	Researching strategies to address underrepresentation and improve equity in health datasets.	Advocating for policies that ensure equitable access to health data and prioritise investment in public sector open-access health data initiatives.	Reduces bias in healthcare improvement initiatives and improves care for underserved populations.
Commercial interests and data ethics	Investigating ethical models, such as data cooperatives, to ensure equitable distribution of benefits in commercial health data use.	Developing governance frameworks to guide ethical commercial use, ensuring fairness and public benefit.	Prevents exploitation of health data and promotes ethical business practices.
Public engagement and communication	Researching best practices for community engagement and the role of media in shaping public perceptions of health data sharing.	Designing educational campaigns and communication strategies to improve health data literacy and public trust.	Increases public awareness, reducing misinformation and resistance to data sharing.

AI: Artificial intelligence.

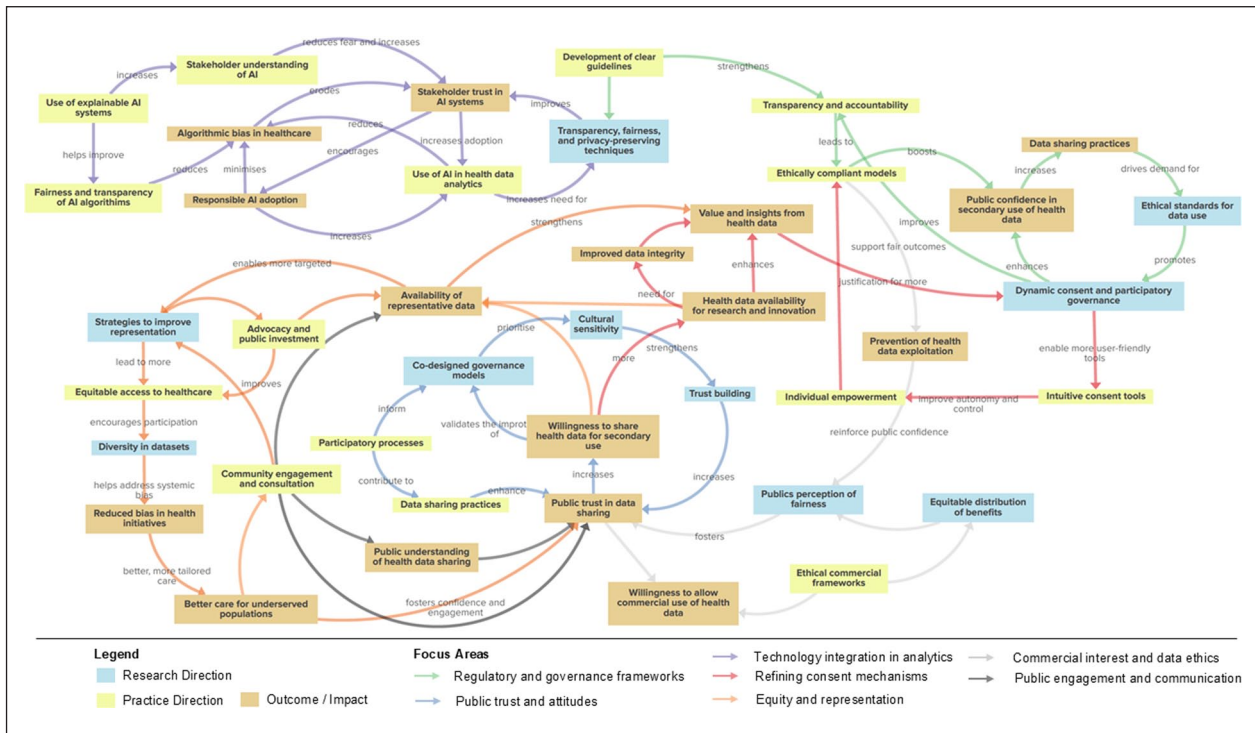


Figure 4. Conceptual map of research and practice directions to improve willingness to share health data for secondary use.

Discussion

This review sought to identify the conceptual themes and topical developments that define the contemporary knowledge base on the public's perspectives on health data sharing for secondary use and articulate emerging research and practice opportunities. Using a scientometric analysis of scholarly output between 2020 and 2023, we identified key themes, contemporary trends, and emerging areas in the field.

Interpretation of results

Conceptual themes in health data sharing for secondary use.

This review identified three areas of research focus in the contemporary knowledge base (AI and secondary data applications, ethical and trust considerations and factors influencing willingness to share data), which have received sustained attention since 2020. The first focus (AI and the secondary use of health data) underscores the expanding role of AI in healthcare, such as predictive analytics, risk stratification, personalised medicine, genomics and epidemiology (Rieke et al., 2020). As AI applications become more prevalent, understanding AI governance will be essential for fostering public trust and addressing ethical concerns (Birkstedt et al., 2023). The second focus (ethical and trust considerations) highlights the complexities of building and sustaining public confidence in data reuse. Transparency in data governance, stringent protection measures and respect for informed consent are all crucial components of ethical data sharing (Kim et al., 2019). While informed consent and de-identification are essential components of ethical health data sharing, they alone may

not adequately address the complexities of large-scale data linkage. Challenges include variations in data quality and standards, the involvement of multiple governments, agencies and private organisations, data custodians' lack of trust in external organisations, differences in legal and privacy regimes and uncertainties about whether community support extends across national borders (Adams et al., 2022). These challenges highlight the importance of robust governance frameworks, transparency and ongoing public engagement to complement consent and de-identification practices and to support ethical, effective data sharing. Additionally, ethical considerations must account for cultural and societal variations, necessitating context-specific approaches to ethical data sharing (Warren et al., 2023). The third focus examines the conditions shaping individuals' willingness to share health data with stakeholders such as healthcare providers, commercial companies and researchers. Trust, privacy concerns and perceived benefits play a critical role in these decisions.

Co-citation analysis highlighted the contributions of four prominent themes (social licence and public acceptability, data sharing and governance, consent, trust and privacy) to the theoretical underpinnings of health data sharing for secondary use. First, social licence and public acceptability play a crucial role in attitudes, as individuals are more likely to support data sharing initiatives when they perceive them as fair, transparent and beneficial (Muller et al., 2021). Historical cases like *care data* in the United Kingdom underscore the risks of inadequate public consultation and perceived commercial exploitation, which can lead to strong opposition (Aggarwal et al., 2021). Second, effective governance structures are critical for balancing the benefits of data sharing with ethical obligations and privacy

concerns (Gross et al., 2022). Policies and regulations that ensure accountability, equitable data access and robust security measures influence public confidence in data sharing and help mitigate fears of data misuse and reinforce public trust. Third, consent mechanisms also remain a central consideration in public attitudes towards data sharing (Kaplow et al., 2024; Sánchez et al., 2023). Traditional broad consent models, where individuals provide a one-time agreement for future data use, have been scrutinised in favour of more flexible approaches (Cumyn et al., 2023; Lee et al., 2023; Richter et al., 2023). Ensuring consent processes are transparent and adaptable to evolving public expectations is key to fostering sustained participation. Finally, trust and privacy concerns are fundamental to public willingness to share health data. Individuals often express hesitancy due to fears of privacy breaches, data misuse, and a lack of control over their personal information. Trust is influenced by perceptions of institutional integrity, the presence of strong privacy safeguards and clear communication about data usage. Regulatory protections, encryption technologies and transparent accountability measures can help address these concerns. Ultimately, fostering public trust requires ethical data stewardship, ongoing public engagement and a commitment to responsible data governance.

Together, the conceptual themes of social licence, governance, consent mechanisms and trust, shape public willingness to share health data for secondary use. Collectively, they frame a broader discourse on how policies and governance frameworks can either enhance or restrict data sharing practices and highlight the need for a balanced approach that prioritises ethical considerations and facilitates the potential benefits of data sharing for secondary use.

Frontiers in health data sharing for secondary use. Keyword analyses conducted in this review identified dominant topics and research frontiers. Privacy emerged as a foundational theme closely intertwined with trust, attitudes, consent and broader societal perspectives. Trust serves as the cornerstone of data sharing, with individuals more inclined to provide consent when assured of secure and ethical data use. Concerns around the protection of privacy are found across all the themes and when lapses in privacy protection occur, trust is eroded. It is therefore crucial that robust safeguards are developed and transparently shared with the public. Attitudes towards data sharing are shaped by a delicate balance of perceived risks and benefits, making transparency in consent mechanisms essential (Baines et al., 2024). As privacy perspectives vary across cultural and societal contexts, data sharing frameworks must be adaptive, ensuring they uphold diverse values while fostering public confidence in ethical research and healthcare innovation (Li, 2022).

Temporal co-word analysis highlighted the evolution of health data sharing for secondary use across three time periods (i.e. exploratory phase, transitional phase and research front). The earliest period represents the foundational work characterised by understanding public perceptions of data sharing, with a focus on attitudes, perceptions and ethical

constructs (i.e. consent, public trust and privacy). This phase laid the groundwork for more specialised research by establishing the fundamental ethical, social and technical concerns related to health data use. The middle period marks a transition towards a more multidisciplinary approach, with research shifting to include advanced technologies and broader governance issues. This phase saw the integration of health informatics, biobanks and the complex challenges in data sharing in both local and global contexts. Research during this time started to move beyond theoretical concerns to address practical, scalable solutions for data sharing. The most recent period focuses on themes related to the application of health data in specific health contexts and the integration of technology. Research during this phase is centred around the adoption and integration of health data and digital health technologies, including commercialisation and e-commerce aspects. This finding is also supported by a scientometric analysis of data sharing for precision medicine (Texier et al., 2019), which observed the emergence of keywords such as cloud, encryption, security and interoperability as newer areas of research. This period also marks the growing use of secondary health data for targeted areas like mental health (Bakken et al., 2022; Kirkham et al., 2022; Watson et al., 2023) and rare diseases (Amorim et al., 2022; Zawistowski et al., 2023), signalling a shift from broad exploratory research to the application of health data in solving specific, context-driven health challenges.

Structure and development of the research. The strategic mapping of thematic clusters offers key insights into the structure and development of research in this field. The most developed research themes focused on trust, informed consent and participation, with particular emphasis on their implications for health research and commercialisation. The presence of keywords such as biobank and DNA suggests a focus on the collection, storage and use of genetic materials in health-related research contexts. This aligns with broader discussions around precision medicine, biomedical innovation and commercialisation of health research. The specialised status of attitudes (cluster 6) towards data use and governance is well-developed but remains isolated from other areas. This highlights an opportunity for greater integration of these dimensions into other areas, such as digital health research to build more trusted, transparent and equitable digital health solutions. Addressing the public's concerns regarding data sharing for secondary use will build public trust and foster engagement with digital health technologies. The foundational themes within Cluster 1, particularly those concerning privacy, AI and data security, highlight critical yet underdeveloped research areas. As digital health technologies become more prevalent, there is a need for further exploration of privacy-enhancing mechanisms, ethical AI implementation and public policy frameworks to ensure secure and equitable data usage. Overall, these findings suggest a research landscape in transition, with well-established themes continuing to shape the field, while emerging topics such as AI and digital privacy signal important future directions.

Future research and practice priorities for health information management

As digital technologies become deeply integrated within health systems, securing and sustaining social licence will be pivotal to the success of data-driven research, innovation and healthcare. In addition to synthesising current trends and offering a comprehensive analysis, this review also seeks to contextualise our findings within the broader framework of health information management. Understanding and addressing the factors that influence individuals' willingness to share their health data for secondary use requires a coordinated, multilevel approach. Box 1 summarises the key focus areas, including corresponding research and practice priorities, while Figure 4 illustrates the interrelationships and interdependence between factors. An online interactive version is also available.¹

Limitations of the study

While scientometric analysis provides a comprehensive overview of publication patterns, trends and research outputs, it inherently focuses on high-level metrics and may not capture the detailed contextual or methodological nuances of individual studies. Consequently, our findings should be interpreted as reflecting broad patterns in the literature rather than detailed qualitative insights. Our future work complements this approach with an in-depth systematic review and meta-analysis (Olsen et al., 2025) to explore the nuanced aspects of research practices and findings. The predominance of systematic and narrative reviews among the most cited publications also highlights a potential citation bias, as reviews tend to attract more citations than primary research. Document and keyword co-occurrence analyses provide valuable structural insights but remain surface-level representations of the underlying landscape. Keyword-based analyses are constrained by database indexing practices, author-defined terms and metadata availability and can limit the ability to capture the depth of emerging or nuanced topics. Furthermore, the dataset is restricted to publications from 2020 to 2023, which should be interpreted as a contemporary snapshot rather than a long-term trend analysis. Despite these limitations, our focused approach – *drawing on a systematically selected dataset* – reduces false positives common in broader scientometric analyses. Future research could expand the temporal scope, include pre-COVID-19 literature and incorporate additional primary research to capture longitudinal trends and emergent topics. Together with our complementary systematic review (Olsen et al., 2025), these findings provide a more comprehensive understanding of public perceptions on health data sharing for secondary use.

Conclusion

This study presents a comprehensive scientometric analysis of the conceptual foundations and topical developments that shape the current literature on public perceptions on health data sharing for secondary use. Integrating theoretical insights with practice and research opportunities will support responsible advancement in a technology-driven

healthcare landscape. By prioritising ethical governance, technological transparency and meaningful community engagement, future efforts can build public trust and secure a social licence for data sharing. These measures together promote more equitable, effective and trustworthy approaches to secondary data use, supporting better health outcomes and increased societal trust.

Author contributions

All authors conducted the systematic literature review to identify the dataset. MK, RE and LW designed the study with assistance from JP, BR, NP, QO and AD. MK conducted the analysis and drafted the manuscript with assistance from RE and LW. All authors contributed to editing and approved the final version.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCID iDs

Michelle A Krahe, BAppSci(Hons), PhD  <https://orcid.org/0000-0002-3223-7825>
 Rebekah Eden, BIT(hons), BAppSc, PhD  <https://orcid.org/0000-0002-6228-7241>
 Jason D Pole, BSc(Hons), MSc, PhD  <https://orcid.org/0000-0002-0413-5434>
 Bernadette Richards, BA, Dip Ed, LLB(Hons), PhD  <https://orcid.org/0000-0001-6448-4954>
 Quita Olsen, MEpi  <https://orcid.org/0009-0007-1451-8408>
 Amalie Dyda, BHLthSc(Hons), MAE, PhD  <https://orcid.org/0000-0003-2806-4834>
 Nalini Pather, BMedSc(Hons), MMedSci., GCULT, Adv Dip(Ethics), PhD  <https://orcid.org/0000-0001-5288-7713>
 Leanna Woods, BN(Hons), PhD  <https://orcid.org/0000-0003-4811-4608>

Data availability statement

All data utilised in this study is available on request from the authors.

Note

1. <https://makrahe.kumu.io/a-scientometric-review-of-health-data-sharing-for-secondary-use-insights-frontiers-and-the-path-ahead>.

References

- Adams C, Allen J and Flack F (2022) *Sharing Linked Data for Health Research: Toward Better Decision Making*. Cambridge: Cambridge University Press.
- Aggarwal R, Farag S, Martin G, et al. (2021) Patient perceptions on data sharing and applying artificial intelligence to health care data: Cross-sectional survey. *Journal of Medical Internet Research* 23(8): e26162.
- Alam P, Bolio A, Lin L, et al. (2024) Stakeholders' perceptions of personal health data sharing: A scoping review. *PLOS Digital Health* 3(11): e0000652.
- Amorim M, Silva S, Machado H, et al. (2022) Benefits and risks of sharing genomic data for research: Comparing the views

- of rare disease patients, informal carers and healthcare professionals. *International Journal of Environmental Research and Public Health* 19(14): 8788.
- Baines R, Stevens S, Austin D, et al. (2024) Patient and public Willingness to share personal health data for third-party or secondary uses: Systematic review. *Journal of Medical Internet Research* 26: e50421.
- Bakken V, Koposov R, Rost TB, et al. (2022) Attitudes of mental health service users toward storage and use of electronic health records. *Psychiatric Services* 73(9): 1013–1018.
- Bastian M, Heymann S and Jacomy M (2009) *Gephi: An open-source software for exploring and manipulating networks* (version 0.10) [computer software]. Gephi Consortium. <https://gephi.org>
- Benevento M, Mandarelli G, Carravetta F, et al. (2023) Measuring the willingness to share personal health information: A systematic review. *Frontiers in Public Health* 11: 1213615.
- Birkstedt T, Minkinen M, Tandon A, et al. (2023) AI governance: Themes, knowledge gaps and future agendas. *Internet Research* 33(7): 133–167.
- Brall C, Berlin C, Zwahlen M, et al. (2021) Public willingness to participate in personalized health research and biobanking: A large-scale Swiss survey. *PLoS One* 16(4): e0249141.
- Braunack-Mayer A, Fabrianesi B, Street J, et al. (2021) Sharing government health data with the private sector: Community attitudes survey. *Journal of Medical Internet Research* 23(10): e24200.
- Braunack-Mayer AJ, Adams C, Nettel-Aguirre A, et al. (2024) Community views on the secondary use of general practice data: Findings from a mixed-methods study. *Health Expectations* 27(1): e13984.
- Cascini F, Pantovic A, Al-Ajlouni YA, et al. (2024) Health data sharing attitudes towards primary and secondary use of data: A systematic review. *eClinicalMedicine* 71: 102551.
- Chan EY and Saqib NU (2021) Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high. *Computers in Human Behavior* 119: 106718.
- Cobo MJ, López-Herrera AG, Herrera-Viedma E, et al. (2011) An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics* 5(1): 146–166.
- Cumyn A, Ménard JF, Barton A, et al. (2023) Patients' and members of the public's wishes regarding transparency in the context of secondary use of health data: Scoping review. *Journal of Medical Internet Research* 25: e45002.
- Du Q, Zhao R, Wan Q, et al. (2024) Protocol for conducting bibliometric analysis in biomedicine and related research using CiteSpace and VOSviewer software. *STAR Protocols* 5(3): 103269.
- Gerke S, Shachar C, Chai PR, et al. (2020) Regulatory, safety, and privacy concerns of home monitoring technologies during COVID-19. *Nature Medicine* 26(8): 1176–1182.
- Gross MS, Hood AJ, Rubin JC, et al. (2022) Respect, justice and learning are limited when patients are deidentified data subjects. *Learning Health Systems* 6(3): e10303.
- Hutchings E, Loomes M, Butow P, et al. (2020) A systematic literature review of health consumer attitudes towards secondary use and sharing of health administrative and clinical trial data: A focus on privacy, trust, and transparency. *Systematic Reviews* 9(1): 235.
- Hvalič-Touzery S, Laznik J and Petrovčič A (2024) "I'm still struggling with it, and it scares me": A qualitative analysis of older adults' experiences with digital health portals during and after COVID-19. *Digital Health* 10: 20552076241282247.
- Kaplow K, Downey M, Stewart D, et al. (2024) Data professionals' attitudes on data privacy, sharing, and consent in healthcare and research. *Digital Health* 10: 20552076241290964.
- Kerasidou A and Kerasidou C (2023) Data-driven research and healthcare: Public trust, data governance and the NHS. *BMC Medical Ethics* 24(1): 51.
- Kim J, Kim H, Bell E, et al. (2019) Patient perspectives about decisions to share medical data and biospecimens for research. *JAMA Network Open* 2(8): e199550.
- Kirkham EJ, Lawrie SM, Crompton CJ, et al. (2022) Experience of clinical services shapes attitudes to mental health data sharing: Findings from a UK-wide survey. *BMC Public Health* 22(1): 357.
- Lee AR, Koo D, Kim IK, et al. (2023) Identifying facilitators of and barriers to the adoption of dynamic consent in digital health ecosystems: A scoping review. *BMC Medical Ethics* 24(1): 107.
- Li Y (2022) Cross-cultural privacy differences. In: Knijnenberg BP, Page X and Wisniewski P, et al., (eds) *Modern Socio-Technical Perspectives*. Cham: Springer, pp.267–292.
- Muller SHA, Kalkman S, van Thiel GJM, et al. (2021) The social licence for data-intensive health research: Towards co-creation, public value and trust. *BMC Medical Ethics* 22(1): 110.
- Murdoch TB and Detsky AS (2013) The inevitable application of big data to health care. *JAMA* 309(13): 1351–1352.
- Naeem I, Quan HD, Singh S, et al. (2022) Factors associated with willingness to share health information: Rapid review. *JMIR Human Factors* 9(1): e20702.
- Olsen Q, Dyda A, Woods L, et al. (2025) Worldwide willingness to share health data high but privacy, consent and transparency paramount, a meta-analysis. *npj Digital Medicine* 8(1): 540.
- Richter JG, Filla T, Acar H, et al. (2023) Sustained agreement rates in the longitudinal assessment of lupus patients to a Broad Consent for personal data and specimen usage in the RHINEVIT biobank. *Frontiers in Medicine* 10: 1208006.
- Rieke N, Hancox J, Li W, et al. (2020) The future of digital health with federated learning. *npj Digital Medicine* 3: 119.
- Sánchez MC, Hernández Clemente JC and García López FJ (2023) Public and patients' perspectives towards data and sample sharing for research: An overview of empirical findings. *Journal of Empirical Research on Human Research Ethics* 18(5): 319–345.
- Saxena S, Mishra SC and Mukerji B (2024) A multi-method bibliometric review of value co-creation research. *Management Research Review* 47(2): 183–203.
- Seltzer E, Goldshear J, Guntuku SC, et al. (2019) Patients' willingness to share digital health and non-health data for research: A cross-sectional study. *BMC Medical Informatics and Decision Making* 19(1): 157.
- Soni H, Grando A, Aliste MP, et al. (2019) Perceptions and preferences about granular data sharing and privacy of behavioral health patients. *Studies in Health Technology and Informatics* 264: 1361–1365.
- Street J, Fabrianesi B, Adams C, et al. (2021) Sharing administrative health data with private industry: A report on two citizens' juries. *Health Expectations* 24(4): 1337–1348.
- Sullivan C, Wong I, Adams E, et al. (2021) Moving faster than the COVID-19 pandemic: The rapid, digital transformation of a public health system. *Applied Clinical Informatics* 12(2): 229–236.
- Texier VL, Henda N, Cox S, et al. (2019) Data sharing in the era of precision medicine: A scientometric analysis. *Precision Cancer Medicine* 2: 30.

- Tosoni S, Voruganti I, Lajkosz K, et al. (2022) Patient consent preferences on sharing personal health information during the COVID-19 pandemic: “the more informed we are, the more likely we are to help”. *BMC Med Ethics* 23(1): 53.
- van Eck NJ and Waltman L (2010) *VOSviewer* (version 1.6.7) [Computer software]. Leiden University, The Netherlands. <https://www.vosviewer.com>
- van Eck NJ and Waltman L (2014) Visualizing bibliometric networks. In: Ding Y, Rousseau R and Wolfram D (eds) *Measuring Scholarly Impact: Methods and Practice*. Cham: Springer International Publishing, pp.285–320.
- Warren V, Critchley C, McWhirter R, et al. (2023) Context matters in genomic data sharing: A qualitative investigation into responses from the Australian public. *BMC Medical Genomics* 15(Suppl. 3): 275.
- Watson E, Fletcher-Watson S and Kirkham EJ (2023) Views on sharing mental health data for research purposes: Qualitative analysis of interviews with people with mental illness. *BMC Medical Ethics* 24(1): 99.
- Zawistowski M, Fritsche LG, Pandit A, et al. (2023) The Michigan Genomics Initiative: A biobank linking genotypes and electronic clinical records in Michigan Medicine patients. *Cell Genomics* 3(2): 100257.