Jul, 2025
GRINS DISCUSSION PAPER SERIES DP N° 22/2025 ISSN 3035-5576

GRINS

Optimizing Population Health Through Strategic Use of Health Data

DP N° 22/2025

Authors:
Vincenzo Atella, Andrea Ganna, Stefano Lombardi

Jul, 2025
GRINS DISCUSSION PAPER SERIES DP N° 22/2025 **ISSN 3035-5576**

Optimizing Population Health Through Strategic Use of Health Data

Vincenzo Atella, Andrea Ganna, Stefano Lombardi

KEYWORDS

Health data | Privacy | EHDS | Digitalization | Research | Italy

ACKNOWLEDGEMENTS

This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

CITE THIS WORK

Author(s): Vincenzo Atella, Andrea Ganna, Stefano Lombardi. Title: Optimizing Population Health Through Strategic Use of Health Data. Publication Date: 2025.

The article explores the transformative potential of health data in improving population health, highlighting its applications in prevention, personalized medicine, care quality, crisis management, and equity. It emphasizes how digital technologies and data integration can enhance clinical decision-making, reduce costs, and drive systemic efficiency. Despite these benefits, strict privacy regulations—particularly in Italy—often hinder data reuse for research, slowing innovation and limiting the impact of public health policies. The authors examine the legal complexities surrounding the GDPR and national legislation, calling for more harmonized and pragmatic frameworks. Case studies, such as the Finnish and Danish models, demonstrate how data access can coexist with robust privacy protection. The chapter also introduces synthetic data and secure data environments as promising solutions to circumvent bureaucratic constraints while preserving privacy. It concludes with a call for centralized coordination, infrastructure development (like the EHDS), and improved data linkage to overcome the persistent “data gap” that impedes the measurement of important health phenomena. Ultimately, the work argues that balancing ethical data use with accessibility is crucial for enabling evidence-based, equitable, and innovative healthcare across Europe.
Keywords: Health data, Privacy, EHDS, Digitalization, Research, Italy.
JEL Code: I18, K29, K39.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
GRINS FOUNDATION

GRINS FOUNDATION

Discussion Paper Series

Optimizing Population Health Through Strategic Use of Health Data

Discussion paper n. 22/2025

Vincenzo Atella, Andrea Ganna and Stefano Lombardi

ISSN 3035-5567

Optimizing Population Health Through Strategic Use of Health Data

Vincenzo Atella, Andrea Ganna and Stefano Lombardi

Discussion Paper DP nº 22/2025

The article explores the transformative potential of health data in improving population health, highlighting its applications in prevention, personalized medicine, care quality, crisis management, and equity. It emphasizes how digital technologies and data integration can enhance clinical decision-making, reduce costs, and drive systemic efficiency. Despite these benefits, strict privacy regulations—particularly in Italy—often hinder data reuse for research, slowing innovation and limiting the impact of public health policies. The authors examine the legal complexities surrounding the GDPR and national legislation, calling for more harmonized and pragmatic frameworks. Case studies, such as the Finnish and Danish models, demonstrate how data access can coexist with robust privacy protection. The chapter also introduces synthetic data and secure data environments as promising solutions to circumvent bureaucratic constraints while preserving privacy. It concludes with a call for centralized coordination, infrastructure development (like the EHDS), and improved data linkage to overcome the persistent "data gap" that impedes the measurement of important health phenomena. Ultimately, the work argues that balancing ethical data use with accessibility is crucial for enabling evidence-based, equitable, and innovative healthcare across Europe.

Keywords: Health data, Privacy, EHDS, Digitalization, Research, Italy.

JEL Code: I18, K29, K39.

This study was funded by the European Union - NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS – Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP D13C22002160001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Optimizing Population Health Through Strategic Use of Health Data

Authors:

Vincenzo Atella, Andrea Ganna and Stefano Lombardi[1]

Abstract

The article explores the transformative potential of health data in improving population health, highlighting its applications in prevention, personalized medicine, care quality, crisis management, and equity. It emphasizes how digital technologies and data integration can enhance clinical decision-making, reduce costs, and drive systemic efficiency. Despite these benefits, strict privacy regulations—particularly in Italy—often hinder data reuse for research, slowing innovation and limiting the impact of public health policies. The authors examine the legal complexities surrounding the GDPR and national legislation, calling for more harmonized and pragmatic frameworks. Case studies, such as the Finnish and Danish models, demonstrate how data access can coexist with robust privacy protection. The chapter also introduces synthetic data and secure data environments as promising solutions to circumvent bureaucratic constraints while preserving privacy. It concludes with a call for centralized coordination, infrastructure development (like the EHDS), and improved data linkage to overcome the persistent "data gap" that impedes the measurement of important health phenomena. Ultimately, the work argues that balancing ethical data use with accessibility is crucial for enabling evidence-based, equitable, and innovative healthcare across Europe.

Keywords: Health data, Privacy, EHDS, Digitalization, Research, Italy.
JEL Code: I18, K29, K39.

[1] Corresponding author: stefano.lombardi@vatt.fi

Index

"The utility of health information, research evidence and knowledge (collectively described as knowledge) is to better inform and thus empower individuals and the public to make the right decisions regarding their health and well-being; influence public health policy and decision making; advance the frontiers of knowledge to develop products and tools for the promotion, maintenance, protection and restoration of health."

The Commission on Health Research for Development

Introduction[2]

In the digital age, healthcare systems generate vast amounts of data every day—from electronic health records (EHRs) to genetic information, clinical studies, and public health databases. When effectively harnessed, this wealth of health data represents a remarkable opportunity to revolutionize healthcare delivery and improve population health outcomes. By enabling researchers to study health trends, assess treatment effectiveness, and identify patterns of disease spread, the proper use of health data can enhance not only individual patient care but also drive systemic improvements across entire health services.

The production, utilization, and strategic governance of health data have emerged as the true transformative forces in contemporary healthcare systems. Far beyond the promise of digital tools themselves, it is the generation, integration, and intelligent use of data that enables systems to become more sustainable, equitable, and responsive. Health data—whether clinical, behavioral, environmental, or genomic—forms the foundation upon which evidence-based decisions can be made, resources can be allocated more efficiently, and interventions can be tailored to population needs. Digital technologies, while instrumental, primarily function as data-generation engines, creating continuous streams of information through tools like electronic health records (EHRs), wearable devices, mobile health platforms, and telemedicine interfaces. Their value lies not solely in their functionality but in their capacity to produce structured, interoperable, and timely data that can be analyzed and mobilized to guide policy and clinical decisions (Topol, 2019; WHO, 2023).

When properly harnessed, this wealth of health data supports early-warning systems, predictive analytics, and performance evaluation, thereby improving surveillance and preparedness, especially in times of health crises (OECD, 2021). Moreover, data-driven insights help address health inequities by identifying underserved populations and monitoring disparities in outcomes and access (Mehrotra et al., 2020).

The personalization of care is also fundamentally a data-driven achievement. By integrating diverse data sources—from genomics to lifestyle metrics—health systems can shift toward precision medicine, offering more targeted and effective treatment options (Torous & Roberts, 2017). Similarly, the use of

[2] This study was funded by the European Union - NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS – Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP D13C22002160001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

data analytics allows for cost control by avoiding redundant procedures, optimizing workflows, and supporting preventive strategies. Operational improvements attributed to digital technologies—such as streamlined documentation or remote consultations—are, at their core, outcomes of more effective data management. For instance, interoperable EHRs and virtual care platforms have enabled up to 15% gains in efficiency by ensuring that relevant patient data is available when and where it is needed (McKinsey & Company, 2021).

There are at least three main reasons suggesting that we have reached a sufficiently mature moment to unlock this value:

1. We are now more capable than ever of producing health data that is ready for use.
2. Innovation is driven by the need to deliver more healthcare services that are better than before and to prepare for future crises with fewer resources, while simultaneously addressing long-standing inequalities.
3. The experiences during the COVID-19 pandemic have shown that digitalization can make a real difference.

Despite these achievements, we still face important problems in transforming this information into knowledge and actions. This is what goes under the name of the "know-do gap" and, in our specific case, refers to the persistent failure to translate available knowledge into actionable strategies that improve health outcomes. This is related to the challenge of sharing and translating health information, research evidence, or knowledge. In today's digital era, this gap is increasingly driven not by a lack of information, but by the inadequate exploitation of the vast amount of data already collected across health systems. Despite the potential of electronic health records, administrative databases, real-world evidence, and digital monitoring tools, these rich data sources often remain underused due to stringent privacy regulations, legal uncertainties, and fragmented governance frameworks (Vayena et al., 2018; Mittelstadt, 2019).

The paradox is that while health systems are data-rich, they are insight-poor. Much of the untapped value lies in the inability to access, link, and analyze datasets across institutions and jurisdictions. Privacy regulations like the General Data Protection Regulation (GDPR), while essential for safeguarding individual rights, can become an obstacle when implemented without sufficient flexibility for research and public health purposes (Shabani et al., 2018). Bridging this gap requires more than technical solutions—it calls for a long-term commitment to strengthen data governance frameworks, build cross-sectoral interoperability, and promote the responsible reuse of data. Strategic coordination between data custodians, researchers, and policymakers is essential to convert health data into meaningful, evidence-based action. Only by overcoming these barriers can we move toward a truly learning and adaptive health system capable of addressing complex and evolving challenges. A close linkage and coordination between fragmented domains such as information systems, health research, and knowledge management is viewed as an essential step in this process.

Ultimately, health data—its production, curation, sharing, and reuse—must be recognized as the backbone of modern healthcare systems. While digital tools are instrumental in capturing and managing this data, it is the strategic governance and intelligent application of data that drives genuine innovation and value. Health systems that are "data mature"—meaning they possess the capabilities to collect high-quality data, ensure interoperability, apply analytics, and support data-informed

decision-making—consistently outperform others in efficiency, responsiveness, and health outcomes (Kelley et al., 2019; OECD, 2021). Moreover, the reuse of health data for secondary purposes, such as public health surveillance, policy evaluation, and research, can generate significant returns by accelerating scientific discovery and informing population-level interventions (Rothstein, 2015). For this potential to be fully realized, there must be not only investments in infrastructure but also robust frameworks for ethical data sharing, privacy protection, and cross-sector collaboration (Vayena et al., 2018; Knoppers & Joly, 2022). In this view, health data becomes a strategic asset, essential not just for treating patients, but for building resilient, equitable, and learning health systems that continuously adapt and improve over time (Friedman et al., 2017).

1. Why the Use of Health Data Is Important in Today's Healthcare

As mentioned in the introduction section, the use of health data has become a cornerstone of modern healthcare systems, revolutionizing how medical decisions are made, how resources are allocated, and how public health challenges are addressed (Appari et al., 2010). In an era shaped by demographic shifts—most notably aging populations—and the rising incidence of chronic diseases, the capacity to collect, analyze, and apply health data is increasingly critical for the sustainability, quality, and efficiency of healthcare delivery. Technological advancements, particularly in artificial intelligence (AI) and big data analytics, have significantly expanded the capabilities of health data, offering transformative potential in personalized medicine, predictive modeling, and evidence-based intervention strategies (Rehman, Naz, & Razzak, 2020). The integration of electronic health records (EHRs), genomic data, and real-time monitoring devices not only strengthens individual clinical care but also fuels large-scale research initiatives targeting multifactorial diseases (Longitools Consortium, n.d.).

Along these lines, Bailey, Currie, and Schwandt (2024) highlight that using individual-level data not only deepen our understanding of health trajectories but also illuminate long-term economic outcomes such as labor market participation, income dynamics, and intergenerational mobility. Their work demonstrates that early-life health shocks and neighborhood-level exposures have measurable consequences on adult economic productivity and social welfare dependence. This evidence underscores the value of linking health and socioeconomic data longitudinally, but it also suggests the untapped potential of real-time data systems. If policymakers could access and analyze individual data streams as events unfold—such as school absences, hospital admissions, or localized environmental stressors—they might anticipate and mitigate negative economic trajectories with targeted, time-sensitive interventions. In this sense, real-time data integration would not only support personalized medicine but also enable dynamic and more equitable health and economic policymaking.

Beyond individual patient outcomes, health data serves as a strategic asset in advancing public health, especially in disease prevention and emergency preparedness. The COVID-19 pandemic made clear the indispensable role of real-time data in tracking disease progression and informing dynamic policy decisions. The work by Chetty et al. (2020) illustrates how real-time, high-frequency individual data, drawn from private sector sources such as credit card processors and payroll firms, can be used to uncover the immediate economic consequences of public health crises. Their analysis of the COVID-19 pandemic revealed how consumer spending, employment, and small business revenue reacted sharply and unevenly to the spread of the virus and associated public health interventions—well before traditional statistics could capture these shifts. This approach highlights the transformative potential of

real-time data not only for understanding economic outcomes but also for informing health policy decisions with economic foresight. If real-time data infrastructures were more tightly integrated with public health systems, they could support more agile, spatially targeted interventions—balancing economic risk and health protection. In this way, individual-level data become a critical lever not only for epidemiological monitoring, but also for anticipating and managing broader socioeconomic disruptions, ultimately enabling more responsive and equitable health and economic policy. Within the framework of Learning Health Systems (LHS), health data is continually leveraged to refine medical practices and elevate patient outcomes through iterative feedback loops (McLachlan, Dube, & Gallagher, 2018). However, while the benefits of health data are substantial, they are accompanied by complex ethical, legal, and social challenges. Concerns around data privacy, ownership, consent, and potential misuse must be carefully navigated to maintain public trust and ensure equitable use (BaHammam, 2023). Thus, the path toward effective integration of health data into contemporary healthcare systems requires not only technical innovation but also robust governance frameworks that prioritize ethical responsibility.

This section examines the multifaceted role of health data in shaping healthcare today, emphasizing its capacity to enhance decision-making, optimize resources, and promote health equity, while acknowledging the risks that must be managed in its application.

1.1 Early Detection and Prevention

Health data plays a fundamental role in the early detection and prevention of disease, serving as a critical tool for reducing morbidity, improving clinical outcomes, and minimizing healthcare costs. The capacity to collect and analyze large-scale health datasets, including electronic health records (EHRs), biometric monitoring, and population-level databases, enables researchers and clinicians to identify subtle physiological and behavioral changes that may signal the onset of chronic diseases such as diabetes, cardiovascular disease, and certain cancers (Topol, 2019; Chen et al., 2022). For example, data from wearable devices—which monitor metrics such as heart rate variability, sleep quality, blood pressure, and physical activity—are increasingly being used to flag early warning signs in real-time. These insights allow for proactive interventions, helping clinicians deliver preventive care before disease symptoms escalate, particularly in high-risk populations (Steinhubl et al., 2015; McKinsey & Company, 2021).

Beyond individual-level prevention, health data enables researchers to analyze risk factors across diverse populations, facilitating a deeper understanding of social, environmental, and genetic determinants of health. These insights inform public health strategies, such as targeted screening programs, behavioral interventions, and resource allocation, which can mitigate risks and promote healthier lifestyles on a community-wide scale (Khoury et al., 2018; Marmot et al., 2020). Thus, by harnessing the predictive and diagnostic power of data, health systems can transition from reactive to preventive care models, ultimately enhancing population health and equity.

1.2 Personalized and Precision Medicine

One of the most promising and transformative applications of health data lies in the domain of personalized and precision medicine. This approach utilizes detailed, individual-level data—including genomic profiles, clinical histories, environmental exposures, and behavioral factors—to tailor diagnostic, preventive, and therapeutic strategies to the unique characteristics of each patient. The

integration of such heterogeneous data enables healthcare providers to move beyond the "one-size-fits-all" model and deliver care that is both more targeted and more effective (Collins & Varmus, 2015; Torkamani et al., 2018).

In oncology, for instance, treatments are increasingly based on the molecular and genetic makeup of specific tumors. This has led to significant improvements in treatment efficacy and a reduction in adverse effects, as therapies are aligned with the biological behavior of the cancer in a given patient (Ashley, 2016). Moreover, by reducing the reliance on trial-and-error prescribing and minimizing ineffective interventions, precision medicine can contribute to cost savings and more efficient resource utilization across health systems (Ginsburg & Phillips, 2018). As health data becomes more comprehensive and interoperable, the potential for real-time learning and adaptive treatment refinement will only increase. This data-driven evolution stands to redefine not only how diseases are treated, but also how they are prevented and understood at a population level.

1.3 Improving the Quality of Care

Health data provides valuable insights into the quality and effectiveness of healthcare services. By systematically analyzing data on clinical outcomes, hospital admissions, treatment efficacy, and patient satisfaction, healthcare providers and researchers can identify inefficiencies, uncover patterns of suboptimal care, and implement targeted improvements (Kruk et al., 2018; Dixon et al., 2016). This evidence-based approach enables the continuous monitoring of health system performance, allowing for benchmarking against standards and peers, and supporting the shift toward more accountable and transparent healthcare delivery (Berwick et al., 2008).

One example of data-driven quality improvement is the analysis of hospital readmission rates. By identifying the underlying causes of avoidable readmissions, health systems can implement preventive strategies, improve discharge planning, and enhance continuity of care (Jencks et al., 2009). Health data can also expose disparities in access, outcomes, and service utilization across different demographic groups, helping systems address inequities and ensure that all populations receive high-quality care (Artiga et al., 2020). Importantly, the use of health data supports a transition from volume-based to value-based care, where performance is judged by health outcomes rather than service quantity. This model promotes more efficient use of resources, better patient experiences, and improved population health (Porter, 2010).

1.4 Predicting and Managing Public Health Crises

The COVID-19 pandemic demonstrated the vital role of real-time health data in managing public health emergencies. Access to up-to-date information on infection trends, hospital capacity, and vaccine deployment enabled authorities to make evidence-based decisions on containment strategies, resource distribution, and vaccination campaigns (Scudellari, 2020; Salathé et al., 2020). Data-driven models were instrumental in slowing the virus's spread and preventing healthcare system collapse in many regions.

Looking ahead, health data will remain essential for anticipating and responding to future crises. Predictive analytics based on historical and real-time data can identify patterns in disease transmission, supporting early interventions and more accurate epidemiological forecasting (Chinazzi et al., 2020). Global integration of health data, paired with early-warning systems, could facilitate rapid detection of

emerging pathogens, giving governments critical lead time to implement containment measures (Kraemer et al., 2020).

1.5 Promoting Health Equity

Health data is a key resource for detecting and addressing health disparities. By integrating information on social determinants—such as income, race, education, and geography—with clinical outcomes, researchers can uncover the structural and environmental drivers of health inequities (Bailey et al., 2017). For instance, data analysis may reveal that asthma prevalence is higher in low-income neighborhoods due to poor air quality, guiding targeted interventions like pollution control and enhanced access to care (Artiga & Orgera, 2019).

Using this knowledge, policymakers and healthcare providers can design equitable health interventions, such as expanding services in underserved areas and tailoring prevention programs to the needs of vulnerable communities (Marmot et al., 2020). The responsible use of such data ensures that interventions are inclusive and aligned with diverse population needs, promoting fairness and improving population health outcomes (Braveman et al., 2011).

1.6 Advancing Research and Innovation

Health data is an invaluable asset for accelerating medical research and driving innovation. Large, high-quality datasets enable studies with broader scope and greater statistical power, supporting discoveries in diagnostics, therapeutics, and public health strategies (Lo & DeMets, 2016). Merging real-world data with clinical trial results enhances external validity and enables the evaluation of treatment performance in routine settings (Sherman et al., 2016).

In particular, the rise of artificial intelligence (AI) and machine learning has expanded the potential of health data. These technologies can process vast datasets rapidly, identifying complex associations that may elude traditional analysis. AI is increasingly used for drug discovery, patient risk stratification, and disease prediction, enhancing the precision and speed of innovation (Esteva et al., 2019; Topol, 2019). By unlocking new insights, data-enabled research is reshaping how health challenges are addressed.

1.7 Challenges and Ethical Considerations

Despite these benefits, the use of health data raises significant ethical and privacy concerns. Protecting patient confidentiality, securing informed consent, and ensuring transparency in data collection and use are foundational to public trust (Floridi & Taddeo, 2016). Strong data governance, guided by principles of fairness, accountability, and transparency, is essential to safeguard against misuse and ensure responsible innovation (Vayena et al., 2018).

Moreover, balancing data accessibility with privacy protection remains a core challenge. Policies must enable data sharing for societal benefit while preserving individual rights. Ethical frameworks and technical safeguards—such as anonymization, data minimization, and secure data environments—can help strike this balance (Shabani et al., 2018). Ultimately, advancing data-driven healthcare requires an ongoing commitment to both innovation and ethical responsibility.

2. Protecting Privacy, Delaying Progress? Health Data Regulation and Its Effects on Population Well-being

Health data, when effectively used, holds the potential to revolutionize healthcare by enhancing disease prevention, diagnosis, and treatment. However, the use of personal health information entails inherent risks related to privacy and security. To mitigate these risks, numerous countries have implemented strict privacy regulations. While these laws are essential to safeguarding individual rights and maintaining public trust, they can simultaneously restrict researchers' access to critical health data, potentially impacting public health outcomes (Vayena & Blasimme, 2017; Mittelstadt, 2019).

2.1 The Importance of Privacy Regulations

Privacy regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union are designed to ensure that personal health data is processed with the highest ethical standards. These frameworks protect patients from data breaches, unauthorized access, and misuse, encouraging individuals to share their health information more openly, thereby improving the quality of care and research participation (Gostin et al., 2019). Moreover, transparency provisions in these regulations empower patients by informing them about data usage and granting them control over their personal information (Shabani & Marelli, 2019).

2.2 Challenges for Researchers

Although privacy regulations are crucial for protecting patients, they also pose significant challenges for researchers who rely on access to health data to conduct studies that benefit public health. These laws often restrict access to large datasets or impose stringent requirements for anonymization and de-identification, which can slow down research or prevent certain studies altogether (Shabani et al., 2018). Researchers may be required to undergo lengthy approval processes or obtain explicit consent from individuals before accessing their health data. This can delay study initiation and reduce sample sizes, making it difficult to draw statistically significant conclusions. In cases where consent is impractical—such as retrospective studies or research involving large populations—researchers may be denied access to vital data (Porsdam Mann et al., 2016). Furthermore, while de-identification is necessary to protect privacy, it can reduce data utility by removing crucial contextual details such as age, location, or socioeconomic background, which are essential for studying health disparities and social determinants of health (Mello et al., 2018).

2.3 Impact on Population Health

Restrictions imposed by privacy regulations can have a direct impact on population health. When researchers lack access to comprehensive datasets, they may miss critical information on disease trends, risk factors, and treatment effectiveness. This can delay the development of new therapies or public health interventions, ultimately affecting health outcomes on a broad scale (Dove et al., 2017). The COVID-19 pandemic underscored the importance of rapid data sharing for crisis management; countries with more flexible data governance frameworks responded more swiftly and effectively (Salathé et al., 2020). In countries where privacy laws were more restrictive, data-sharing efforts were sometimes hampered, slowing the response and endangering lives. Moreover, restricted data access hinders the ability to identify and address health disparities, exacerbating existing inequities and limiting targeted policy responses (Bailey et al., 2017).

2.4 Balancing Privacy and Research Needs

The challenge for policymakers and healthcare organizations is to strike a balance between protecting individual privacy and enabling the use of health data for research that benefits public health. Several strategies can help achieve this equilibrium. First, ensuring that data is properly anonymized or pseudonymized can help safeguard privacy while preserving research utility (Shabani & Marelli, 2019). However, it is essential that the de-identification process does not eliminate information critical to meaningful analysis. Second, developing clearer consent frameworks—such as broad consent models that allow individuals to authorize the use of their data for multiple research purposes under secure conditions—can streamline data access for researchers (Grady et al., 2015). Third, secure data-sharing platforms that meet regulatory requirements can facilitate research without compromising patient privacy. These platforms should include encryption, access controls, and audit mechanisms to ensure that data is used only for authorized purposes (Knoppers, 2014). Fourth, engaging and educating the public about the benefits of health data research and how their privacy is protected can build trust and encourage participation. Transparency about data use and the measures taken to protect it reassures individuals that their information is handled responsibly (Vayena et al., 2015).

As we will see, these strategies, while ambitious, are not merely theoretical: several Nordic countries have already operationalized this balance through well-established frameworks that combine robust data protection with high-quality, ethically governed data access for research—demonstrating that privacy and responsible data openness can, in fact, successfully coexist in practice.

2.5 The Future of Health Data and Privacy Regulations

As healthcare becomes increasingly data-driven, privacy regulations must evolve to accommodate emerging technologies like artificial intelligence and machine learning, which depend on access to large, high-quality datasets (Topol, 2019). Future privacy frameworks must enable the ethical and secure use of such technologies in health research. Emerging technologies like blockchain may also offer new ways to create more secure and transparent systems for managing health data (Agbo et al., 2019). By giving patients control over their own data and allowing them to selectively share it with researchers or providers, these systems could enhance privacy while enabling valuable research. Ultimately, while protecting privacy remains a fundamental obligation, rigid regulatory barriers must be addressed to avoid stifling scientific discovery and delaying health innovations that could benefit society at large (Floridi, 2020). By embracing secure data-sharing solutions, reforming consent models, and fostering public dialogue, healthcare systems can achieve a sustainable balance between safeguarding rights and advancing medical science.

3. The Legislation on Health Data in Europe: Does Privacy Regulations Hinder Their Reuse?

The conduct of biomedical, clinical, epidemiological, and digital health research in Europe is profoundly influenced by the GDPR legislation. The GDPR provides specific provisions for processing sensitive health data for research purposes, such as the "research exemption" under Article 9(2)(j), but its application is subject to interpretation by national and regional authorities, resulting in substantial fragmentation across member states (Péloquin et al., 2020; Scheibner et al., 2020). This divergence

manifests in variable consent requirements, differing standards for pseudonymisation and anonymisation, and a lack of harmonized guidance for secondary data use, particularly impacting multicenter, cross-border, and registry-based studies (Péloquin et al., 2020; Scheibner et al., 2020; Doetsch et al., 2021; van der Wel et al., 2019).

Empirical studies provide evidence of these regulatory barriers. Surveys and interviews with clinical research stakeholders across the EU highlight persistent legal uncertainty, administrative delays, and operational burdens arising from the interplay between GDPR, the Clinical Trials Regulation, and national laws—obstacles often unrelated to crisis-specific measures such as those introduced during the COVID-19 pandemic (Lalova-Spinks et al., 2022; Bak et al., 2023; Richards, 2022). In Finland, for example, the implementation of stricter data access laws was associated with a 47% reduction in new registry data permits, illustrating the measurable negative impact of privacy regulations on research capacity (Brück et al., 2024).[3] Comparative legal analyses show that countries such as Portugal, Finland, Norway, and the Netherlands impose widely varying requirements for the linkage of cohort and routine health data, leading to administrative delays from as little as seven days to as much as 300 days (Doetsch et al., 2021). Similarly, in Italy, restrictive and heterogeneous interpretations by data protection authorities and ethics committees create uncertainty and have led to the suspension of epidemiological projects, placing Italian researchers at a competitive disadvantage in European collaborative efforts (Cagnazzo et al., 2023; Bisceglia et al., 2023).

The increasing focus on cross-border collaboration and data sharing, including international transfers to non-EU countries such as the United States, has highlighted additional legal and operational challenges. The uncertainty surrounding adequacy decisions—especially after developments like Schrems II—and the complexity of applying Standard Contractual Clauses continue to impede transatlantic research (Bradford et al., 2020; Lalova-Spinks et al., 2024; Molnár-Gábor & Korbel, 2020). The rise of new EU acts and proposals, such as the Data Governance Act (DGA) and the European Health Data Space (EHDS), further complicates the landscape. While initiatives like data altruism and centralized data access bodies are intended to promote harmonization and data sharing, legal analyses and stakeholder interviews suggest they may add new layers of uncertainty unless carefully coordinated and clarified (Lalova-Spinks et al., 2023; Slokenberga, 2022; Rak, 2024).

Overall, the literature demonstrates a persistent disconnect between the theoretical flexibilities available under the GDPR and the fragmented, often burdensome reality faced by researchers in practice (Péloquin et al., 2020; Lalova-Spinks et al., 2022; Doetsch et al., 2021). Below we report in synthesis the Key Mechanisms and Barriers Identified for a smooth utilization of the data:

- Legal Fragmentation and Divergence EU-level rules (GDPR, sectoral directives). These rules allow broad "research exemptions" but are inconsistently interpreted and implemented at national and even subnational levels, creating practical uncertainty and administrative burden

[3] It is worth mentioning that the analysis presented by Brück et al. (2024) may be misleading, as it appears to suggest that the introduction of FinData led to a reduction in access. However, their study focuses solely on the immediate years following the implementation of the service, without considering longer-term trends. Moreover, their comparison lacks robustness, as it conflates hospital permits with other types of permits, thereby limiting the validity of their conclusions. Following interaction with FinData, we got access to the number of applications and permits from 2021 to 2025, and the downward trend discussed in Brück et al. (2024) did not show up.

for researchers.[4] Italy stands out as especially restrictive, with heterogeneous and sometimes subjective interpretations by regional authorities and data protection offices, hampering observational and epidemiological studies.

- Consent and Secondary Use Challenges. Despite GDPR art. 9(2)(j) and "broad consent" notions, many Member States—including Italy—still require narrow, project-specific consent, re-consenting, or new ethics approvals for uses of existing data, particularly for registries and biobanks. The lack of harmonized guidance leaves researchers exposed to variable criteria on when consent can be waived or what technical/organizational safeguards suffice (Smit et Al., 2023a,b).
- Data Access and Administrative Delays. Substantial delays (ranging from weeks to years), repetitive ethics/data-access reviews, and increased costs are empirically documented in multi-country comparisons, especially for record linkage and registry-based studies (van der Wel (2019), Doetsch et Al. (2021)). Centralized data authorities (e.g. Finland) can ease linkage, but new privacy acts have sometimes reduced data access rather than expanded it (van der Wel (2019)).
- Cross-Border & International Data-Sharing Barriers. Both intra-EU and transatlantic (EU–US) research are hampered by uncertainty over adequacy decisions, Standard Contractual Clauses, and shifting regulatory environments following Schrems II; these challenges are particularly acute for health and genomic data (Scheibner et al., 2020; Bradford et al., 2020; Lalova-Spinks et al., 2024; Molnár-Gábor & Korbel, 2020). Efforts toward harmonization remain elusive, and European consortia are frequently required to undergo repeated legal review and technical adaptation in response to divergent national and international data protection standards (Lalova-Spinks et al., 2022; Bradford et al., 2020; Richards, 2022; Cathaoir et al., 2021; Molnár-Gábor & Korbel, 2020).
- Emerging Regulatory Trends (EHDS, DGA, Data Altruism). The proposed European Health Data Space (EHDS) and Data Governance Act (DGA) are intended to facilitate future harmonization but currently introduce new ambiguities—such as the scope of Health Data Access Bodies, data altruism organizations, and technical standards—and risk adding complexity unless their implementation is thoroughly coordinated (Lalova-Spinks et al., 2023; Slokenberga, 2022; Rak, 2024; Kertesz, 2024; Terzis & Santamaria Echeverria, 2023; Terzis, 2022). Initial stakeholder perspectives indicate skepticism that these frameworks alone will rapidly resolve entrenched legal and operational fragmentation (Lalova-Spinks et al., 2023; Richards, 2022; Slokenberga, 2022).

Ongoing reforms aim to address these issues, but empirical and conceptual studies emphasize the need for further harmonization, streamlined administrative processes, and clearer EU-level implementation guidance to ensure that regulatory protections do not continue to hinder vital health research across Europe. The recent publication of the European Health Data Space (EHDS) Regulation in the Official Journal of the European Union marks a pivotal development in the European Union's efforts to create a unified digital health ecosystem (European Commission, 2025). As part of a broader strategy to enhance cross-border healthcare and stimulate data-driven innovation, the EHDS establishes a legal framework for the access, exchange, and use of electronic health data across EU Member States. The

[4] This difference in interpretation is quite significant. For example, even within Finland, different ministries operate under varying rules. Health data, overseen by the Ministry of Health and Welfare, is treated differently from socioeconomic registers, which fall under the Ministry of Finance.

Regulation aims to reinforce the EU's leadership in digital health technologies while addressing urgent systemic challenges such as population aging and healthcare workforce shortages (European Union, 2025).

From a theoretical perspective, a cornerstone of the EHDS is the empowerment of individuals through greater control over their personal health data. Citizens will benefit from seamless access to their medical records across borders, facilitating the delivery of high-quality healthcare anywhere within the Union (European Commission, 2025). This patient-centered model promotes continuity of care and ensures that individuals can exercise their health data rights uniformly throughout the EU. In addition to enhancing primary care delivery, the EHDS supports the secondary use of health data—including anonymized and pseudonymized information—for purposes such as scientific research, innovation, public health planning, and evidence-based policymaking. These secondary uses are tightly regulated to ensure compliance with the EU's stringent data protection, ethical, and cybersecurity standards (European Union, 2025).

The Regulation envisions a phased implementation strategy to ensure operational viability. It officially enters into force on 26 March 2025, with data exchanges for the initial priority categories—such as patient summaries—beginning by March 2029. By that time, rules on the secondary use of data will also become applicable to most categories, and additional expansions are expected by March 2031 (European Commission, 2025). To support this complex rollout, over twenty Implementing Acts are expected, alongside the establishment of dedicated EHDS governance structures. These will coordinate with Member States, healthcare providers, researchers, and industry actors to promote adherence, technical compatibility, and trust across national systems.

Ultimately, the EHDS Regulation signifies a foundational shift in how health data is conceptualized and utilized in the EU. By balancing individual privacy rights with the societal benefits of data reuse, the EHDS lays the groundwork for a secure, efficient, and innovative health system that is better equipped to respond to current and future health challenges (European Union, 2025). This initiative stands as a testament to the EU's commitment to fostering a data-driven, inclusive, and resilient digital health future.

4. Health Data in Italy: Privacy Regulations Hinder Their Reuse

The regulation of access to and use of health data in Italy has undergone significant evolution over the past two decades, reflecting broader shifts at the European level as well as domestic legal, ethical, and political considerations. Initially, Italy's regulatory framework was built around strong privacy protections, influenced by the enactment of the "Codice in materia di protezione dei dati personali" (Personal Data Protection Code) in 2003 (Legislative Decree No. 196/2003). This Code established detailed provisions for the handling of personal and sensitive data, including health data, emphasizing informed consent, data minimization, and purpose limitation as key principles (Garante per la Protezione dei Dati Personali, 2003).

The General Data Protection Regulation (GDPR), implemented in 2018, significantly reshaped the Italian data protection landscape. While the GDPR introduced harmonized rules across the European Union for processing personal data, including health data under its special categories (Article 9), Italy adapted its national legislation through Legislative Decree No. 101/2018, which amended the 2003

Code to align with the GDPR (Garante per la Protezione dei Dati Personali, 2018). However, Italy's Data Protection Authority (Garante per la Protezione dei Dati Personali) maintained a cautious and often conservative approach, particularly concerning the secondary use of health data for research purposes. This conservatism led to frequent requirements for explicit, project-specific informed consent, even when the GDPR would have permitted broader consent models or derogations under appropriate safeguards (Florindi et al., 2023).

As a result, researchers in Italy faced substantial administrative burdens and uncertainty. Delays in obtaining approvals, divergent regional practices, and inconsistent interpretations of what constituted adequate anonymization or pseudonymization often hindered the efficient use of existing health datasets for scientific research. Comparative studies found that Italian researchers were disadvantaged relative to their counterparts in countries with more streamlined health data governance models, such as Denmark or the Netherlands (Gkotsi & Gasser, 2021).

In response to mounting pressure from the scientific community and in anticipation of the forthcoming European Health Data Space (EHDS), Italy has initiated reforms to facilitate greater access to health data while maintaining robust privacy protections. From 2024 onward, new guidelines allow the reuse of health data without requiring prior explicit consent when contacting data subjects would be impossible or would risk compromising the scientific objectives of the research (Garante per la Protezione dei Dati Personali, 2024). In line with the GDPR, health data reuse must be based on one of several recognized legal grounds, including explicit consent from the data subject (which must be informed, freely given, specific, written, and revocable), contractual necessity related to care or employment, protection of vital interests, public interest in the field of public health, or scientific research (Regulation (EU) 2016/679). In such cases, researchers must conduct a Data Protection Impact Assessment (DPIA) and publicly justify their decision-making processes, promoting transparency and accountability. The DPIA remains mandatory whenever high risks to individual rights and freedoms are identified, in compliance with Articles 35 and 36 of the GDPR. Depending on the case—especially when intellectual property rights are concerned—the DPIA can be fully or partially published and must be available for consultation by the Garante in cases of significant risk.

For institutions such as the IRCCS (Scientific Institutes for Research, Hospitalization, and Healthcare), research activities are considered part of their core institutional functions (Garante per la Protezione dei Dati Personali, 2022). Nonetheless, compliance with Article 9 of the GDPR, which prohibits the processing of sensitive health data unless specific conditions are met, remains essential. When required, consent must relate specifically to each research project and be renewed as needed.

At the European level, additional legislative initiatives have been launched to strengthen the broader ecosystem for data sharing and security. The Data Governance Act (Regulation (EU) 2022/868), currently active, establishes mechanisms for the safe sharing of public sector data, sensitive data, and data held by companies for altruistic purposes. It introduces the concept of "data intermediaries" to facilitate voluntary data sharing under trustworthy conditions (European Commission, 2022b). This act sets the groundwork for fostering a European data economy by building trust between data holders and users. The forthcoming Data Act, still under negotiation, aims to further regulate access and use rights for data generated by connected devices and services, thus facilitating the functioning of data-driven initiatives like the EHDS (European Commission, 2022b). The Data Act will address issues such as

data portability, data access obligations, and fairness in contracts regarding data sharing, especially critical for ensuring equitable participation of all stakeholders in health innovation.

Complementing these efforts are new regulations focused on cybersecurity. The NIS2 Directive, adopted in December 2022, seeks to strengthen cybersecurity resilience across sectors considered critical, including healthcare, by imposing more stringent security requirements and incident reporting obligations on essential service providers (European Commission, 2023a). Furthermore, the proposed Cyber Resilience Act aims to ensure that connected digital products and services meet mandatory cybersecurity requirements throughout their lifecycle (European Commission, 2023b).

Given the increasing digitalization of healthcare systems, Italy is investing in technical infrastructure to support secure data sharing environments that align with EU standards, such as the use of certified secure processing environments where data cannot be downloaded, and access is limited to pseudonymized information unless anonymization is impossible. These developments are intended to prepare Italy for integration into the EHDS framework, which aims to create a unified European system for primary and secondary health data use (European Commission, 2025).

Nevertheless, challenges persist. The Italian healthcare system's decentralization into regional administrations complicates efforts to harmonize practices nationwide. Variations in digital maturity, interoperability of electronic health records (EHRs), and regional data governance policies continue to create disparities in researchers' access to health data (Osservatorio Innovazione Digitale in Sanità, 2023). Moreover, ongoing public skepticism about data sharing and privacy risks, partly fueled by historical concerns over state data surveillance, necessitates continued efforts to build public trust through transparency, engagement, and ethical governance.

In conclusion, the evolution of Italian regulation governing access to health data reflects a complex balancing act between protecting individual rights and enabling scientific advancement. The vast repositories of health data already collected and stored in databases by hospitals and clinical centers represent a "gold mine" for scientific research—a potential treasure trove for advancing medical knowledge and serving the public good. However, in Italy, much of this data remains inaccessible due to what many researchers consider overly restrictive privacy protections. These regulations, though well-intentioned in safeguarding individual rights, often prevent researchers from utilizing already-available data, collected at no additional cost, for projects with no commercial interest but significant societal benefit (Florindi et al., 2023; Shabani et al., 2018). Moreover, inconsistencies in the approval processes for data access—even among institutions with similar characteristics—further frustrate Italian researchers, who perceive themselves at a disadvantage compared to peers in countries where data reuse protocols are more flexible (Vayena & Blasimme, 2017). Recent reforms, coupled with the alignment to broader European digital strategies, signal a transition toward a more research-enabling environment. Nevertheless, achieving the full potential of health data for innovation and public health improvement will require not only regulatory updates but also deep systemic coordination, infrastructure investments, and sustained efforts to promote public trust and ethical stewardship of personal health information.

5. The Digitalization of Healthcare in Italy: Where Do We Stand?

Centralizing health data (and services) at the European level requires foundational steps in digitalizing the national healthcare system—steps that, despite the acceleration triggered by the pandemic and the funding provided by the National Recovery and Resilience Plan (PNRR, Mission 6: "Health"), appear to be progressing slowly. In Italy, one of the key drivers of healthcare innovation is the Electronic Health Record (Fascicolo Sanitario Elettronico, FSE), which allows citizens to access their medical history while enabling healthcare professionals to view patient information in a comprehensive way to guide medical decisions.

Despite its potential, the FSE remains underutilized. According to the latest Digital Innovation in Healthcare Observatory, only 38% of the population is aware of it, and just 12% knowingly use it. Beyond awareness, one of the main inefficiencies lies in the lack of central coordination, which has hindered integrated and consistent regional adoption. Healthcare is a matter of concurrent jurisdiction between the State and the Regions in Italy. To date, regional governments have widely exercised regulatory discretion, implementing national guidelines inconsistently and without a clearly defined interoperability framework.

The PNRR has allocated specific investments ("Strengthening technological infrastructure and tools for data collection, processing, analysis, and simulation") to harmonize Regional Health Records and ensure widespread adoption nationwide. The goal is to achieve interoperability and connect Italy's digital health infrastructure with the European dimension (MyHealth@EU), as outlined in the forthcoming European Health Data Space (EHDS).

Interoperability and FSE management are already in motion. The next step, aligned with EU-level planning, is to implement a new technological infrastructure under the Ministry of Health, through the New Health Information System (NSIS), which will host Italy's health data assets and enable advanced analyses for the EHDS's secondary purposes—namely, research, innovation, and health policy development.

5.1 – PNRR Objectives and EHDS Implementation: The Need for Central Coordination

The PNRR stipulates that all Italian regions must adopt and implement the Electronic Health Record system in accordance with standardized national criteria by mid-2026. This ambitious timeline will require a strong push from the Ministry of Health and active cooperation from implementing bodies. According to a study by The European House – Ambrosetti's PNRR Observatory, as of December 31, 2022, only 6% of available funds had been spent and just 1% of projects completed. This resulted in delays of over €20 billion in originally scheduled spending from 2020–2022. The Italian Court of Auditors, in its second report on the PNRR, identified the health mission as one of the least advanced in terms of expenditure, noting critical risks that could delay targets for the first half of 2023.

Even if deadlines are met, the EHDS regulation will only become applicable 12 months after its formal entry into force. Given the lack of a coherent and uniform national health system, full implementation in Italy remains unlikely without structural reforms. Adding to this challenge are growing calls from some regions for greater autonomy over healthcare organization—an approach that may increase quality for non-essential services but risks undermining national governance mechanisms.

In this context, AGENAS (the National Agency for Regional Healthcare Services), acting as the operational arm of the Ministry of Health in charge of digital healthcare transformation, must enhance its guidance and oversight capacities. It must ensure that regional implementations align with national goals, while remaining attentive to the European dimension and the EHDS criteria and standards.

The PNRR could serve as a strategic compass, outlining objectives that reflect a vision of healthcare rooted in European integration. It represents a once-in-a-generation opportunity to transform Italy's healthcare system. However, this transformation will require strong political will, centralized leadership, and robust governance mechanisms to coordinate digital health initiatives across all levels.

6. The "Data Gap" and the Challenge of Measuring Health Phenomena

The collection of accurate and comprehensive health data is widely recognized as a fundamental prerequisite for advancing, for example, the understanding of health inequalities, informing the design of effective public policies, and evaluating the efficacy of health interventions. Nevertheless, a persistent challenge within the healthcare sector remains the inadequacy of sufficiently detailed data necessary to measure and analyze critical health phenomena. This deficiency not only impedes the progress of empirical research but also exerts profound consequences on the policymaking process, given that policymakers depend heavily on reliable data to identify and address disparities in health outcomes. The absence of robust and integrated health databases substantially limits the capacity to uncover causal mechanisms underlying health disparities, to monitor long-term epidemiological trends, and to formulate targeted interventions aimed at mitigating vulnerabilities within specific population groups. Inadequate data systems thus pose a significant barrier to the development of evidence-based policies designed to promote health equity. Drawing upon recent scholarly contributions, the following analysis explores the nature of the data gap in health research, assesses its impact on the comprehension of health disparities, and evaluates its implications for the advancement of data-driven, equitable policymaking.

6.1 – The Nature of the "Data Gap" in Health Research

The study of health outcomes inequalities relies on data drawn from administrative records, surveys, and longitudinal studies. Yet, as Case and Kraftman (2024) highlight, many of the available datasets are incomplete or lack key variables needed for comprehensive analysis. One major issue is the fragmentation of data collection efforts. Many health surveys focus primarily on clinical indicators and collect only basic economic data, while economic surveys provide detailed income and wealth information but lack health variables. This division limits the ability to understand the interactions between socioeconomic and health factors.

Another significant challenge is the difficulty of tracking health trajectories over time. For instance, mortality data are often recorded long after the socioeconomic conditions that may have influenced outcomes, making it hard to link deaths to earlier life circumstances and limiting the ability to draw strong conclusions about long-term health determinants. Furthermore, studies focused on specific causes of death or rare conditions often operate with small samples, which undermines statistical significance and generalizability.

Linking diverse datasets could provide a more complete picture of health disparities, but such efforts remain limited. For example, Chetty et al. (2016) in the United States linked tax records with mortality data to study the relationship between income, geographic location, and life expectancy. Their findings revealed substantial life expectancy gaps between income groups and highlighted the importance of economic conditions as health determinants. However, even large-scale studies face limitations, as many administrative datasets lack fundamental variables like educational attainment or ethnicity—both strongly associated with health outcomes.

6.2 – Challenges in Linking Health and Socioeconomic Data

A core obstacle in health inequality research is the challenge of linking individual health data with long-term socioeconomic indicators. Ideally, this would involve merging census data with mortality records to track health trajectories over time (Case & Kraftman, 2024). However, this approach is rarely implemented systematically, and even when attempted, it often relies on small samples that limit analytical depth.

Another recurring problem is ecological fallacy, where researchers infer individual health outcomes from aggregate data. For instance, many studies use geographic units such as states or municipalities as the level of analysis. While this can reveal regional disparities, it often misses individual-level variation. Low-income individuals do not necessarily live in the poorest areas, and regions with high income inequality may contain pockets of poverty within generally affluent communities. This complicates the identification of causal mechanisms and highlights the limitations of relying solely on geographic data (Andrasfay & Goldman, 2021).

Moreover, demographic variables such as race and education are not consistently recorded in health databases, posing further challenges for research on health disparities. In the U.S., for example, educational attainment has been included on death certificates only since 1989, allowing researchers to explore growing mortality gaps by education level. Recent studies show that life expectancy has declined among Americans without a college degree, particularly due to suicide, substance abuse, and cardiovascular disease (Sasson & Hayward, 2019). However, such data practices are not widely adopted in other countries, limiting international comparisons and broader insights into education-health relationships.

6.3 – Implications for Policy and Decision-Making

The lack of comprehensive and integrated data has profound consequences for policymakers, restricting the ability to design targeted and effective health interventions. Without accurate measurements of health inequalities, it becomes difficult to identify vulnerable populations and implement policies addressing the structural causes of poor health outcomes. For example, if existing data fail to capture the long-term effects of childhood conditions on adult health, governments may underfund early-life interventions despite strong evidence of their long-term benefits.

Likewise, the absence of detailed demographic information in health databases creates blind spots in policy design. In many European countries, education level is not systematically recorded on death certificates, hampering evaluations of how education influences health disparities. Improving the quality

and granularity of available data is therefore essential to ensure that policy decisions are grounded in robust evidence.

Additionally, the lack of demographic identifiers in mortality records—such as race and ethnicity—has historically limited the UK's ability to analyze health disparities among different population groups. Only recently has the Office for National Statistics (ONS) begun to address this gap (Case & Kraftman, 2024). Similarly, missing education data in many European countries prevents the identification of mortality trends by social class, further hindering the development of equity-focused health policies (Mackenbach, 2019).

6.4 – Potential Solutions and Future Prospects

Addressing the data gap in health research requires a coordinated effort to improve data collection, integration, and accessibility. A crucial step is expanding initiatives that link administrative data sources, such as tax records, census data, and health registries. This approach would allow researchers to conduct deeper analyses of health inequalities and assess the long-term effects of socioeconomic factors on health outcomes.

Another key strategy is investing in longitudinal studies that follow individuals over time. These studies provide valuable insights into the relationships between early-life conditions, socioeconomic status, and long-term health. Longitudinal cohort studies, such as those conducted in the UK, offer a detailed understanding of health trajectories and can help identify causal relationships that cross-sectional studies cannot capture (Hendi & Ho, 2021).

Policymakers should also prioritize enhancing the granularity of existing datasets by ensuring that critical demographic variables—such as education and ethnicity—are consistently recorded. This would support more nuanced research into health disparities and enable the design of interventions tailored to the needs of specific population groups. Finally, to avoid ecological fallacies, regional studies should be complemented with individual-level analyses wherever possible.

In conclusion, the data gap remains a significant barrier to advancing health research and tackling health inequalities. Incomplete datasets, fragmented collection efforts, and reliance on aggregate-level analysis highlight the urgent need to improve data integration and expand longitudinal research. Addressing these issues would empower policymakers to design more precise and impactful health interventions, ultimately reducing disparities and improving population health. Investing in more accurate data collection and analysis would enable governments and research institutions to base their policies on solid, reliable evidence, contributing to the development of more equitable and efficient healthcare systems.

7. Best Practices in Europe: The Cases of Finland and Denmark

7.1 Accessing health registers in Finland: Findata

Access to individual health register data in Finland is provided by Findata, the Finnish data permit authority for the social and health care data, operating under the Ministry of Social Affairs and Health.[5] Findata issues permits to process: i) data maintained by several public social and health sector controllers, including those that transferred the right to issue permits to Findata itself; ii) register data of private social and health service providers; iii) data stored in Kanta services.[6] Findata is responsible to pre-process and combine data covered by a permit, which the pseudonymization and anonymization of registers.

Findata maintains an expanding list of ready-made registers. Ready-made registers are pre-compiled and pre-processed datasets ready to be quickly made accessible without the need for cost estimates or extraction fees from controllers. Currently ready-made registers at Findata fall under two groups: FinRegistry and COVID-19. FinRegistry consists of the registry data collected in the FinRegistry research project and the research data generated from them. The material includes data from Digital and Population Data Services Agency (DVV), Cancer Registry, Finnish Centre for Pensions (ETK), Kanta services, Kela, THL, and Statistics Finland, insofar as the data is covered by the Act on Secondary Use. Overall, Findata contains over 20 datasets and covers data from several decades.

On top of ready-made registers, Findata handles applications to access non-ready-made social and health care data maintained by the social and health controllers that fall under Findata permits.[7] Moreover, Findata can combine external data sources under the Secondary Use Act.[8] These include data collected by researchers or data accessed via another valid data permit. See section 7.3 for more information on linking external data sources.

7.2 Data permits

Findata and individual data controllers can issue and amend data permits. The first case applies whenever the application involves combining data from multiple controllers covered by the Act. Data requests are submitted to Findata via its online portal. Importantly, some public controllers have transferred the right to issue permits to Findata, so that Findata can issue permits on their behalf under the Act on Secondary Use. In these cases, Findata processes all permit applications related to these controllers' register data. A major controller that falls into this category is the Finnish Institute for Health and Welfare (THL), which provides several registers accessible via a Findata permit.[9] In practice, the system is highly centralized, with Findata acting as a central node when it comes to applications to access health data registers.

Of equal importance, and particularly relevant for linking external data sources, is the legal limitation that prevents Findata from receiving permit authority from all public controllers. In other words,

[5] Findata activities are defined by the Act on the Secondary Use of Health and Social Data (Ministry of Social Affairs and Health, 2019).
[6] Kanta services are a set of digital services that store and use citizens' social welfare and health care data.
[7] See Findata' Data page.
[8] Sensitive data, such as self-collected survey data or data generated via experiments, require an ethical approval from an institutional review board.
[9] THL reserves the right to issue data permits for internal administration and the THL Biobank data permits.

Findata cannot issue permits for data from all controllers covered by the Secondary Use Act. A leading example of is Statistics Finland. In practice, researchers interested in linking Statistics Finland's registers to Findata's registers need to obtain a permit from Statistics Finland, not from Findata.[10] We will return to the dual nature of the Finnish environment in terms of accessing register data when discussing data linkage in Section 7.3.

A permit can be issued exclusively for the purposes laid down in law, which for individual-level data are scientific research, statistics, education, and planning and reporting duty of an authority. As a rule, the Secondary Use Act applies to register-based studies, that is, studies that use register data collected for other purposes or national registers.[11] Moreover, every data permit needs to apply the GDPR minimization principle, which requires to disclose only for data essential to answer to the question included in the proposal. While the application documents are confidential, decisions and permits granted by Findata are public. This is also to facilitate the exercise of the right to opt out on the part of project participants.

Data protection officers at Findata process the data application and evaluate it based exclusively on legal grounds. They can request additional information or modifications to the application or reject it altogether (to which decision the applicant has the right to appeal). When a permit is issued, then the data is extracted by the controllers and securely sent to Findata. As mentioned, Findata pre-processes, pseudonymize and links the data sources covered by the permit. This is done by removing all direct identifiers (e.g., name and surname, tax-authority personal identity numbers) and by creating a pseudo identifier. Pseudo identifiers uniquely identify individuals over time and across registers; they can be used to link information from several registers and to construct family networks within and across generations.

7.3 Linking data sources

Registers whose access is covered by Findata can be combined with external data sources that the applicant is in possess. These can be other registers. The researcher needs to have the right to process the external data sources via a separately issued data permit from the data controller that specifies the use of the data in relation to the Findata registers combined in the same application. Having such data permit from the data controller does not automatically guarantee access to the combined data, as this decision must be taken by Findata after making the data protection (and data minimization) considerations required to release a Findata permit.

Importantly, if the data controller is Statistics Finland, then the whole permit must be applied from Statistics Finland. Suppose for instance that for specific research question a researcher needs to access individual data on hospitalizations and drug purchases (covered by Findata) and link these to individual data on employer (covered by Statistics Finland). In this example, it is not possible to have a permit issued by Findata that comprises the use of Statistics Finland's registers as external data source. Instead, the researcher needs to i) obtain a permit to use the health data from Findata, ii) apply for a data permit at Statistics Finland, asking Statistics Finland to pseudonymize and combine Findata's registers, which will be access from Statistics Finland's servers. In other words, a permit at Statistics Finland can cover

[10] Other examples of public controllers that cannot transfer their permit authority to Findata are the Digital and Population Data Services Agency (DVV) and the Finnish Centre for Pensions (ETK).
[11] Clinical and medical trials do not fall under the Secondary Use Act.

the access to Findata registers (and those covered other controllers, such as DVV, ETK), but the opposite is not true.

Statistics Finland's data access process resembles that described here, but there are some notable differences.[12] For instance, Statistics Finland's interpretation of the GDPR is stricter than Findata's. Register data can only be accesses from within the EU and EEA in the case of Statistics Finland, whereas under specific conditions it can occur from outside of it in case of Findata permits (see next section). It is also worth stressing that registers at Statistics Finland can generally never leave Statistics Finland's secure environments. On the other hand, as explained below, Findata permits can in principle allow data access from local servers if the applicant can prove that the level of data protection meets the same protection existing via Findata's remote access system. While setting up local systems for projects covered by Findata permits is a rather complex, costly, and time-consuming process, it shows a fundamental difference between the rules covering the permits issued by Findata and Statistics Finland.

7.4 Accessing and processing the data

Individual-level data must always be accessed and analyzed in a secure environment. The primary way to access individual-level data is through Findata's secure processing environment, Kapseli. However, under the Act on the Openness of Government Activities, data can also be disclosed to other approved secure environments if necessary. In this case, the processing environment must follow Findata's regulation specifying the information security requirements used for secondary use of social and health data. In practice, the processing environment must be certified by a data security assessment body. The current regulation allows for alternative options, ranging from a secure space with an isolated computer to cloud-based solutions. Foreign researchers' processing environments must comply with these requirements and obtain internationally recognized security certifications, verified by an approved Finnish assessment body. As a rule, the processing environment should have the same level of information security as Findata's own operating environment.

By default, the processing of personal data from abroad is considered a transfer of personal data, even if the data is in a remote access environment. However, this assessment can vary depending on the affiliation of the data processor and of the data controller. If the *data processor* is employed by a *data controller* located within the EU and EEA (EU Member States, Norway, Liechtenstein, and Iceland), then processing from abroad is *not* considered a data transfer, and the processor may access the data from outside the EU/EEA.

Under the EU's General Data Protection Regulation (GDPR), data can be transferred within the EEA under the same conditions as within Finland. For data transfers or processing outside the EEA (third countries), there must be a legal basis as outlined in Chapter V of the GDPR. Consider the following example:

- Findata has granted a data permit covering datasets from HUS, Pirha, and Varha.
- The permit states that the *data processors* are employees of HUS, Pirha, and Varha, whereas the *data controller* of the dataset is HUS.
- The data is processed within Findata's secure Kapseli environment.

[12] See also Lombardi (2025a) for more information on the rules and process to access register data at Statistics Finland, and Lombardi (2025b) for a parallel description of the Swedish system.

- Then, if the data processors travel to the United States for a conference:
    - An employee of HUS can process the data remotely via Kapseli from the U.S., since they are employed by the data controller.
    - Employees of Pirha and Varha, however, are not employed by the data controller (HUS), meaning they cannot process the data in Kapseli from the U.S. without a legal basis under Chapter V of the GDPR.

When it comes to data processing and exporting, data must always be processed to preserve anonymity. and Findata must ensure that results are anonymous before exporting them from a secure environment. Researchers that want to export data from a secure environment must make sure that each cell in a table or underlying a graphical output is based on frequencies of at least 3 units. Moreover, results cannot identify any individuals, either directly or indirectly.

As such, the Finnish system offers a clear example of how data protection requirements can be fulfilled while granting access to register data. This balance is achieved in agreement with National and EU laws, while allowing researchers to access very detailed, individual-level information. As Finland, Italy and the other EU member states must abide by the GDPR and could even take inspiration from the Finnish system when considering whether to open administrative data access for research purposes.

It is important to recall that the current status-quo was reached in a stepwise fashion, building trust on the institutions. Therefore, importing the Nordic model in other settings should likely be made in steps. Fortunately, the Finnish system itself embeds and suggests alternative ways to reach the end goal of granting access to administrative registers. First, data access can alternatively occur via the Kapseli system or by sending the data to a secure environment. If policy makers deem too impractical (or even risky) the latter option, then having a fully centralized data access system appears to be a viable alternative. Second, Finland offers an example of dual system, where different broad categories of registers can be accessed via either a Findata permit or via a Statistics Finland's one. There are pros and cons for having a dual system. For instance, from the perspective of a researcher it is not always practical to navigate a dual system. At the same time, data experts at Statistics Finland and at Findata can fully specialize on (and improve the offer of) specific data sources in ways that would not be achieved in a fully unified setting.

7.6 Accessing health registers in Denmark

Although in the European Union (EU) the General Data Protection Regulation (GDPR) has introduced stringent protections for personal data, with significant implications for biomedical research and precision health initiatives, Denmark represents a leading model in the responsible and innovative use of individual health data for scientific and clinical purposes. The Danish healthcare system is widely regarded as a global leader in the digitalization of health services and the effective governance of personal health data. Its success is rooted in decades of strategic investment, technological innovation, and a strong societal commitment to transparency and trust in public institutions (Healthcare Denmark, 2024; OECD, 2019).

At the core of Denmark's digital health ecosystem is the Civil Registration Number (CPR) system, implemented in 1968. This unique identifier enables the linkage of individual data across diverse public

sector databases, including healthcare, social services, and demographic registries. The CPR system allows healthcare providers to access comprehensive, longitudinal patient histories, thus enabling continuity of care across general practice, hospital services, and municipal care (Healthcare Denmark, 2024; Schmidt et al., 2019).

Sundhed.dk, Denmark's national e-health portal, plays a central role in citizen engagement, offering individuals secure, online access to their health records, laboratory results, prescriptions, referrals, and appointment schedules. Citizens also can track who accesses their health data, bolstering transparency and reinforcing public trust (Healthcare Denmark, 2024; Andreassen et al., 2020). The Shared Medication Record (FMK) further enhances safety by ensuring that all prescribing healthcare professionals operate from a single updated record, dramatically reducing medication errors and adverse drug interactions (Healthcare Denmark, 2024).

Interoperability is a fundamental principle of Denmark's healthcare system. MedCom, a non-profit organization owned by the Ministry of Health, regional authorities, and municipalities, ensures that IT systems across healthcare sectors are compatible and that data can be securely exchanged between GPs, hospitals, pharmacies, and municipal services (MedCom, 2022). The National Service Platform facilitates secure, standardized access to national databases, ensuring that healthcare providers and researchers can utilize rich, integrated data sources without compromising data security (Healthcare Denmark, 2024).

In addition to clinical applications, Denmark has excelled in enabling the secondary use of health data for research and policy planning. The Danish Health Data Authority manages a national research platform where de-identified data can be analyzed in secure environments without individual-level data export. These systems are compliant with GDPR requirements and ensure ethical oversight through dedicated review committees (Legido-Quigley et al., 2024). Denmark's model thus aligns with the FAIR (Findable, Accessible, Interoperable, Reusable) principles critical for modern health data management (Wilkinson et al., 2016).

Moreover, decentralized clinical trials have become a priority through initiatives like the Personalized and Decentralized Clinical Trials (PACT) project. This approach enables participants to contribute to clinical research remotely, using telemedicine, wearables, and digital health platforms, thereby increasing diversity and inclusivity in research cohorts (Healthcare Denmark, 2024; Dorsey & Topol, 2020).

To enable healthcare data exchange and advanced analytics, Denmark relies on the National Service Platform, which connects local IT systems to national registries such as the National Patient Registry, the Cancer Registry, and the Prescription Registry. These national registries are accessible to healthcare professionals under strict governance rules, allowing them to retrieve a comprehensive view of patients' health profiles while ensuring data security and compliance with privacy legislation (Healthcare Denmark, 2024).

These advancements are framed within Denmark's broader vision for 2024–2027, emphasizing a "digital and technological first" strategy. This strategy aligns national objectives with European initiatives, preparing Denmark to actively participate in the European Health Data Space (EHDS), which aims to create a pan-European infrastructure for primary and secondary health data use while

ensuring privacy and security through mechanisms such as secure processing environments, patient opt-out options, and standardized interoperability requirements (European Union, 2025). Recent strategies, such as Denmark's Digital Health Strategy 2022–2026, emphasize a "digital and technological first" vision to address healthcare system challenges, including workforce shortages, aging populations, and the growing burden of chronic diseases (Danish Ministry of Health, 2022). This strategy aims to expand telemedicine, integrate artificial intelligence (AI) into clinical workflows, and prepare the national infrastructure for participation in the forthcoming European Health Data Space (EHDS). The "Coherent Health Network for All" initiative focuses on integrating services across hospital, general practice, and municipal care to create seamless patient pathways. Meanwhile, the PACT project (Personalized and Decentralized Clinical Trials) aims to decentralize clinical trial participation, leveraging digital technologies to allow patients to contribute data from their homes through wearables, teleconsultations, and digital platforms, thus expanding research inclusivity and diversity (Healthcare Denmark, 2024).

International comparisons further highlight Denmark's achievements. According to the OECD (2019), Denmark ranks among the top countries globally in terms of electronic health record adoption, citizen access to digital health services, and secure secondary use of health data for research. However, experts caution that maintaining public trust will require continuous transparency about data use, robust cybersecurity protections, and clear governance structures (van Panhuis et al., 2014).

In conclusion, Denmark's healthcare system offers a model of how national policies, technological interoperability, and citizen-centric governance can create a resilient, efficient, and equitable digital health ecosystem. Through the strategic integration of primary clinical data and secondary research data, Denmark continues to foster innovation, support evidence-based policymaking, and lead global efforts toward realizing the potential of data-driven healthcare.

8. Conclusions

The COVID-19 pandemic underscored the critical importance of reliable, accessible, and high-quality public health data for shaping effective health policies and managing crises. As demonstrated by Ortiz-Prado et al. (2023) and Zhang et al. (2023), the availability of timely and robust data directly influences the development of healthcare strategies, particularly in emergency contexts. Johannesson et al. (2023) further emphasize the need for globally coordinated and comprehensive data systems that transcend national boundaries, enabling more effective responses to global health challenges.

The strategic utilization of health data stands at the intersection of technological innovation, public health advancement, and ethical governance. As this chapter has shown, health data holds transformative potential across multiple domains, from enhancing early detection and personalized medicine to promoting health equity and improving systemic resilience. Yet, despite technological advancements and the proliferation of digital infrastructures, substantial barriers remain to fully exploiting this potential, particularly in contexts where privacy regulations impose significant constraints on data reuse.

The European and Italian regulatory frameworks, while founded on the legitimate goal of protecting individual rights, have often introduced complexities that inhibit the secondary use of health data for research and innovation. The General Data Protection Regulation (GDPR), the Data Governance Act, and the upcoming European Health Data Space (EHDS) represent critical steps toward harmonizing

data governance across Europe. However, as evidenced by the Italian experience, national interpretations and administrative fragmentation continue to hinder progress, requiring additional efforts toward operational clarity, infrastructural investment, and institutional coordination.

Moreover, successful models such as those in Finland and Denmark demonstrate that it is possible to reconcile strong privacy safeguards with dynamic, research-friendly health data ecosystems. These examples underscore the importance of centralized governance, interoperable systems, and secure data access platforms in enabling evidence-based policymaking and scientific discovery while maintaining public trust. Initiatives such as Findata in Finland and the comprehensive digital health infrastructure in Denmark reveal that a balance between data protection and data utility is not only achievable but essential for the future of healthcare.

Though not discussed in this article, emerging solutions—including the use of synthetic data, secure processing environments, and federated learning models—further suggest pathways for overcoming the longstanding "know-do gap." These innovations offer opportunities to leverage data without compromising privacy, enabling more inclusive, equitable, and effective health interventions. These approaches enable analytical insights to be drawn from real-world data while minimizing the risk of privacy breaches. By allowing researchers to work with realistic data representations or to train models across decentralized datasets without transferring them, such methods can significantly mitigate legal and ethical constraints. Although not discussed in detail here, these innovations represent a critical frontier in enabling the responsible and scalable use of individual-level health data for public interest research.

Ultimately, achieving a sustainable and ethically sound health data ecosystem requires not only regulatory adjustments but also cultural change: a renewed commitment to data stewardship, transparency, and cross-sector collaboration. If adequately supported by coherent policies and robust infrastructures, health data can become the foundation of a resilient, adaptive, and learning healthcare system capable of addressing both current and future health challenges.

References

Agbo, C. C., Mahmoud, Q. H., & Eklund, J. M. (2019). Blockchain technology in healthcare: A comprehensive review and directions for future research. *Applied Sciences*, 9(9), 1736. https://doi.org/10.3390/app9091736

Andrasfay, T., & Goldman, N. (2021). Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on racial and ethnic minorities. *Proceedings of the National Academy of Sciences, 118*(5), e2014746118.

Andreassen, H. K., Dyb, K., & May, C. R. (2020). Digitized patient–provider interaction: How does it matter? *Social Science & Medicine, 258*, 113023. https://doi.org/10.1016/j.socscimed.2020.113023

Appari, A., & Eric Johnson, M. (2010). Meaningful Use of Electronic Health Record Systems and Process Quality of Care: Evidence from a Panel Data Analysis of US Acute-Care Hospitals. *Health Services Research*, 45(4), 1128-1148.

Artiga, S., & Orgera, K. (2019). Key facts on health and health care by race and ethnicity. *Kaiser Family Foundation*. https://www.kff.org/report-section/key-facts-on-health-and-health-care-by-race-and-ethnicity-issue-brief

Artiga, S., Orgera, K., & Pham, O. (2020). Disparities in health and health care: Five key questions and answers. *Kaiser Family Foundation*. https://www.kff.org/report-section/disparities-in-health-and-health-care-issue-brief

Ashley, E. A. (2016). The precision medicine initiative: A new national effort. JAMA, 315(7), 713–714. https://doi.org/10.1001/jama.2016.0297

BaHammam, A. S. (2023). Unlocking the Power of Health Datasets and Registries: The Need for Urgent Institutional and National Ownership and Governance Regulations for Research Advancement. *arXiv preprint arXiv:2306.06580*.

Bailey, M. J., Currie, J., & Schwandt, H. (2024). *The Long-Term Effects of Neighborhoods on Children: New Evidence from Linked Administrative Data. Quarterly Journal of Economics, 139*(2), 829–880. https://doi.org/10.1093/qje/qjad063

Bailey, Z. D., Krieger, N., Agénor, M., Graves, J., Linos, N., & Bassett, M. T. (2017). Structural racism and health inequities in the USA: Evidence and interventions. *The Lancet*, 389(10077), 1453–1463. https://doi.org/10.1016/S0140-6736(17)30569-X

Bak, M.A.R., Ploem, M.C., Tan, H.L. *et al.* (2023). Towards trust-based governance of health data research. *Med Health Care and Philos* **26**, 185–200. https://doi.org/10.1007/s11019-022-10134-8

Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs*, 27(3), 759–769. https://doi.org/10.1377/hlthaff.27.3.759

Bisceglia L, Caranci N, Giorgi Rossi P, Zengarini N. (2023). La privacy e l'epidemiologia: troviamo una soluzione condivisa. *Recenti Prog Med*; 114(6):332-336. doi 10.1701/4042.40224

Bradford L., Aboy M., Liddell K. (2020). International transfers of health data between the EU and USA: a sector-specific approach for the USA to ensure an 'adequate' level of protection, *Journal of Law and the Biosciences*, Vol. 7, Issue 1, Jan-Jun, lsaa055, https://doi.org/10.1093/jlb/lsaa055

Braveman, P., Egerter, S., & Williams, D. R. (2011). The social determinants of health: Coming of age. *Annual Review of Public Health*, 32, 381–398. https://doi.org/10.1146/annurev-publhealth-031210-101218

Brück, O., Sanmark, E., Ponkilainen, V. *et al.* (2024). European health regulations reduce registry-based research. *Health Res Policy Sys* **22**, 135. https://doi.org/10.1186/s12961-024-01228-1

Cagnazzo, C., Ficara, B., Palermo, F., & Fagioli, F. (2024). Mind the gap with Europe. Why public interest cannot be a good reason to perform observational and epidemiological research?. *Tumori*, *110*(1), 6–9. https://doi.org/10.1177/03008916231190956

Case, A., & Deaton, A. (2021). *Deaths of despair and the future of capitalism*. Princeton University Press.

Case, A., & Kraftman, L. (2024). Challenges for empirical analysis: Data deficits. *Oxford Open Economics, 3*(S1), 503-504.

Cathaoir, K. Ó., Gunnarsdóttir, H. D., & Hartlev, M. (2021). The journey of research data: Accessing nordic health data for the purposes of developing an algorithm. *Medical Law International*, *22*(1), 52-74. https://doi.org/10.1177/09685332211046179

Chen, M., Wang, L., Zhang, L., & Zhang, Y. (2022). Predictive modeling in healthcare using health data: Applications and challenges. *Frontiers in Public Health*, *10*, 890134. https://doi.org/10.3389/fpubh.2022.890134

Chetty, R., et al. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA, 315*(16), 1750-1766.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., & the Opportunity Insights Team. (2020). *The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data. Quarterly Journal of Economics*, 137(4), 1911–1956. https://doi.org/10.1093/qje/qjaa037

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., … & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395–400. https://doi.org/10.1126/science.aba9757

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. New England Journal of Medicine, 372(9), 793–795. https://doi.org/10.1056/NEJMp1500523

Danish Ministry of Health. (2022). *A Stronger and More Coherent Digital Health Strategy 2022–2026*. Retrieved from https://sum.dk/

Dixon, B. E., Grannis, S. J., & Gibson, P. J. (2016). Leveraging data to improve population health: The importance of evaluation. *Public Health Reports*, 131(1_suppl), 54–59. https://doi.org/10.1177/00333549161310S110

Doetsch JN, Dias V, Indredavik MS et al. (2021). Record linkage of population-based cohort data from minors with national register data: a scoping review and comparative legal analysis of four European countries [version 2; peer review: 3 approved]. Open Res Europe, **1**:58 (https://doi.org/10.12688/openreseurope.13689.2)

Dorsey, E. R., & Topol, E. J. (2020). State of Telehealth. *New England Journal of Medicine, 375*(2), 154–161. https://doi.org/10.1056/NEJMra1601705

Dove, E. S., Townend, D., Meslin, E. M., Bobrow, M., Littler, K., Nicol, D., & Knoppers, B. M. (2017). Ethics review for international data-intensive research. *Science*, 357(6349), 653–654. https://doi.org/10.1126/science.aan7158

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., … & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

European Commission. (2021). Proposal for a Regulation on harmonized rules on fair access to and use of data (Data Act). COM(2022) 68 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0068

European Commission. (2022a). *Regulation (EU) 2022/868 on European data governance (Data Governance Act)*. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R0868

European Commission. (2022b). *Proposal for a Regulation on harmonised rules on fair access to and use of data (Data Act)*. COM/2022/68 final.

European Commission. (2022c). Proposal for a Regulation on the European Health Data Space. COM(2022) 197 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0197

European Commission. (2023a). *Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2 Directive)*. Official Journal of the European Union.

European Commission. (2023b). Proposal for a Regulation on horizontal cybersecurity requirements for products with digital elements (Cyber Resilience Act). COM(2022) 454 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0454

European Commission. (2025). *Regulation on the European Health Data Space published*. Health and Food Safety e-News. Retrieved from https://ec.europa.eu/newsroom/sante/newsletter-archives/60931

European Union. (2025). *Regulation (EU) 2025/XXX of the European Parliament and of the Council on the European Health Data Space*. Official Journal of the European Union. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32025RXXXX

Floridi, L. (2020). Artificial intelligence, privacy, and data protection: Challenges and opportunities. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2169), 20190162. https://doi.org/10.1098/rsta.2019.0162

Florindi, F., Rea, F., & Fagherazzi, G. (2023). Unlocking health data for research: Challenges and opportunities in the GDPR era. *Journal of Medical Internet Research*, 25, e45632. https://doi.org/10.2196/45632

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), 20160360. https://doi.org/10.1098/rsta.2016.0360

Friedman, C. P., Rubin, J. C., & Sullivan, K. J. (2017). Toward an information infrastructure for global health improvement. *Yearbook of Medical Informatics*, *26*(1), 16–23. https://doi.org/10.15265/IY-2017-010

Gao Q, Hu Y, Liao Z and Yao L (2023) Cyber-infrastructure and epidemic precautionary policy: evidence from China. *Front. Public Health*. 11:1198928. doi: 10.3389/fpubh.2023.1198928

Garante per la Protezione dei Dati Personali. (2003). *Codice in materia di protezione dei dati personali* (Legislative Decree No. 196/2003). Retrieved from https://www.garanteprivacy.it

Garante per la Protezione dei Dati Personali. (2018). *Decreto Legislativo 10 agosto 2018, n. 101*. Retrieved from https://www.garanteprivacy.it

Garante per la Protezione dei Dati Personali. (2022). Linee guida in materia di trattamento dei dati personali per scopi di ricerca scientifica. https://www.garanteprivacy.it

Garante per la Protezione dei Dati Personali. (2024). Aggiornamento delle regole per il riutilizzo dei dati sanitari a fini di ricerca. https://www.garanteprivacy.it

Ginsburg, G. S., & Phillips, K. A. (2018). Precision medicine: From science to value. Health Affairs, 37(5), 694–701. https://doi.org/10.1377/hlthaff.2017.1624

Gostin, L. O., Halabi, S. F., & Wilson, K. (2019). Health data and privacy in the digital era. *JAMA*, 322(23), 2305–2306. https://doi.org/10.1001/jama.2019.17126

Gkotsi, G. M., & Gasser, U. (2021). Balancing privacy and data reuse in the European Health Data Space: Challenges and solutions. *Health Policy*, 125(10), 1345–1350. https://doi.org/10.1016/j.healthpol.2021.08.002

Grady, C., Eckstein, L., Berkman, B., Brock, D., Cook-Deegan, R., Fullerton, S. M., … & Wendler, D. (2015). Broad consent for research with biological samples: Workshop conclusions. *American Journal of Bioethics*, 15(9), 34–42. https://doi.org/10.1080/15265161.2015.1062162

Healthcare Denmark. (2024). *Digitalisation of the Danish Healthcare Sector*. Retrieved from https://www.healthcaredenmark.dk

Hendi, A. S., & Ho, J. Y. (2021). Immigration and the narrowing education gap in US adult mortality. *Demography, 58*(3), 905-929.

Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the Medicare fee-for-service program. *New England Journal of Medicine*, 360(14), 1418–1428. https://doi.org/10.1056/NEJMsa0803563

Johannesson JM, Glover WA, Petti CA, Veldman TH, Tsalik EL, Taylor DH, Hendren S, Neighbors CE, Tillekeratne LG, Kennedy SW, Harper B, Kibbe WA, Corbie G, Cohen-Wolkowiez M, Woods CW and Lee MJ (2023). Access to COVID-19 testing by individuals with housing insecurity during the early days of the COVID-19 pandemic in the United States: a scoping review. *Front. Public Health*. 11:1237066. doi: 10.3389/fpubh.2023.1237066

Kelley, T. A., Brown, A., Perkins, R. M., & Adil, M. T. (2019). Data maturity in healthcare: A model to measure and advance data capabilities in health systems. *Healthcare Management Forum*, *32*(1), 24–29. https://doi.org/10.1177/0840470418806402

Kertesz F. (2024) Collaboration in Healthcare: Implications of Data Sharing for Secondary Use in the European Union. Eur J Health Law. 2024 Aug 22;31(5):497-517. https://doi: 10.1163/15718093-bja10133.

Khoury, M. J., Iademarco, M. F., & Riley, W. T. (2018). Precision public health for the era of precision medicine. *American Journal of Preventive Medicine*, *54*(3), 398–401. https://doi.org/10.1016/j.amepre.2017.11.017

Knoppers, B. M. (2014). Framework for responsible sharing of genomic and health-related data. *The HUGO Journal*, 8(1), 3. https://doi.org/10.1186/s11568-014-0003-1

Knoppers, B. M., & Joly, Y. (2022). Building a responsible health data ecosystem. *Nature Medicine*, *28*, 1114–1116. https://doi.org/10.1038/s41591-022-01843-2

Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., … & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497. https://doi.org/10.1126/science.abb4218

Kruk, M. E., Gage, A. D., Joseph, N. T., Danaei, G., García-Saisó, S., & Salomon, J. A. (2018). Mortality due to low-quality health systems in the universal health coverage era: A systematic analysis of amenable deaths in 137 countries. *The Lancet*, 392(10160), 2203–2212. https://doi.org/10.1016/S0140-6736(18)31668-4

Lalova-Spinks T, De Sutter E, Valcke P, Kindt E, Lejeune S, Negrouk A, Verhenneman G, Derèze JJ, Storme R, Borry P, Meszaros J, Huys I. (2022). Challenges related to data protection in clinical research before and during the COVID-19 pandemic: An exploratory study. Front Med (Lausanne). Oct 10;9:995689. doi: 10.3389/fmed.2022.995689.

Lalova-Spinks T, Meszaros J, Huys I. (2023). The application of data altruism in clinical research through empirical and legal analysis lenses. Front Med (Lausanne). Mar 30;10:1141685. doi: 10.3389/fmed.2023.1141685. PMID: 37064041; PMCID: PMC10098212.

Lalova-Spinks T, Valcke P, Ioannidis JPA, Huys I. (2024.)EU-US data transfers: an enduring challenge for health research collaborations. NPJ Digit Med. Aug 16;7(1):215. doi: 10.1038/s41746-024-01205-6.

Legido-Quigley, C., Wewer Albrechtsen, N. J., Blond, M. B., Corrales Compagnucci, M., Ernst, M., Herrgård, M. J., Minssen, T., Ottosson, F., Pociot, F., Rossing, P., & Sulek, K. (2024). Data sharing restrictions are hampering precision health in the European Union. *Nature Medicine*. https://doi.org/10.1038/s41591-024-03437-1

Lo, B., & DeMets, D. L. (2016). Incentives for clinical trialists to share data. *New England Journal of Medicine*, 375(12), 1112–1115. https://doi.org/10.1056/NEJMp1608343

Lombardi S. (2025a). Accessing Administrative Data for Research: Lessons from Finland, in F. Berton and P. Paruolo (eds.), "Data-driven learning: using administrative sources to create public value - an EU perspective", Springer, *forthcoming*.

Lombardi S. (2025b). Accessing Administrative Data for Research: Lessons from Sweden", in F. Berton and P. Paruolo (eds.), "Data-driven learning: using administrative sources to create public value - an EU perspective", Springer, *forthcoming*.

Longitools Consortium. (n.d.). *Longitudinal data for tools to predict disease*. Retrieved from https://www.longitools.org

Mackenbach, J. P. (2019). *Health inequalities: Persistence and change in modern welfare states*. Oxford University Press.

Marmot, M., Allen, J., Boyce, T., Goldblatt, P., & Morrison, J. (2020). *Health equity in England: The Marmot review 10 years on*. Institute of Health Equity. https://www.instituteofhealthequity.org

McLachlan, S., Dube, K., & Gallagher, T. (2018). Using health data for a learning health system: The perspective of healthcare stakeholders. *Learning Health Systems*, 2(3), e10047. https://doi.org/10.1002/lrh2.10047

McKinsey & Company. (2021). *Telehealth: A quarter-trillion-dollar post-COVID-19 reality?* https://www.mckinsey.com/industries/healthcare/our-insights

McKinsey & Company. (2021). *The future of wellness: The role of digital health*. https://www.mckinsey.com/industries/healthcare

MedCom. (2022). *Annual Report: Standardizing Digital Communication in Healthcare*. Retrieved from https://www.medcom.dk

Mehrotra, A., Chernew, M., Linetsky, D., Hatch, H., & Cutler, D. (2020). The impact of the COVID-19 pandemic on outpatient visits: A rebound emerges. *The Commonwealth Fund*. https://doi.org/10.26099/5eg0-0919

Mello, M. M., Francer, J. K., Wilenzick, M., Teden, P., Bierer, B. E., & Barnes, M. (2018). Preparing for responsible sharing of clinical trial data. *New England Journal of Medicine*, 369(17), 1651–1658. https://doi.org/10.1056/NEJMhle1309073

Ministry of Social Affairs and Health (2019). Act on the Secondary Use of Health and Social Data. Helsinki. Available at: https://stm.fi/en/secondary-use-of-health-and-social-data

Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Molnár-Gábor F, Korbel JO. (2020). Genomic data sharing in Europe is stumbling-Could a code of conduct prevent its fall? EMBO Mol Med. Mar 6;12(3):e11421. doi: 10.15252/emmm.201911421. Epub 2020 Feb 18.

Naz O, Ibrahim M, Mohiuddin AF, Khan AA and Samad Z (2023) Public health data quality and evidence use in developing countries: a call to action. *Front. Public Health* 11:1194499. doi: 10.3389/fpubh.2023.1194499

OECD. (2019). *Health in the 21st Century: Putting Data to Work for Stronger Health Systems*. OECD Health Policy Studies. https://doi.org/10.1787/e3b23f8e-en

OECD. (2021). *Health at a Glance 2021: OECD Indicators*. OECD Publishing. https://doi.org/10.1787/4dd50c09-en

Ortiz-Prado E, Izquierdo-Condoy JS, Vasconez-Gonzalez J, GarciaBereguiain MA. (2023). Upholding academic freedom: a call to protect freedom of expression and science in Ecuador and beyond. Front Public Health. 11:1259110. doi: 10.3389/fpubh.2023.1259110.

Osservatorio Innovazione Digitale in Sanità. (2023). *Rapporto annuale sull'innovazione digitale in sanità*. Politecnico di Milano.

Péloquin, D., Barnes, M., et al. (2020). Disruptive and avoidable: GDPR challenges to secondary research uses of data. European Journal of Human Genetics.

Porsdam Mann, S., Savulescu, J., & Sahakian, B. J. (2016). Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. *Philosophy & Technology*, 29(2), 97–120. https://doi.org/10

Porter, M. E. (2010). What is value in health care? *New England Journal of Medicine*, 363(26), 2477–2481. https://doi.org/10.1056/NEJMp1011024

Rak R. (2024). Anonymisation, Pseudonymisation and Secure Processing Environments Relating to the Secondary Use of Electronic Health Data in the European Health Data Space (EHDS). *European Journal of Risk Regulation*;15(4):928-938. doi:10.1017/err.2024.67

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*, L119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Rehman, A., Naz, M., & Razzak, I. (2020). Leveraging big data analytics in healthcare enhancement: Trends, challenges and opportunities. *Journal of Big Data*, 7(1), 1–20. https://doi.org/10.1186/s40537-020-00327-1

Richards R. (2022). Barriers on cross-border sharing of health data for secondary use and options to overcome these. Eur J Public Health. Oct 25;32(Suppl 3):ckac129.367. doi: 10.1093/eurpub/ckac129.367.

Rothstein, M. A. (2015). Secondary use of health data: Are we asking the right questions? *Journal of Law, Medicine & Ethics*, *43*(2), 350–355. https://doi.org/10.1111/jlme.12251

Salathé, M., Althaus, C. L., Neher, R., Stringhini, S., Hodcroft, E., Fellay, J., … & Low, N. (2020). COVID-19 epidemic in Switzerland: On the importance of testing, contact tracing and isolation. *Swiss Medical Weekly*, 150, w20225. https://doi.org/10.4414/smw.2020.20225

Sasson, I., & Hayward, M. D. (2019). Education and the increasing mortality disparity in the United States. *Demography, 56*(3), 977-1002.

Scheibner J, Ienca M, Kechagia S, Troncoso-Pastoriza JR, Raisaro JL, Hubaux JP, Fellay J, Vayena E. (2020). Data protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies. J Law Biosci. May 6;7(1):lsaa010. doi: 10.1093/jlb/lsaa010.

Schmidt, M., Schmidt, S. A. J., Sandegaard, J. L., Ehrenstein, V., Pedersen, L., & Sørensen, H. T. (2019). The Danish National Patient Registry: A review of content, data quality, and research potential. *Clinical Epidemiology, 7*, 449–490. https://doi.org/10.2147/CLEP.S91125

Scudellari, M. (2020). How the pandemic might play out in 2021 and beyond. *Nature*, 584(7819), 22–25. https://doi.org/10.1038/d41586-020-02278-5

Shabani, M., Dyke, S. O., Joly, Y., Borry, P., & Knoppers, B. M. (2018). Controlled access under review: Improving the governance of genomic data access. *PLoS Biology*, 16(12), e2006649. https://doi.org/10.1371/journal.pbio.2006649

Shabani, M., & Marelli, L. (2019). Re-identifiability of genomic data and the GDPR. *EMBO Reports*, 20(6), e48316. https://doi.org/10.15252/embr.201948316

Sherman, R. E., Anderson, S. A., Dal Pan, G. J., Gray, G. W., Gross, T., Hunter, N. L., … & Califf, R. M. (2016). Real-world evidence—what is it and what can it tell us? *New England Journal of Medicine*, 375(23), 2293–2297. https://doi.org/10.1056/NEJMsb1609216

Slokenberga, S. (2022). Scientific research regime 2.0? Transformations of the research regime and the protection of the data subject that the proposed EHDS regulation promises to bring along. *Technology and Regulation*, *2022*, 135-147.

Smit, J. R., Mostert, M., van der Graaf, R., Grobbee, D. E., & van Delden, J. J. M. (2024). Specific measures for data-intensive health research without consent: a systematic review of soft law instruments and academic literature. *European journal of human genetics : EJHG*, *32*(1), 21–30. https://doi.org/10.1038/s41431-023-01471-0

Smit, Julie-Anne & Graaf, Rieke & Mostert, Menno & Vaartjes, Ilonca & Zuidgeest, Mira & Grobbee, Diederik & Delden, Johannes. (2023). Overcoming ethical and legal obstacles to data linkage in health research: stakeholder perspectives. International Journal of Population Data Science. 8. 10.23889/ijpds.v8i1.2151.

Steinhubl, S. R., Muse, E. D., & Topol, E. J. (2015). The emerging field of mobile health. *Science Translational Medicine*, *7*(283), 283rv3. https://doi.org/10.1126/scitranslmed.aaa3487

Terzis, P. (2022). Compromises and Asymmetries in the European Health Data Space. *European Journal of Health Law*, *30*(3), 345-363.

Terzis, P., & Santamaria Echeverria, O. E. (2023). Interoperability and governance in the European Health Data Space regulation. *Medical Law International*, *23*(4), 368-376.

Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.

Torous, J., & Roberts, L. W. (2017). Needed innovation in digital health and smartphone applications for mental health: Transparency and trust. *JAMA Psychiatry, 74*(5), 437–438. https://doi.org/10.1001/jamapsychiatry.2017.0262

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics, 19(9), 581–590. https://doi.org/10.1038/s41576-018-0018-x

Van Der Wel, K. A., Östergren, O., Lundberg, O., Korhonen, K., Martikainen, P., Andersen, A. N., & Urhoj, S. K. (2019). A gold mine, but still no Klondike: Nordic register data in health inequalities research. *Scandinavian journal of public health*, *47*(6), 618–630. https://doi.org/10.1177/1403494819858046

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., … & Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health, 14*, 1144. https://doi.org/10.1186/1471-2458-14-1144

Vayena, E., & Blasimme, A. (2017). Biomedical big data: New models of control over access, use and governance. *Journal of Bioethical Inquiry*, 14(4), 501–513. https://doi.org/10.1007/s11673-017-9809-6

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLoS Computational Biology*, 11(2), e1003904. https://doi.org/10.1371/journal.pcbi.1003904

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

World Health Organization. (2023). *Global strategy on digital health 2020–2025: Mid-term review*. https://www.who.int/publications/i/item/9789240077740

Zhang M, Zou K, Liu Z, Liu D, Wang X, Shi Y, Chen Z, Cheng X, Lang B, Li H, Zeng L, Tang Y, Zhao S, Jiang Y, Choonara I and Zhang L (2023) Availability of essential medicines, progress and regional distribution in China: a systematic review and meta-analysis. *Front. Public Health* 11:1149838. doi: 10.3389/fpubh.2023.1149838.