Using Health Data for Research: Evolving National Policies

January 2021 (Version 1.0)

Editor: **Limor Shmerling Magazanik**

Managing Director, Israel Tech Policy Institute

ISRAEL TECH POLICY INSTITUTE
An Affiliate of the Future of Privacy Forum

FUTURE OF PRIVACY FORUM

The Israel Tech Policy Institute (ITPI) is a policy think-tank and research institute founded in Tel-Aviv, Israel, in 2018. Unique among Israeli non-academic research institutions, the ITPI provides a platform for policy discourse across a range of various stakeholders, including government, industry, academia and civil society. It aims to support technological innovation by helping ensure respect for human and civil rights. Its mission is to create an effective channel for policy debates around the ethical, regulatory, and societal implications of emerging technologies.

The Future of Privacy Forum (FPF), a Washington, DC, based think tank established ITPI as its Israeli chapter. FPF is a catalyst for privacy leadership and scholarship advancing responsible data practices in support of emerging technologies. Its advisory board comprises leading figures from industry, academia, and advocacy groups. The FPF and ITPI are supported by more than 200 companies specializing in various fields, including technology, internet, communications, finance, health, retail, and more. FPF receives significant support from the US National Science Foundation , the Alfred P. Sloan Foundation, the Bill & Melinda Gates Foundation, the Chan Zuckerberg Initiative, and the Robert Wood Johnson Foundation.

The views expressed in this paper are not the views of any member or supporter and none were provided any opportunity to influence the work.

ITPI and FPF gratefully acknowledge the support of Intel Corporation for this project.

This report includes valuable contributions by:

Ann Waldo, Sandra Azria, Caoimhe Stafford, Noam Rosen, Dr. Sivan Tamir, Minna Paltiel, Kirsi Talonen and Omer Tene.

This report was also supported by FPF staff including:

Dr. Rachele Hendricks-Sturrup, Katelyn Ringrose, Dr. Sara Jordan, Jules Polonetsky and Hannah Schaller.

Executive Summary

The COVID-19 pandemic has brought to the fore the crucial role that data collection, analysis, sharing, and dissemination play for governments, academic institutions, and private sector businesses racing to advance scientific research to help combat the virus. It also illustrates that data protection safeguards are essential to build public trust for the swift adoption of data-based solutions, such as the myriad efforts related to pandemic related research. The interactions between data protection and scientific research are complex, with privacy and data protection enhancing individuals' trust and ensuring respect of fundamental rights and ethical standards, while at the same time setting parameters and boundaries for data collection and sharing across organizations and borders. Nowhere is this balancing of interests and rights clearer than in the context of secondary use of healthcare data for scientific research. Countries around the world are charting new paths to seek the insights that health data can reveal while at the same time respecting individual rights.

Even before the pandemic, public sentiment strongly supported healthcare data based research. For example, a 2017 survey showed that 93% of Australians support using medical records for research, and more than 95% had moderate, high, or very high trust that researchers would use health data responsibly. The EU General Data Protection Regulation (GDPR) recognizes scientific research as an issue of prime interest to the public. Recital 159 of the GDPR reads, "For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research." Explicit allowances are made for secondary use of personal data for scientific research in Article 89 of the GDPR as well as Article 5(1)(b) of the regulation, which sets forth the purpose limitation principle, while at the same time providing, "further processing for .... scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes."

However, concerns about misuses of health data and skepticism about uses for research are also widely documented. Machine learning techniques that are a priority for many researchers require access to vast datasets and raise fears of bias and discrimination. Use of real world evidence casts a wider net for health research far beyond the traditional data collected for medical care. These factors and others raise the stakes for data research benefits and the risks at an individual and community level.

In this survey, the Israel Tech Policy Institute and the Future of Privacy Forum examined

the legal frameworks for secondary use of healthcare data, including demographics, diagnoses, symptoms, prescriptions, immunizations, tests and other medical conditions, for research purposes in eight countries, including Australia, England, Finland, France, India, Ireland, Israel and the US. The research demonstrates large commonality across legal systems and regimes, permitting secondary use of healthcare data for research purposes under certain conditions, including review by ethical boards, proper de-identification and additional administrative, technical and contractual safeguards.

Legal frameworks typically extend the definition of secondary research to academic, government, and pharmaceutical research, including pharmacovigilance, but excluding uses of data for clinical purposes or for the operation of the healthcare system. These purposes, in turn, are based on other legal mechanisms. The Health Insurance Portability and Accountability Act (HIPAA), for example, defines a sphere of healthcare data uses commonly referred to as Treatment, Payment, and Healthcare Operations (TPO). Under HIPAA, organizations may use patient data for TPO, as opposed to research purposes, even without specific patient consent. Irish law excludes from the obligations that apply to research purposes activities such as service evaluations, clinical audits, and "usual practice" (e.g. investigations of the health of a population and the causes of disease).

The legal frameworks governing secondary use of healthcare data for research purposes are complex. They typically include at least three layers of laws and regulations: general privacy and data protection laws, which regard healthcare data as sensitive information subject to strict protections; patient privacy laws; and public health regulation. In certain countries, such as the US and Australia, a federal system adds complexity with state and provincial laws providing additional layers of regulation. Consider the recent California Consumer Privacy Act of 2018 (CCPA), which sets new obligations for health data that is exempt from HIPAA, when conducted by commercial entities. In England, for example, secondary use of healthcare data is governed by a complex legal framework, including common law, statutory law, and government decisions and recommendations from health, ethics, and data protection authorities. And in this context, GDPR provides latitude for national derogations that have led to extensive member state legislation.

Several of the examined countries have created or are in the process of creating centralized national hubs for healthcare data. In 2019, Finland, for example, enacted a law explicitly intended to clarify the legal basis for broader secondary use of healthcare data. The law aims to encourage use of healthcare data for socially valuable purposes by enabling data users to have a "one-stop-shop" for access requests to such data. The law also created a new government agency, specifically tasked with collecting

and managing healthcare data from numerous disparate sources, securing the data, and efficiently processing access requests from researchers. One key benefit of such a centralized system is the simplification of bureaucratic hurdles. Where researchers once had to wait two years or more to access information, they can now clear the process within three months.

France, for example, established another model, whereby the national data protection authority (CNIL) published reference methodologies for those wishing to access the national health data system. Researchers that do not certify to these standards must undergo a CNIL review process before accessing the data.

Consent

All legal frameworks explored in this project generally[1] empower patients to authorize healthcare data research with explicit consent.[2] This, however, is a basis that may often be infeasible, especially where data is used for purposes unexpected at the time of collection or for population level research. All countries surveyed therefore also permit data research on other grounds[3] and with additional safeguards, such as ethical reviews, pseudonymization and security measures. In some countries, to be approved without consent, research must also be deemed to be in the public interest. In others, such as the US, research must contribute to generalizable knowledge. The manner in which these legal grounds and safeguards are structured or prioritized and the extent to which consent is privileged creates variations across countries. And these divergences, in turn, pose limitations for research and scientific collaboration.

Under Irish law, for example, a data controller seeking an alternative to consent must apply to a special central committee to assess whether the public interest in conducting health data research "significantly outweighs" the interest in obtaining patient consent. In the US, HIPAA allows several pathways for healthcare data to be used or disclosed for research, including pursuant to a waiver of consent by a local Institutional Review Board or Privacy Board; for limited purposes "preparatory to research"; as part of

1. Under GDPR consent is often considered not freely given if the processor is in a position of power, such as a public authority, employer, or in clinical trials. An alternative to consent is required in such cases.
2. Whether consent can be general for wide areas of research or must be specific is an important distinction between jurisdictions, but outside the scope of this paper.
3. Under GDPR scientific research is a compatible secondary use that doesn't require an additional legal basis.

a "Limited Data Set" subject to a Data Use Agreement prohibiting re-identification; and more. English law authorizes secondary use of healthcare data if a researcher shows that obtaining patient consent is not practical and that anonymized information cannot be used; the activity has a medical purpose, such as medical research that has received ethics approval from a Research Ethics Committee; and that the activity is in the public interest or in the interests of improving patient care. In Australia, research is permitted where obtaining consent is impracticable; the information is "necessary" for research; the research outcome is relevant to public health or public safety; a Human Research Ethics Committee has approved the project; and sufficient de-identification is deployed.

Some countries provide patients with a right to opt out from having their health data analyzed for scientific research. Critics argue that in some cases general opt out rights could unintentionally skew research results injecting bias into public health decision making or into machine-learning algorithms. In England, for example, government operations, including NHS resource allocation and treatment protocols, rely on analyses of patient data made available under data research mandates. The risk is that if opt outs concentrate in certain groups or communities, such analyses may become less accurate and useful, leading to inappropriate treatment and deployment of government services or critical adverse impacts on certain populations.

De-identification

Under most data protection frameworks, anonymized data is not subject to legal protections. For data to be considered anonymized, some countries require that re-identification be practically impossible; others accept anonymization if re-identification is unlikely or risks are remote. France and Finland are guided by stricter de-identification standards, according to the European Data Protection Board (EDPB) guidance under GDPR. In Finland, the law focused on secondary use of health and social data creates a dedicated agency to centralize and manage access. Under Finnish law, only aggregate data is freely available to the public, whereas all other forms of personal data, including pseudonymized data, are available only in a secure user environment and subject to contractual restrictions. The UK has historically taken the position that de-identification risk can be remote, but maybe aligning more closely with EDPB guidance. Under US law, data that is "de-identified" under the HIPAA privacy rule is no longer considered Protected Health Information and is thus not subject to any restrictions. This is the case where health information does not identify – and there is no reasonable basis to believe that it can be used to reidentify – an individual patient. HIPAA provides two methods for de-identification, a Safe Harbor method, whereby an organization removes 18 direct and indirect identifiers and does not have actual

knowledge that the information could be used alone or in combination with other information to identify an individual; and an Expert Determination method, whereby an expert with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods of de-identification determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify a data subject.

Ethics

Ethical review is generally required for all human subject research, independently from data protection law. Review panels are tasked with protecting the rights, safety, dignity and well-being of research participants. Ethics committee guidance, however, may also set parameters for consent to participate in research or may provide a waiver of consent requirements. If a research project involves further sharing of data for research, the ethics committee may require or waive consent with regard to such data sharing. In the US, the criteria for waivers of consent under HIPAA are that the use or disclosure of healthcare data involves no more than minimal risk to the privacy of individuals and that the research could not practicably be conducted with patient consent or without access to and use of the data. In Australia, a committee approving the waiver of consent for a research project must be satisfied that several preconditions are met, including that involvement in the research carries no more than low risk to participants; the benefits from the research justify any risks of harm associated with not seeking consent; it is impracticable to obtain consent; there is no known or likely reason for thinking that participants would not have consented if they had been asked; and there is sufficient protection of patient privacy and data confidentiality.

However, it is important to note that the EDPB has advised that the informed consent that ethics committees often require for research data sharing may not always align with the legal basis under data protection law, due to concerns that in a clinical trial relationship data protection consent may not be freely given. Thus, another legal basis other than consent is required under data protection law for access to the same research data.

Additional safeguards

Legal and ethical frameworks ensure additional safeguards for healthcare data research. In England and Finland, researchers must execute a data-sharing agreement and comply with strict information governance protocols. In the US, in contrast to de-identified data, which is outside the remit of the law, data in a Limited Data Set

remains Protected Health Information subject to HIPAA restrictions[4]. While it may be used and disclosed for research without consent, such use or disclosure require entry into a Data Use Agreement containing specific re-identification obligations and other prohibitions. In Finland, the central data hub allocates access rights in line with the issued permits; monitors access to devices, systems, and office sites; deploys security controls to prevent unauthorized access; and monitors and restricts data communications. It also deploys extensive logging systems that record all data-processing events. In Ireland, processes and procedures relating to the management and conduct of health data research projects include: an assessment of the data protection implications of the research; a data protection impact assessment where an initial assessment indicates a high risk; measures that demonstrate compliance with GDPR data minimization requirements; controls to limit access to health data to prevent unauthorized consultation, alteration, disclosure or erasure; controls to log whether and by whom health data have been accessed; measures to protect data security; arrangements to anonymize, archive or destroy health data once the research project has been completed; and other technical and organizational measures designed to ensure that the processing is carried out in accordance with the GDPR, together with processes to test and evaluate the effectiveness of such measures.

Cross Border Data Transfers

Data protection frameworks apply generally applicable restrictions to the transfer of health data for research purposes abroad, seeking to continue to guarantee a high level of protection. Because remote access to personal data is considered a transfer under the GDPR, the regulation's restrictions on cross-border data transfers apply to the consultation of national health repositories. Researchers thus face a range of obstacles to participation in cross border projects, and may be further limited by the consequences of the Court of Justice of the European Union decision in Schrems II. Some countries apply specific additional transfer limitations to health data. Under draft regulations, Israel, which typically applies liberal data transfer rules, requires localization of health data when used for secondary purposes, including research. India too has introduced draft legislation which would require data localization, in the form of retaining a copy of any dataset in the country if it is transferred abroad. Australia also requires localization of health records made available for research.

4. The scope of health data protected by HIPAA which only applies to certain covered entities such as health care providers and health plans, while a broad range of health data is covered by other jurisdictions.

The US generally does not restrict international transfers of data, but extra-territorial risks to compliance must be considered.

Conclusion

The fight against COVID-19 has highlighted the urgency of facilitating the utility of health data for scientific research while protecting individual rights to ensure trust and safety. This report examines the legal frameworks governing secondary use of health data in a number of countries. The legal and ethical frameworks are complex, comprising several layers of privacy, patient rights and public health regulations, at the national and state level. But the commonalities between frameworks outweigh the differences. All countries allow data research for compelling public interests such as public health. This includes repurposing data without patients' consent, subject to protections that include review by ethical committees, de-identification, contractual arrangements, and elevated security obligations as well as general data transfer restrictions and in certain jurisdictions, localization requirements. Significant variations in the manner in which these legal and ethical safeguards are structured and interpreted pose difficulties for research and collaboration. These differences may be bridged by a deeper mutual understanding of the commonalities.

Contents

Australia

What is the legal basis for secondary use of health data for research?

This section addresses the legal basis under both (a) general privacy law, and (b) law applicable to My Health Record.

General Privacy Law

Australia recognizes the right to privacy as a fundamental right, as reflected in the 13 Australian Privacy Principles (APPs) embodied in the 1988 Commonwealth Privacy Act (Privacy Act.) However, the right to privacy is not absolute; it is to be weighed against the rights of others and the importance of matters—such as medical research—that benefit society as a whole.

While medical research generally involves the informed consent of participants, Australian law does allow secondary use of identifiable data for research without consent pursuant to certain exceptions under APP 6.2 and 6.3 and in accordance with a defined legal framework established under the Privacy Act. The public interest in the research activity must substantially outweigh the public interest in maintaining the level of privacy protection afforded by the Australian Privacy Principles.

The Privacy Act sections 16B(2) and (3) permit the collection, use, and disclosure of health information for the purpose of research (or compilation or analysis of statistics) relevant to public health, without the data subject's consent, if (a) it is impracticable to obtain such consent, and (b) de-identified information will not achieve the purpose of the research or statistical compilation or analysis. Such collection, use, and disclosure must be in accordance with Guidelines issued by the National Health and Medical Research Council (NHMRC) under the Privacy Act s95[5] (for public sector) or s95A[6] (for private sector). Useful flowcharts for analyzing the applicability of the Guidelines are also available for public sector research and for private sector research. Where

5. **Guidelines under Section 95 of the Privacy Act 1988**, National Health and Medical Research Council, available here
6. **Guidelines under Section 95A of the Privacy Act 1988**, National Health and Medical Research Council, available here

applicable, the Guidelines are to be applied in addition to the National Statement on Ethical Conduct in Human Research.[7] In addition to being conducted in accordance with the Guidelines and the National Statement, the research project must be approved by a Human Research Ethics Committee.

In summary, to be permissible without consent, the collection, use, and disclosure of personal information for research relevant to public health requires that:

- The information is "necessary" for the research ("necessary" meaning more than helpful or convenient, but rather necessary in an objective, practical sense);
- The research outcome is relevant to public health or public safety;
- De-identified data is insufficient;
- Consent is impracticable (such as where contact information is unavailable, or consent would introduce selection bias);
- Specified detailed information about the proposed research has been submitted by the researcher to a Human Research Ethics Committee, and the Committee has approved the project (see below);
- Reasonable steps have been taken to de-identify the data before it is disclosed; and
- The researcher reasonably believes, and can justify that belief, that the recipient will not disclose the information or personal information derived from it.[8]

My Health Record

The My Health Record (MHR) system is a database administered by the Australian Digital Health Agency containing summary health information about all Australians, except those who have opted-out or cancelled their records. Individuals and healthcare providers have online access to the system. The My Health Records Act (MHRA)[9] includes privacy protections similar to those in the comprehensive Privacy Act, but the MHRA provisions are stricter and carry more onerous civil and criminal penalties for violations. As of late 2019, approximately 90% of Australians have MHRs.[10]

7. **National Statement on Ethical Conduct in Human Research, 2007 (updated 2018)**, available here
8. **Guide to Health Privacy**, Office of the Australian Information Commissioner, Sept. 2019, Ch 9, available here
9. **My Health Records Act 2012 No. 63.2012**, as amended, available here
10. **Implementation of the My Health Record System**, Australian National Audit Office, Nov. 25, 2019, available here

By default, secondary use of de-identified MHR data for public health and research is permissible on an opt-out basis, subject to a number of regulatory constraints applicable to the research project and researchers. An individual wishing to opt out of having her de-identified data used for public health and research can change her default settings in her MHR.[11]

For identifiable MHR data to be put to secondary use for research, specific consent is required. As of now, the functionality to consent to research involving identifiable data has not yet been built into the MHR system, so consent for identifiable data research must be granted through conventional research consent processes. In time, a dynamic consent process may be created so individuals could grant consent for defined secondary purposes on a case-by-case basis.[12]

Secondary use of MHR data for research is governed by the Framework to Guide the Secondary Use of My Health Record System Data,[13] which is interpreted and applied by the MHR Secondary Use of Data Governance Board. Using designated protocols, the Board makes decisions about granting researchers access to MHR data.

Secondary use of MHR data is subject to a number of restrictions:

- It cannot be used for solely commercial or non-health related purposes;
- It can't be provided to insurance companies or used for benefits eligibility;
- Health Research Ethics Committee approval must be granted (see below);
- MHR data released for secondary purposes must not be sold;
- Geographic limitations apply .

Does secondary use of health data for research require review boards' approval?

General Privacy Law

Before research can be conducted using personal information without the consent of data subjects, it must be approved by a Human Research Ethics Committee (HREC).

11. **How secondary use of My Health Record data can improve health outcomes for Australians**, Australian Digital Health Agency, available here
12. **Framework to guide the secondary use of My Health Record system data**, Australian Government Dept of Health, May 2018, available here
13. **Ibid**

HRECs conduct reviews of proposed research projects in accordance with the Guidelines under the Privacy Act s95[14] (for public sector) or s95A[15] (for private sector), as well as under the National Statement (especially paragraphs 2.3.9-2.3.12),[16] as described above.

The HREC approving the waiver of consent for a research project must be satisfied that:

- Involvement in the research carries no more than low risk to participants;
- The benefits from the research justify any risks of harm associated with not seeking consent;
- It is impracticable to obtain consent (for example, due to the quantity, age, or accessibility of records);
- There is no known or likely reason for thinking that participants would not have consented if they had been asked;
- There is sufficient protection of their privacy;
- There is an adequate plan to protect the confidentiality of data;
- In case the results have significance for the participants' welfare there is, where practicable, a plan for making information arising from the research available to them (for example, via a disease-specific website or regional news media);
- The possibility of commercial exploitation of derivatives of the data or tissue will not deprive the participants of any financial benefits to which they would be entitled; and
- The waiver is not prohibited by state, federal, or international law.[17]

My Health Record

The My Health Record (MHR) Secondary Use of Data Governance Board receives and assesses applications from applicants wishing to access MHR data. The Board's assessments are based primarily on the use of the data, not the user. The Board uses the "Five Safes" principles to evaluate applications.

14. **Guidelines under Section 95 of the Privacy Act 1988**, National Health and Medical Research Council, available here
15. **Guidelines under Section 95 of the Privacy Act 1988**, National Health and Medical Research Council, available here
16. **National Statement on Ethical Conduct in Human Research, 2007 (updated 2018)**, available here
17. **National Statement on Ethical Conduct in Human Research, 2007 (updated 2018)**, available here

For applications involving identifiable MHR data, the Board requires ethics approval from the AIHW Ethics Committee before data can be accessed or released. For applications involving de-identified data, the Board may require ethics approval. In addition, the Board requires an approved applicant to agree to the Conditions of Use Agreement (CUA) before MHR data can be released.

Who can access health data for research?

Any Australian entity (except an insurance agency) can apply to access MHR system data for secondary use, subject to meeting the criteria in the Framework. Applicants outside Australia may, in limited circumstances, be involved in the use of MHR system data for secondary purposes. They must be working in collaboration with an Australian applicant and the application shall generate public health benefits for Australians. The MHR data can only be directly accessed by the Australian entity and must be stored in a facility within Australia. The foreign party shall be subject to a Use Agreement and must meet Australian data management and security requirements.

Can health data be transferred outside the country?

MHR data that has been made accessible for secondary use must not leave Australia. However, there is scope for data analyses and reports produced from MHR system data to be shared internationally.[18]

What is the standard for anonymizing or de-identifying health data?

Data is considered de-identified where the risk of an individual being re-identified in the information is very low in the relevant release context (or data access environment). In other words, data is considered de-identified where there is no reasonable likelihood of re-identification.

De-identification involves two steps: (1) The removal of direct identifiers; and (2) Taking one or both of the following additional steps: First, removing or altering other information in the data set that could re-identify an individual; Second, using controls and safeguards in the data access environment to prevent re-identification.[19]

18. **Ibid**
19. **De-Identification and the Privacy Act**, Office of the Australian Information Commissioner available here

In 2017 the Australian government published a book containing extensive guidance to help organizations appropriately de-identify data. Adapted from a UK anonymization framework, the Australian framework sets forth a process for data custodians to identify and address the key factors relevant to their particular data sharing or release situation, including privacy risk analysis and control, shareholder engagement, and impact management.[20]

Other noteworthy aspects of secondary use of health data for research

- State and territorial laws may apply in addition to federal privacy laws. While these state and territorial laws generally apply only to the public sector, they may be relevant in terms of requirements applicable to public hospitals and universities and for entities receiving public funding. In addition, certain state and territory privacy laws extend specifically to private sector health data. In particular, note the New South Wales Health Records and Information Privacy Act, the Victoria Health Records Act 2001, and the Australian Capital Territory Health Records (Privacy and Access) Act 1997.[21]

- Australian authorities are increasingly aware of the potential value through data linkage, which is the linking and integration of disparate data sets to create a more robust understanding of health factors than would be available from single data sets. For example, researchers at the University of New South Wales are linking Medicare, pharmaceutical, mortality, and cancer data sets to analyze reasons for poor outcomes associated with prescription opioid use. Because of the inherent privacy risks involved with data set linkage, the government allows only a small number of accredited integrating authorities to conduct data linkage subject to strictly controlled safeguards. The Australian Institute of Health and Welfare is one of these accredited entities permitted to do high-risk data integration projects. Research data sets are frequently adjusted before release to researchers to further reduce privacy risks. [22]

---

20. **De-Identification Decision-Making Framework**, Office of the Australian Information Commissioner and Commonwealth Scientific and Industrial Research Organisation Data 61, 2017, available here
21. **Australia – Health and Pharma Overview**, Alexandra Wedutenko and Mathew Baldwin, 2020, available through One Trust Data Guidance
22. **Australia's Health 2018, ch. 2.5, Secondary use of health information**, Australian Institute of Health and Welfare, available here

- Studies indicate that Australians strongly support using medical records for research. A 2017 survey showed that 93% of Australians support using medical records for research, and more than 95% had moderate, high, or very high trust that researchers would use health data responsibly. While the public is willing to make its health information available for research, they do expect some control over how the data will be used. [23]

23. **Ibid**

England

What is the legal basis for secondary use of health data for research?

Scope

The United Kingdom has four separate national health systems—NHS England, NHS Wales, NHS Scotland, and NHS Northern Ireland. This analysis focuses primarily on the law in England governing secondary research uses of clinical data generated by NHS England.

High-Level Summary

Secondary use of clinical data for research is governed by a complex legal framework, including common law, statutory law, and government decisions and recommendations from health, ethics, and data protection authorities. Secondary use of identifiable medical data for research without consent can indeed be permissible, but only if the research fits within the statutory requirements that the research is in the public interest, the requisite process has been followed, the GDPR is complied with, and the data subjects have not exercised their new national right to opt-out of secondary uses of their health data.

Data Sharing in Practice

The majority of patient care is provided by NHS England. NHS Digital compiles certain patient data in a national database called "the Spine." NHS Digital allows authorized NHS staff, such as hospital and ambulance workers, to see a summary of each patient's information, called the Summary Care Record. The Spine links each Summary Care Record to a unique NHS number. Subject to multiple restrictions, discussed below, NHS Digital can make patient information available for research and other secondary uses through its Secondary Uses Service (SUS).[24]

24. **Secondary Uses Service (SUS)+ data: GDPR Information**, NHS Digital, available here

Hospital data from electronic health records and administrative data is collected nationally by NHS Digital,[25] which makes it available to researchers in depersonalized form[26] as Hospital Episode Statistics (HES). HES data releases are subject to strict disclosure rules, including a rule that the data must be fully anonymized before it can be publicly released. This hospital data has served as the evidence base for thousands of scientific journal articles.[27]

In contrast, data from doctors' offices has not been widely shared for research, at least not until recently. Although general practitioners (GPs) have electronic medical records (EHRs) containing extensive clinical data, this information has not been routinely collected at the national level and thus has not been widely available for research.

However, clinical data is rapidly becoming more available. Some local programs have been collecting limited GP data and making it available for secondary uses in depersonalized format subject to safeguards.[28] More importantly, large-scale anonymized data from GP clinicians is now available for pharmacovigilance (i.e., safety) research through the Clinical Practice Research Datalink (CPRD). The CPRD is the largest research database in the UK and one of the largest in the world. Submitted by GPs who choose to provide it, the data includes longitudinal patient information collected from the EHRs of approximately 15% of the UK population. This data is individually linked to data from other high-quality sources, including specialty care, HES hospital statistics, disease registries, and death records. These linkages of original clinical data with outcomes data are possible because of the unique NHS identifier number assigned to each NHS patient.[29]

---

25. NHS Digital was formerly the Health and Social Care Information Centre (HSCIC)
26. In UK parlance, depersonalized data is sometimes referred to synonymously with pseudonymized data, as defined by EU law. At times, though, depersonalized data seems to refer to data that has had direct identifiers removed, in contrast to pseudonymized data, which has had direct identifiers removed but still is attached to a code or key tying it to fully identifiable data. Regardless of the nuanced (and inconsistent) meanings of the terms, depersonalized data and pseudonymized data occupy a middle ground between Confidential Patient Information and anonymous data. Both types are still regulated by the GDPR, the Data Protection Act 2018, and the Common Law Confidentiality Duty.
27. **Using data in the NHS: the implications of the opt-out and GDPR**, King's Fund, May 25, 2018, available here
28. **Ibid**
29. **How Clinical Practice Research Datalink data are used to support pharmacovigilance**, Rebecca E. Ghosh, Elizabeth Crellin, Sue Beatty, Katherine Donegan, Puja Myles, and Rachel Williams, Therapeutic Advances in Drug Safety, 2019, available here

The CPRD supports international public health research by providing anonymized data to academic, government, and pharmaceutical researchers who can use it to assess safety risks for drugs and vaccines. The data includes "Real World Evidence" about demographics, diagnoses, symptoms, prescriptions, immunizations, and tests. For example, data on 1.3 million young children was analyzed by researchers who concluded that a meningitis vaccine was not linked to seizures, as had been suspected. Another extensive review of the data helped debunk the discredited reports linking the measles/mumps/rubella (MMR) vaccine to autism. Certain medication use linked to breast cancer has been identified. Because the linked data sets are so large and of such high quality and accuracy, they facilitate near real-time safety surveillance capability for new drugs and vaccines. The more quickly the huge CPRD data sets can be generated and analyzed, the faster safety signals can be acted upon. Even though the data is anonymized, the research is subject to the same scientific, ethical, and governance requirements applicable to all pharmacovigilance research.[30]

Given recent plans by NHS Digital to create a centralized consent mechanism for research consents,[31] pseudonymized patient data from GP records may become more available for research in the future. Such data will still be subject to the strict section 251 controls for secondary uses and the new national opt-out policy.[32]

Consent as the Default Requirement for Secondary Uses

In England, the general requirement that consent is required for secondary uses of medical data comes from the common law (i.e., case law made by courts.) Under the Common Law Duty of Confidentiality (CLDC), if someone shares personal information in confidence, that data may not be shared without her explicit consent unless another valid lawful exception permits the disclosure. As for the secondary use of health information, a potential exception exists to share that information if the public interest overrides confidentiality interests.[33] Numerous statutory and procedural controls define and limit when the public interest can override the CLDC consent requirement.

30. **Ibid**
31. **The Guide to the Sandbox (beta phase)**, ICO, 2020, available here
32. **Compliance with the national data opt-out**, NHS Digital, Ap. 27, 2020, available here; **The Guide to the Sandbox (beta phase), op.cit**.
33. **The common law duty of confidentiality**, UK Caldicott Guardian Council, available here

Disclosure for Secondary Research Uses Without Consent under Section 251

As stated above, English common law requires explicit patient consent for research involving Confidential Patient Information (CPI) unless the legal basis of an overriding public interest applies. The conceptual basis for such exception is codified by statute in section 251 of the NHS Act 2006 and in the Control of Patient Information Regulations. Section 251 allows CPI to be disclosed for research without consent if a Confidentiality Advisory Group (CAG) of the Health Research Authority (HRA) has, considering the legal requirements, approved the disclosure for the specific research project.

The safeguards that the CAG will consider when assessing an application for a section 251 disclosure include:

- The applicant must show that obtaining patient consent is not practical and anonymized information cannot be used;
- The activity must have a medical purpose, such as medical research that has received ethics approval from a Research Ethics Committee or the management of health and social care services;
- The activity must be in the public interest or in the interests of improving patient care;
- The activity must be consistent with the GDPR 2018;
- The application must undergo an annual review; and
- A mechanism to register and respect patient objections must be in place.[34]
- Section 251 authority does not apply to:[35]
- Disclosures for medical care (where consent is implied by law);
- Disclosures pursuant to explicit consent (such as participation in a clinical trial);
- Disclosures required by law;
- Disclosures for communicable disease surveillance and other risks to public health;
- Disclosures of anonymized data; and
- Disclosures for non-health services or purely profit-making purposes.

Data Protection Law Requirements

While approval by the CAG consistent with section 251 satisfies the exception for common law confidentiality duties, compliance with statutory data protection law is

34. **FAQs about the law**, NHS Health Research Authority, available here
35. **Ibid.; Using data in the NHS, op.cit.; National Data Opt-Out**, NHS Digital, available here

still required. The Data Protection Act 2018 and the GDPR thus apply to secondary uses of medical data for research. In fact, the GDPR's strict standard of "explicit consent" requires organizations to specifically describe the purpose for which data is being collected, makes it unlikely that consent can be the basis for secondary uses. Other legal bases should be relied on to justify the research use (generally, private controllers will rely on "legitimate interest" and public controllers will rely on "public interest" or, in the case of NIH, "legal obligation").[36] All other GDPR and DPA 2018 requirements for processing data, particularly those applicable to sensitive data, also apply to CPI released under section 251.

In addition, the recent draft Data Sharing Code of Practice from the UK Information Commissioner's Office (ICO)[37] recommends completing a Data Protection Impact Assessment (DPIA)—even if not mandated by law—and using formal data sharing contracts.

The New National Opt-Out

Introduction of Opt-Out. Since May 25, 2018, patients with medical records in the NHS England system have been able to use a new national opt-out system if they want to block the use of their CPI for secondary purposes, including research, subject to certain exceptions.

Scope of Opt-Out. The new opt-out only applies to secondary uses and disclosures, including for research, that would otherwise be permissible on a non-consented basis subject to all section 251 approvals and controls. In other words, a patient can now opt-out of having her CPI used in research even if that research is approved by the CAG, complies with section 251, and complies with GDPR and DPA 2018, as described above.

36. **Consent in Research—GDPR Guidance—What the Law Says**, NHS Health Research Authority, 2019, available here; **Secondary Uses Service (SUS)+ data, op.cit**.
37. **Data sharing code of practice: Draft code for consultation**, Information Commissioner's Office, 2019, available at here

The following are outside the scope of the opt-out:

- CPI used in individual health care (**i.e.**, the detailed local patient records or the national Summary Care Record);
- Disclosures of CPI to NHS Digital (such disclosures are mandated by the Health and Social Care Act 2012);
- Disclosures for invoice and payment purposes;
- Disclosures made pursuant to express informed consent, such as for participation in a clinical trial;
- Disclosures for the protection of public health, including CPI used to diagnose or control communicable disease, deliver and monitor vaccination programs, and manage risks from environmental, food, or water contamination;
- Disclosures made pursuant to legal mandate;
- Disclosures in exceptional circumstances where the public interest overrides confidentiality interests, subject to requirements and guidelines from the NHS Information Governance Alliance;
- Disclosures for specific national purposes, including disease registries and patient experience surveys (some of which have their own opt-outs); and
- Data that has been anonymized in accordance with the ICO's Code of Practice on Anonymization or is in aggregate form.

How opt-outs are managed. Any patient (age 13 or older) who has received care through NHS England can record her opt-out through one of several channels: online, by phone or paper, or through the NHS app.[38] A patient's opt-out is recorded and associated with her NHS number on the NHS Spine. Unless changed by the patient, the opt-out remains in place, even after the death of the patient. Once NHS England receives an opt-out, the entire medical record associated with that patient must be fully removed from the data set otherwise available for section 251 uses, including research; simply removing identifiers is not sufficient.

Who must comply with the opt-out. All organizations providing or coordinating publicly funded health or adult social care in England must comply, even if the organization is headquartered outside England. This includes publicly funded care, private care given in NHS settings, and private care coordinated or commissioned by NHS. While NHS Digital and certain other large organizations have been compliant

38. The opt-out question presented to patients can be viewed at **National Data Opt-Out Operational Policy Guidance Document**, vers. 4.0, NHS, 2019, p. 15, available here

with the opt-out since 2018, the deadline for GPs to comply was set for March 2020. That deadline was extended to September 30, 2020, because of the COVID-19 pandemic.[39] Failure to comply with the opt-out can be deemed a violation of the GDPR and DPA 2018, sanctionable by the ICO.

History of the opt-out. Prior to the new national opt-out, there were certain other national and local opt-outs, which have been transitioned into the new national opt-out. An initiative called care.data was launched in 2013 to upload records from GPs into national health databases in order to make them more available for care coordination, planning, and—on an anonymized basis—research. But the program generated considerable controversy related to privacy concerns, ineffective public communications, and doctors' criticisms. Because of the opposition, care.data was postponed in 2014 and eventually cancelled in 2016.[40] The National Data Guardian, Dame Fiona Caldicott, was tasked by the government with recommending a new model for national health data governance. In 2017, Dame Fiona published her review. Her report stressed the vital importance of sharing medical data for research to improve medical care, while also recommending a new national opt-out model for secondary uses, which is what has now been operationalized.[41]

Does secondary use of health data for research require review boards' approval?

Review of research projects by Research Ethics Committees (RECs) is generally required. RECs are tasked with safeguarding the rights, safety, dignity, and well-being of research participants. Independent of research sponsors, RECs review research proposals and opine on whether the research projects meet ethical standards, including data confidentiality and security.[42]

39. **National Data Opt-Out**, NHS Digital, available here
40. **Using data in the NHS, op. cit**.
41. **National Data Guardian for Health and Care 2017 report: Impact and influence for patients and service users**, National Data Guardian, 2017, available here
42. **Research Ethics Committee – Standard Operating Procedures**, NHS Digital Health Research Authority, 2019, available here

The NHS Health Research Authority has a convenient online tool for determining if review by a REC is required. REC reviews are separate from and in addition to reviews by the CAG, which are required if identifiable or depersonalized/pseudonymized data is to be used, pursuant to section 251.

Who can access health data for research?

Even though most research by NHS and others uses depersonalized/pseudonymized data, safeguards and restrictions still apply. Researchers seeking access to NHS data must submit an application to NHS Digital's Data Access Request Service (DARS). They must execute a data-sharing agreement and comply with strict information governance protocols. Researchers can review requirements and apply for NHS data online.[43] Pursuant to new legal restrictions in the Care Act 2014, NHS data can be made available only for research that supports health or social care, not for purely commercial reasons.[44]

Can health data be transferred outside the country?

As stated above, the GDPR and the DPA 2018 still apply, even where all requirements of section 251 have been met and the patients whose data is in the research data set have not opted out of secondary uses. Therefore, personal data used in research can be transferred out of the European Economic Area (EEA) only to the extent permissible under the GDPR and the DPA 2018.

Note: Although Brexit occurred January 31, 2020, the GDPR still applies in the UK during the transition period through December 31, 2020. Since the ICO expects the GDPR to be replaced by a UK GDPR, the geographic restrictions on research data will likely still apply.[45]

What is the standard for anonymizing or de-identifying health data?

Anonymized data is not subject to the CLDC and thus can be disclosed for research without adhering to the strict safeguards and requirements of the section 251 process. Such data is thus not subject to any of the research or privacy laws discussed here –

43. DARS applications are received here
44. **Using data in the NHS, op. cit.**
45. **Information Rights at the end of the transition period FAQs,** ICO, July 10, 2020, available here

the CLDC, section 251, the national opt-out, or the GDPR/DPA 2018. Similarly, data that is in aggregate form is not covered by these laws.

According to the GDPR, anonymous information does not "relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."[46] But the ICO recommends caution when attempting to anonymize data because the mere removal of identifying data elements will rarely be sufficient. If the data holder has any reasonably available means of re-identifying the individual data subjects, the data has been merely pseudonymized, not anonymized. As such, it remains subject to data protection laws and other limitations on sharing for secondary uses.[47]

In 2016, the ICO published a Code of Practice for Anonymization that lays out principles of appropriate anonymization. Data is anonymized only if it does not itself identify any individual, and if it is unlikely to allow any individual to be identified by being combined with other data. The Code includes an explanation that an organization conducting anonymization should consider the "motivated intruder" test, which assumes that data remains identifiable if a motivated intruder determined to re-identify individuals in a data set, who has access to extensive public information in libraries and online and is willing to make inquiries, would likely be able to succeed in re-identifications. The Code also recommends regular re-identification testing for anonymized data sets.[48]

To be clear, the ICO does not require that the risk of re-identification for anonymized data be zero, only that the risk be remote. The ICO points out that anonymization is a valuable Privacy-by-Design practice that protects individuals while making vital data publicly available. In fact, anonymization is crucial for government bodies, such as health entities, that are required to both protect privacy and publish information about health service operations, disease incidence, and patient outcomes.[49]

The ICO recognizes that identifiability exists on a spectrum from fully identifiable at one end to fully anonymous at the other end, with depersonalized information in between. If data has had personal identifiers removed, but that process could

46. Recital 26, General Data Protection Regulation, European Union, 2016
47. **What is personal data?,** ICO, 2020, available here
48. **Anonymisation: managing data risk code of practice,** ICO, 2012, available here
49. **Ibid.**

be reversed in order to re-identify someone in the data set, the data is considered depersonalized (or pseudonymized), and it thus still must be subject to safeguards and legal protections.[50] While such data is not anonymized, it may still be available for secondary use under the legal framework discussed in this paper, subject to controls appropriate to the setting.

The UK has created the UK Anonymisation Network through which experts can share best practices and insights about anonymization techniques. The Network emphasizes that organizations disseminating anonymized data make best efforts to ensure that re-identification will be "an extremely hard problem for the data intruder." They recommend that those disseminating data follow state-of-the-art best practices of statisticians and computer scientists, including security measures, minimizing the risk of re-identification while maximizing the value of the data released, and having plans to deal with the small risk of a re-identification event. The Network cautions that an excessive degree of anonymization can make data less useful or, worse, lead researchers to erroneous conclusions or inferences.

The Network concludes that anonymization is "a very valuable tool that allows data to be shared, thereby exploiting its huge social and economic value, whilst preserving privacy."[51]

Other noteworthy aspects of secondary use of health data for research

- Surveys in the UK show that while the public is concerned about the privacy of medical records, there is also a broad understanding of the importance of using information in medical records for valuable public purposes like research and epidemiology.[52] Citizens want their information to be put to good use, but they also care about risks to their privacy and want to be kept informed about how the data is used. In addition, since the NHS is the source of most health data in the country, health data is subject to data transparency imperatives and pressures, such as the government's 14 Public Data Principles. These Principles include that government bodies should actively encourage the re-use of their public data,

50. **Understanding Patient Data: Identifiability Demystified,** Wellcome Trust, 2020, available here
51. **What is anonymisation,** UK Anonymisation Network, 2020, available here
52. **Anonymisation Standard for Publishing Health and Social Care Data, Supporting Guidance: Drawing the line between identifying and non-identifying data,** NHS Information Centre for Health and Social Care, 2013, available here

publishing the data itself on a non-personal basis under open licenses that enable free re-use, including commercial re-use.[53]

- As the national opt-out for secondary uses has been implemented, considerable efforts have also been underway to educate the public about how their health data is protected and used. In addition to the significant public educational materials on the NHS Digital and ICO websites, the Welcome Trust has, with government funding, created extensive consumer-friendly materials at a website called Understanding Patient Data. This site contains infographics, tools, and explanations about how patient data is protected by law and how it is put to valuable societal purposes. NHS England also publishes extensive materials explaining the vital importance of patient data for beneficial purposes such as understanding disease, improving diagnosis, advancing safety, and planning and measuring the effectiveness of NHS services.[54]

  Based on a clinical treatment model sometimes called "the learning healthcare system," Understanding Patient Data has also recently proposed a new data research governance model where data that has been accessed for research could follow an iterative pathway, shifting from a one-way pipeline from clinical sources to a feedback cycle involving the reporting back of outcomes. The idea is that data-based research outcomes would be reported back to a panel that scrutinizes previous data access decisions and their outcomes in order to better inform new decisions about how to appropriately share data based on real-world evidence and outcomes.[55]

- Efforts are being made to expedite data sharing for secondary use. In addition to the national opt-out whereby patients can set controls to prevent the sharing of their data for research and other secondary purposes, NHS Digital has just been selected to participate in the ICO Sandbox in order to develop a central mechanism for collecting and managing patients who wish to express their consent for research and regulated clinical trials.

53. **Anonymisation Standard,** NHS Information Centre, **op.cit.**
54. **Using data in the NHS, op.cit.**
55. **A new approach to decisions about data**, Understanding Patient Data, Welcome Trust, July 8, 2020, available here

- Some policymakers and researchers are raising concerns about inadvertent and potentially harmful consequences of the national opt-out for secondary uses of patient data. Numerous government operations, including NHS resource allocation and treatment protocols, rely on analyses arising from patient data made available under section 251. The risk is that such analyses may become less accurate and useful, leading to inappropriate treatment and deployment of government services. Selection bias due to variation in opt-out rates is well-documented; see, for example, a recent peer-reviewed analysis finding that substantial geographic variation in opt-out rates limits research accuracy, particularly for small-area studies involving environmental hazards.[56] Disparate opt-out rates based on age, location, condition, and ethnicity can all cause distortion.[57] Particular concerns have been expressed by UK cancer charities about how patient opt-outs may block improvements in cancer care.[58] Researchers at the King's Fund, noting that even a low opt-out rate can have a detrimental impact on certain kinds of analysis, express concern that easier online access to the national opt-out mechanism may increase the opt-out rate. They caution that skewed opt-outs could introduce bias into machine-learning algorithms, which could unintentionally serve to provide poorer care and services to the populations with a higher opt-out rate. Rising opt-outs may also undermine the accuracy of findings derived from linked, pseudonymized data sets. The King's Fund notes that a high degree of public trust in NHS and research is needed to help minimize opt-out rates - and the resulting scientific inaccuracies.[59]

56. **The challenge of opt-outs from NHS data: a small-area perspective,** Frédéric B. Piel, Brandon L. Parkes, Hima Daby, Anna L. Hansell, Paul Elliott, Journal of Public Health, 2019, available here
57. **What the national data opt-out means for researchers,** NHS, 2019, available here; King's Fund, op.cit.
58. **Exclusive: Patient record opt-outs risk blocking improvements in cancer care,** Sharon Brennan, Health Service Journal, 2017, available here
59. **Using data in the NHS,** op.cit.

Finland

What is the legal basis for secondary use of health data for research?

High-Level Summary

Finland aims to be a leader within the EU regarding the secondary use of clinical data. In 2019, Finland enacted a law explicitly intended to clarify the legal basis for the broader secondary use of clinical data under the GDPR. The law aims to encourage broad use of its medical data for socially valuable purposes by enabling data users to have a "one-stop-shop" for requesting access to such data in an efficient, consolidated, and speedy fashion. The new data management and access system is designed to be GDPR-compliant and highly secure, with almost all access being allowed only on a remote basis. Prior to this law, data access for secondary use was cumbersome and slow, which led to the under-utilization of Finland's exceptionally comprehensive and high-quality data resources.[60] The new legal framework and data access system are intended to benefit the public with better, safer, and more cost-effective medical treatment, technology, and services by advancing research more accurately and quickly through the unlocking of complex and varied data sets.[61]

The New Law and Data Permit Authority (Findata)

The Finnish Act on the Secondary Use of Health and Social Data (the Act) took effect May 1, 2019, after years of development. The law was crafted to enable efficient and secure secondary processing of health and social data by creating a better-defined legal basis for such processing under the GDPR and national data protection act (DPA).[62] The other key objective of the Act was to create a new government agency, Findata, and task it with collecting and managing personal data from numerous

60. **Frequently asked questions,** Finnish Ministry of Social Affairs and Health (MSAH), available here
61. **How Finland is Pioneering the Use of Health Data for Secondary Purposes,** Mario Romao, Intel, July 22, 2019, available here; **New legislation will speed up the use of Finnish health data,** Saara Malkamäki, Sitra, May 14, 2019, available here
62. **Data Protection Act (1050/2018),** [unofficial translation available here]

disparate sources, securing the data, and efficiently processing requests from third parties to access the data.[63]

The secondary uses of health and social data that are enabled by the Act are:

- Statistics
- Scientific research
- Development and innovation activities (including in the private sector)
- Education
- Knowledge management (broadly defined to include processing by service providers to support operations, production, financial control, management, and decision-making)
- Steering and supervision of social and health care by authorities; and
- Planning and reporting duties of authorities.[64]

Operating as a new, separate Data Permit Authority within the National Institute for Health and Welfare, Findata collects and coordinates data from different sources. These extensive sources include Kanta (the national database containing prescription and patient data and health and social welfare data archives) and various healthcare and other registers, including the Finnish Institute for Health and Welfare (THL), the Social Insurance Institution of Finland (Kela), the Population Register Centre, and the Finnish Centre for Pensions, and Statistics Finland.[65]

Findata offers a "one-stop-shop" for entities seeking data for the secondary purposes specified in the Act. If the researcher is seeking statistical data only, it follows Findata's process for data requests and is granted the data if qualified. If the researcher needs pseudonymized personal data, it files a Findata application for a data permit. If a data permit is granted, Findata gathers the data requested from multiple sources, combines the data, pseudonymizes (or anonymizes) it, and attaches logging and other metadata. The permit holder then enters into a data agreement with Findata and is allowed remote access to the pseudonymized or anonymized data on virtual machines. The permit holder does not receive a copy of the data requested, except in rare cases where absolutely necessary and subject to strict security controls.[66]

63. **Act on the Secondary Use of Health and Social Data (552/2019),** available here
64. **Ibid**
65. **FAQ,** Findata, available here
66. **Services for customers**, Findata, available here

If the data permit application only concerns data stored in the personal data registers of a single organization, the organization will be responsible for making the decision on granting access to it. In this case the organization will be responsible to collect the data, and as necessary combine, pre-process and pseudonymize it before disclosing the data to the permit holder. But, if the data requested needs to be anonymized, this will be the responsibility of Findata.

Increased speed and improved scientific accuracy

Prior to the new law and the existence of Findata, researchers and others who needed to analyze data from a number of sources had to apply to each data controller separately, a complicated and bureaucratic process that often took years.[67] Now, Findata processes data permits within just three months, with one possible three-month extension for exceptionally complex data requests. Moreover, by streamlining the consolidation of numerous disparate data sets, Findata will make vastly more data available—essential for technologies like artificial intelligence that rely on voluminous, high-quality data in order to discover insights needed for disease prevention and treatment.[68]

Restrictions on Data Access

The Act requires Findata to comply with the GDPR requirement of data minimization. Accordingly, after Findata gathers and combines the data requested, it makes it available to the applicant in the following order of priority:

- Aggregate data;
- Individual-level data that has been anonymized;
- Personal data that does not contain direct identifiers;
- Pseudonymized personal data that does not contain direct identifiers but does contain codes enabling linking to direct identifiers; or
- Data containing personal identifiers—which can only be supplied in rare and well-justified circumstances.[69]

Only aggregate data is freely available to the public. All forms of personal data, including pseudonymized data, are available only in a secure user environment,

67. **FAQ,** Findata, **op. cit.**
68. Romao, **op. cit.**
69. **Frequently asked questions,** MSAH, **op. cit.**

subject to contractual restrictions. Companies requesting data for innovation and development activities can only be given access to aggregate, not individual-level, data. Data cannot be used for marketing or insurance purposes.

Findata does not sell data. It does, however, charge fees for its collection, extraction, processing, pseudonymization, and customization services.[70]

Legal basis for processing

Consent of the data subject is not the legal basis for processing for secondary uses under the Act. Instead, the legal basis for Findata to process and grant restricted access to data for secondary purposes is GDPR Article 6(1)(e)—processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, and in the case of health data, Article 9(2)(g)—substantial public interest, with suitable data protection safeguards.[71] Additionally, sections 4 and 6 of the national data protection act add safeguards.

The Act does not apply to data involved in clinical trials, nor is Findata involved with such data.[72]

Security Controls

Findata is required to disclose data in a way that maximizes data protection and minimizes disclosure. After approving a data permit, it compiles the data stored from original sources, enters into a contract granting temporary data access to the applicant, and then provides the data inside Findata's secure data environment. After the processing is complete, Findata erases the data compiled. In addition, Findata sets access rights in line with the issued permits, monitors access to devices, systems, and office sites, deploys security controls to prevent unauthorized access, and monitors and restricts data communications. Findata also has extensive logging systems that record all data-processing events.[73]

70. **FAQ,** Findata, **op. cit.**
71. Data Protection Act (1050/2018), available here
72. **FAQ,** Findata, **op. cit.**
73. **Data protection and the processing of personal data,** Findata, available here

Development of the Act and Findata

The Act and the Findata data access system benefitted from a collaborative development process involving academia, government, business, and associations. In contrast to the ubiquitous approach where legislation is first drafted and enacted, and then public and private entities implement and react to it, in the case of the Act, private and public entities collaborated on implementation planning simultaneously with the legislative process. As stated by a staff person at Sitra[74], which played a key role in developing the new framework, this novel parallel track of implementation and legislation "took a lot of courage, some risk-taking and a strong belief in the ability to succeed."[75] Sitra was also instrumental in the Isaacus project, which spent several years piloting and testing service models, metadata descriptions, data lakes, and collaborative models with authorities and stakeholders. The process included international involvement as well, with EU member states and other countries providing valuable recommendations about best practices.[76]

Does secondary use of health data for research require review boards' approval?

The Act itself does not require use of review boards to approve secondary use of clinical data. However, other law frequently does require approval of research projects by review boards. In the case of research involving health or medical research, the Medical Research Act (488/1999) should be consulted to determine if ethical review is required. For research involving human participants in the natural sciences, technology, the arts, and in some cases, non-invasive health research, the national guidelines established by the Finnish National Board on Research Integrity TENK should be consulted. TENK's ethical principles, which have been adopted by almost all universities and research institutions require research to be reviewed by a human sciences ethics committee under specified circumstances.[77]

74. Sitra is an independent public foundation operating under the authority of the Finnish Parliament.
75. **New legislation will speed up the use of Finnish health data, op. cit.**
76. **Ibid**
77. **The ethical principles of research with human participants and ethical review in the human sciences in Finland,** Finnish National Board on Research Integrity TENK guidelines, 2019, available here

Who can access health data for research?

Permit holders approved by Findata may access data under a time-limited license granting access under the conditions required by Findata. Such access is almost always through Findata's secure remote site. Other restrictions apply, including that companies seeking data for innovation and development activities may also gain access to aggregated data.

A data permit is not needed for aggregate data. Applicants seeking aggregate data only need to provide a compliant data request to Findata.[78]

Can health data be transferred outside the country?

The Act allows permit holders to be outside Finland. However, because remote access to personal data is considered a transfer under the GDPR, the GDPR's restrictions on cross-border data transfers apply. While anonymized data can be accessed or transferred without regard to borders, pseudonymized and other personal data can only be accessed from or transferred to countries in the European Economic Area[79], unless one of the GDPR's possible grounds for transfer exists (**i.e.**, adequacy, standard contractual clauses with appropriate security controls, Binding Corporate Rules, or rarely, an Article 49 exception).[80]

What is the standard for anonymizing or de-identifying health data?

The GDPR's definitions of anonymized and pseudonymized data apply. The DPA does not apply to anonymized data. As described above, Findata will seek to supply aggregate or anonymized data where such can suit the needs of the data applicant.

78. **Frequently asked questions,** MSAH, **op. cit.**
79. The European Economic Area is comprised of EU countries, plus Norway, Liechtenstein, and Iceland.
80. **FAQs,** Findata, **op. cit.**

Other noteworthy aspects of secondary use of health data for research

- In addition to compiling data per specific requests by applicants, Findata is preparing "ready data"—data that is pre-collated based on the expectation that it will be useful to data seekers. This will make data access quicker and more efficient for researchers and analysts who can use this data.[81]

- The Act and Findata are just a part of a broad reform effort in Finland intended to advance medical research. The 2012 Biobank Act has been heralded in international medical circles for enabling a broader use of biological samples than in other European countries, while protecting the rights of sample donors. The Biobank Act allows broad consent for medical research, rather than separate consents for each research project.[82] Amendments are being considered to further increase the availability of biological samples for genetic research on an opt-out basis.[83] The FinnGen project aims to collect genome data, integrate it with health registry data on about 10% of the population, and make it available to generate therapeutically relevant insights into the root causes of common diseases and their response to therapies.[84]

- Sitra, the Finnish think tank and foundation mentioned above, has moved on from laying the foundation for Findata to other health-related data projects. Sitra's IHAN initiative aims to build legal, technological, cultural, and organizational underpinnings for modern, privacy-protective, and ethical data exchange. Among the IHAN health projects is its Patient-Centric Real-World Data effort, which is testing the ability of pharmacies to act as trusted collectors of information between patients, data controllers, and pharmaceutical companies to augment prescription data with clinical data and self-generated patient feedback about their experiences.[85]

---

81. **Frequently asked questions,** MSAH, **op. cit.**
82. **Finnish biobanks one step ahead of others,** Good News from Finland, Oct. 16, 2014, available here
83. **Secondary use of health data – the new Finnish Act,** Roschier, available here
84. **Finland: A framework for genetic research,** Open Access Government, June 1, 2020, available here
85. **Update of the European Data Market—SMART 2016/0063—Story 6, The Secondary Use of Health Data and Data-driving Innovation in the European Healthcare Industry,** Jan. 21, 2020, available here

France

What is the legal basis for secondary use of health data for research?

In France, the use of personal data is governed by several legal sources including the General Data Protection Regulation (GDPR), the French Data Protection Act[86] (FDPA) and the Public Health Code (CSP) law as amended on January 26, 2016.

Since 2016, France has adopted an innovative system to provide broader access to health data for research purposes. France's stated goal is to "modernize our health system" and create "the conditions for direct access to health data".

The French Data Protection Act (FDPA)

Articles 64 et seq. of the FDPA specify additional conditions for the processing of personal health data to those defined by the GDPR. These include processing for a public interest in compliance with certain standard methodologies. Article 65 of the FDPA excludes from these conditions the processing of personal health data referred to in Article 9.2 of the GDPR, including processing for healthcare research.

The Public Health Code

The CSP incorporates all of the databases in the health sector into a single centralized system known as the "National Health Data System" (SNDS) and authorizes data collection and processing to improve care through research and innovation. Created on April 10, 2017, the SNDS pools pre-existing databases in order to contribute to two main goals:

- To advance knowledge in healthcare;
- To broaden research in healthcare for the public good.

Based on the French Data Protection Authority's (CNIL) opinion, a decree of December 26, 2016, governs the conditions for access to data and establishes a procedure for authorization. The CSP also establishes the Health Data Hub (HDH – previously INDS), which is

86. **Loi "Informatique et Libertés"** no. 78-17, 6 Jan. 1978

tasked with the management of the SNDS.

The SNDS does not contain direct identifiers about patients, such as names, addresses or social security numbers.

Notice about the establishment and purposes of the SNDS and the possible re-use of healthcare data for research purposes is provided:

- On the websites of hospitals, health insurers and mutual insurance companies, and other healthcare providers;
- On posters physically posted on the premises of healthcare providers or through documents provided to patients.

Individuals have a right to opt-out of having their SNDS data be used for research purposes, except for certain public goals such as the mitigation of an epidemic or monitoring of health services. In addition, individuals may exercise their GDPR rights of access, rectification and objection to the processing of personal data against their health insurer.

There are two types of access to the SNDS:

Permanent access is granted to a specified list of public bodies, including the General Directorate of Health, the Regional Health Agencies, the National Agency for the Safety of Medicines and Health Products, the National Cancer Institute, etc.

Restricted and occasional access may be granted to other private organizations or public organizations by (a) self-declaration of compliance with the CNIL's published Reference Methodology; or (b) authorization of the CNIL for research purposes that meet a public interest.

Access to Health Data under a declaration of compliance with a reference methodology

Data processing for the purposes of health research may be conducted without prior authorization from the CNIL under the following cumulative conditions:[87]

87. L. no. 78-17, 6 Jan. 1978, art. 73

- the process must be in compliance with one of the reference methodologies[88] approved and published by the CNIL in consultation with the HDH;
- the data controller must send the CNIL a signed declaration attesting to its compliance with the reference methodologies.

Access granted through an approval process

To approve a research project that does not comply with the reference methodologies, the CNIL will require an applicant to produce an impact assessment to examine authorization requests with a public interest purpose.

The CNIL's authorization will be provided only after opinions by:[89]

- the competent committee for the protection of persons (CPP) when the request relates to human subject research;
- the ethical and scientific committee for research, studies and evaluations in the health field (CESREES), when the requests relate to research not involving a human subject.

Assessment of the Public Interest by the CESREES

The Comité Éthique et Scientifique pour les Recherche (CESREES) is authorized to assess the scientific relevance and public interest of a research project requiring access to health data prior to its submission to the CNIL for approval. The CESREES will consider criteria such as feasibility, quality, potential bias, the expectations of the data subjects etc. Once CESREES has given its opinion, the file may be submitted to the CNIL for authorization.

88. MR-00: involving the human person for interventional research. MR-001 requires the express and informed consent of the patient or that of his/her legal representatives. MR-003: involving the human person for non-interventional research. This methodology does not imply the consent of the patient, but information to the patient is mandatory. MR-004: research not involving the human person, related to studies, evaluations in the health field.
89. L. no. 78-17, Jan. 6, 1978, art. 76

Assessment of the Public Interest by the CNIL

The law does not define the concept of "public interest", which justifies the processing of sensitive data.[90] Nevertheless, recent regulatory developments, particularly in the context of health data research projects, provide additional guidance and interpretation. [91]

The CNIL and the HDH have not provided explicit guidelines or a definitive opinion on the exact criteria. However, the INDS, after having commissioned a study on this concept,[92] proposed to formulate criteria;[93] and the CNIL had the opportunity to recognize the existence of a public interest in several of its decisions.

The CNIL held that the pursuit of a public interest is not reserved to only public entities.[94] The INDS shared this position, stating that a commercial interest is not a priori incompatible with a public interest. Hence, the nature of an applicant as a health establishment, consultancy, academic research unit, or healthcare company is not dispositive. The INDS considered that "studies, research or evaluations at the request of or for the benefit of public authorities are presumed to be in the public interest".[95]

With regard to the applications for access to the SNDS, the INDS has recognized the following as research of public interest:[96]

- the improvement of public health;
- the improvement of the health care system;
- research and the contribution to scientific knowledge;
- the potential contribution to the public interest of a study pursuing private interests.

The CNIL considered databases set up by public or private organizations for research

90. L. n° 78-17, 6 janv. 1978, ancien art. 31
91. Conv. Constitutive du GIP national « Institut national des données de santé » (INDS), art. 11, prévoyant un Comité d'Expertise sur l'Intérêt Public (CEIP)
92. INDS, Rapp. D'expertise juridique sur l'intérêt public dans le contexte des données de santé, 29 juin 2017 www.indsante.fr
93. INDS, Principes d'appréciation de l'intérêt public, 11 févr. 2019 ; Éléments de réflexion pour la définition des finalités d'intérêt public, 12 juill. 2019 : www.indsante.fr
94. CNIL, délib. n° 2017-285, 26 oct. 2017. – CNIL, délib. n° 2017-347, 21 déc. 2017. – CNIL, délib. n° 2018-289, 12 juill. 2018
95. INDS, **Principes d'appréciation de l'intérêt public, 11 févr. 2019,** p. 5
96. INDS, **Éléments de réflexion pour la définition des finalités d'intérêt public,** 12 juill. 2019 : www.indsante.fr

purposes as advancing a public interest. It noted that studies facilitated by such data warehouses can enable "both private and public players to conduct research in the medical field or to enlighten public authorities in their decision-making in terms of health policy". [97]

Similar to GDPR[98], French law provides that data from the SNDS may not be processed for the purposes of:

- the promotion of health products to health professionals or health institutions;
- the exclusion of insurance coverage or modification of insurance contributions or premiums for an individual or a group.[99]

Does secondary use of health data for research require review boards' approval?

Yes, under the conditions defined above and in particular:

- the CESREES on the nature of the public interest of the research project under consideration;
- the CPP in the case of human subject research.

Who can access health data for research?

Any person or entity, public or private, for-profit or not-for-profit, will be able to access SNDS data with prior CNIL authorization or a declaration of compliance with the Reference Methodology, to carry out a study, research or evaluation in the public interest.[100]

Can health data be transferred outside the country?

97. CNIL, délib. n° 2017-013, 19 janv. 2017. – CNIL, délib. n° 2017-285, 26 oct. 2017. – CNIL, délib. n° 2017-347, 21 déc. 2017. – CNIL, délib. n° 2018-289, 12 juill. 2018. – CNIL, délib. n° 2018-295, 19 juill. 2018. – CNIL, délib. n° 2018-369, 20 déc. 2018. – CNIL, délib. n° 2019-103, 5 sept. 2019. For example, monitoring and conducting epidemiological and medico-economic studies, analyzing current medical practice in oncology; changes in the consumption of anti-cancer treatment, pricing and access to treatment by patients; better knowledge of innovative anti-cancer treatments; evolution of an epidemic; changes in drug consumption; evolution of the management of a population suffering from the same pathology.
98. Reg. (EU) 2016/679, 27 Apr. 2016, cons. 54
99. CSP, art. L. 1461-1
100. Article L1460-1 Code de Santé Publique

Yes, transfers are possible under the rules of the GDPR as transposed into French law.

Authorization from the CNIL is required when appropriate safeguards are provided through contractual clauses or provisions to be included in administrative arrangements between public authorities.

What is the standard for anonymizing or de-identifying health data?

Since the SNDS contains sensitive health data, the CNIL requires data protection guarantees in accordance with the FDPA.

The decree of March 22, 2017, sets out the rules for the secure management of the SNDS relating in particular to pseudonymization and identifiablity. For example, data contained in the SNDS must be pseudonymized in an irreversible manner in order to preserve the privacy of individuals. The CNIL is authorized to certify the anonymization process.[101]

Anyone processing non-anonymous SNDS data is subject to a duty of professional secrecy under the penalties provided in Article 226-13 of the Penal Code.

Users must make enforceable commitments to comply with the general conditions of use of the SNDS, including commitments of:

- confidentiality, in particular the non-dissemination of non-anonymous data;
- not re-identifying the data;
- maintaining security;
- not pursuing a prohibited purpose of the SNDS.

Adequate sanctions must be provided for in the case of non-compliance with these commitments, including the termination of access to the data.

Furthermore, only anonymous datasets may be exported from the SNDS to a system that is not part of the extended SNDS (through the establishment of a specific agreement ensuring data security). If non-anonymous data from the extended SNDS is transmitted over an uncontrolled network, it must be encrypted according to the findings of a risk analysis.

101. L. no. 78-17, 6 Jan. 1978, art. 8, I, 2o, i

India

What is the legal basis for secondary use of health data for research?

Two draft data protection and health data bills are currently pending in India and are the subject of considerable attention. This chapter will distinguish between the legal framework for secondary research uses under existing law and under the draft legislation.

Existing Law Framework

The collection, use, management, storage, and transfer of personal data in India is currently governed by several major laws:

- The Information Technology Act, 2000 (the IT Act)
- The Information Technology (Amendment) Act, 2008
- The Information Technology (Intermediaries Guidelines) Rules, 2011
- The Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011 (SPDI Rules)

In addition to the safeguards, rights, and restrictions that this legal framework applies to "personal data", existing law imposes extra controls on "sensitive personal data or information (SPDI)". SPDI includes physical, physiological and mental health conditions, as well as medical records and history. SPDI do not include personal data that is freely available or accessible in the public domain or furnished under any law. [102]

In general, and in particular for SPDI, Indian law relies heavily on consent as the legal basis for data processing. In Indian data protection law (both existing law and the proposed new legislation, consent serves a more essential function than in the GDPR, which relies also on other legal bases for processing.

102. **Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules,** 2011, Ministry of Communications and Information Technology, (the SPDI Rules), Ap. 11, 2011, available here

Specifically, a body corporate[103] collecting SPDI must obtain specific written consent of the data subject regarding the purpose of the SPDI, and ensure that the data subject has knowledge of:

- The fact that SPDI is being collected;
- The purpose of the collection;
- The intended recipients of SPDI; and
- The name and address of the agency collecting SPDI and the agency retaining it.[104]

Important as consent is, it alone is not sufficient. Even with consent, a body corporate may collect SPDI only for a lawful purpose connected with a function or activity of the body corporate or any person on its behalf, and the SPDI must be considered necessary for that purpose.[105]

The SPDI Rules also require that the SPDI collector give the data subject an option to not provide SPDI, withdraw her consent, and access and correct her SPDI. The SPDI collector must provide a detailed privacy policy, appoint a grievance officer, and adhere to prescribed security standards. SPDI may be retained only as long as required by law or for the purposes for which it was lawfully collected. Violations of the IT Act and SPDI Rules are subject to civil fines, imprisonment for up to three years, and the payment of damages to data subjects.[106]

Transfers of SPDI (whether inside or outside India) are restricted. The transferor must ensure that the recipient maintains the same level of data protection as the transferor, and the transfer is allowed only if necessary, for the performance of a lawful contract or where the data subject has consented to the transfer.

These SPDI Rules apply to doctors and other healthcare providers, since they are collecting and processing medical data, a form of SPDI. In addition, where health information is to be used in research, extensive additional requirements apply. In 2017, the Indian Council of Medical Research updated its detailed Research Guidelines. They provide that informed consent is required for almost all research involving human participants. The Guidelines lay out a detailed description of the types of research

---

103. A "body corporate" is defined under section 43A of the IT Act as "any company and includes a firm, sole proprietorship or other association of individuals engaged in commercial or professional activities."
104. **The SPDI Rules, op. cit.**
105. **Ibid**
106. **Ibid**

consent processes and their implications, including:

- Blanket or broad consent (which is to be used only for biological samples where the donor wants it to be used for any future research);
- Tiered consent;
- Specific consent;
- Delayed consent;
- Dynamic consent;
- Withdrawal of consent or destruction of biological sample;
- Waiver of consent;
- Re-consent (for secondary or extended uses of stored samples or datasets).[107]

There are, however, circumstances in which the standard requirement for informed consent for research can be waived. Only the Ethics Committee for the research project can grant a waiver. A researcher can apply to the Ethics Committee for a consent waiver if the research involves less than minimal risk to participants and the waiver will not adversely affect the rights and welfare of the participants. An Ethics Committee can approve a consent waiver only in the following situations:

- The research cannot practically be carried out without the waiver and the waiver is scientifically justified;
- Retrospective studies, where the participants are de-identified or cannot be contacted;
- Research on anonymized biological samples/data;
- Certain types of public health studies, monitoring programs, or program evaluation studies;
- Research on data available in the public domain; or
- Research during humanitarian emergencies and disasters, when the participant cannot give consent.[108]

The Research Guidelines provide additional detail about when an Ethics Committee may waive consent for public health research.

The Guidelines note that increasing digitization of health records, along with IT advances and decreasing costs, offer huge potential for secondary uses of health

---

107. **National Ethical Guidelines for Biomedical and Health Research Involving Human Participants,** Indian Council of Medical Research, 2017, ("Research Guidelines"), available here
108. **Ibid**

data for research and commercialization. Nonetheless, the Guidelines warn that secondary research or commercialization uses must follow the same requirements as any other health research, following the same rules for informed consent, due diligence, and Ethics Committee review.[109]

In 2017, the Supreme Court of India delivered a landmark decision dramatically affecting Indian privacy law. In Puttaswamy,110 the Court unanimously held that individuals have an intrinsic Constitutional right to privacy, which includes a negative obligation to not violate their right to privacy. As noted by leading Indian legal scholars, Puttaswamy changed the contours of Indian privacy law, the interpretation of the existing privacy rules, and raised the spectre of a common law tort of violation of privacy."[111] The Court laid out a new Constitutional standard for scrutinizing any law that encroached on privacy. Moreover, the Court imposed a positive obligation on the government to enact legislation that protects the right to privacy.[112]

Pending Legislation

The Personal Data Protection Bill

While comprehensive privacy legislation had long been under consideration in Indian, the Puttaswamy decision accelerated the impetus to enact an updated and broader version of the IT Act. A draft law titled The Personal Data Protection Bill, 2018 (PDPB) was introduced in 2019 and is under review by Parliament, although recent reports indicate that the pandemic has delayed its progress.[113]

109. **Ibid**
110. **Justice K. S. Puttaswamy (Retd.) and Anr. vs Union of India And Ors,** Aug. 24, 2017, available here
111. **India – Data Protection Overview,** Mathew Chacko, Aadya Misra, and Purushotham Kittane, OneTrust DataGuidance, Feb. 2020, ("DP Overview") available here
112. **Ibid**
113. **Coronavirus delays the passage of the world's most important new privacy law,** Glyn Moody, Privacy News Online, Mar. 26, 2020, available here

Because numerous summaries of PDPB exist,[114] this chapter mentions only a few salient features:

- PDPB's scope is very broad, considerably broader than that of the IT Act. It would apply to almost all businesses in India, even small businesses;
- PDPB would define "personal data" more broadly than the IT Act does;
- PDPB would focus on consent as the primary legal basis for data processing, with fairly narrow exceptions for necessary government functions, compliance with court orders, medical emergencies, disasters, breakdowns of law and order, and reasonable purposes like whistleblowing, corporate transactions, credit scoring, debt recovery, etc;
- Similar to the GDPR, PDPB would impose requirements related to notice, consent, withdrawal of consent, data collection, data erasure, data portability, and Privacy-by-Design;
- Similar to the GDPR, PDPB would authorize penalties that could reach 4% of the worldwide annual turnover (**i.e.**, gross revenue) of a company (and its affiliates);
- Unlike the GPDR, PDPB would grant broad power to the central government to exempt any government agency from its requirements (analogous government exemption authority is more limited in the EU);
- Unlike the GDPR, PDPB would authorize the government to require businesses to share their nonpersonal data with the government to assist with government service delivery or formation of evidence-based policies;
- PDPB's data localization requirements - obligations to process and retain data inside India -are stricter than those emanating from the GDPR.[115]

For the processing of sensitive personal data, such as health information potentially to be used for secondary research purposes, PDPB would require explicit consent. To be valid, such explicit consent must be free, informed, clear, specific, and capable of being withdrawn. PDPB does not provide guidance on how such explicit consent varies from regular consent.[116] It is unclear how, or if, these heightened consent requirements would be reconciled with the authority that Ethics Committees have today to waive consent in accordance with the Research Guidelines. Further analysis is needed to

---

114. **See, e.g., What is in India's Sweeping Personal Data Protection Bill,** Anirudh Burman and Suyash Rai, Carnegie India, Mar. 9, 2020, available here; **Key Global Takeaways From India's Revised Personal Data Protection Bill,** Arindrajit Basu, Justin Sherman, LawFare, available here
115. **The Personal Data Protection Bill,** 2019, Bill No. 373 of 2019 ("the PDP Bill"), available here
116. **Ibid**

determine if, under PDPB, Ethics Committees could grant waivers of informed consent for low-risk, high-value research under the same conditions as they can today.

The Digital Information Security in Healthcare Act

Alongside the development of the comprehensive PDPB, India has been focusing on health-related privacy as part of its ambitious plans to accelerate a digital health technology ecosystem. This initiative included a 2012 mandate that all "clinical establishments" (i.e., healthcare providers) use Electronic Health Records (EHRs). In 2016, the government released its EHR Standards that laid out non-binding standards for hospitals to follow regarding system architecture, encryption, data elements to be captured, and internationally standard data formatting requirements.

Looking ahead, the National Health Policy, 2017, proposed striking advances in digital health, including an integrated health information system, the use of Aadhaar (a unique 12-digit identification number issued by the government to each resident), big data/public health registries, Health Information Exchanges, and improved national and personal technology to capture real-world data.[117] To enhance health privacy, the Ministry of Health & Family Welfare proposed draft legislation entitled the Digital Information Security in Healthcare Act, 2018 (DISHA), which would create a comprehensive health privacy regulatory scheme.

DISHA takes an approach that is even stricter and more rigorously consent-based than that of PDPB. In fact, DISHA requires both patient consent and a permitted legal purpose for processing specified by DISHA. Any other use of health data, even with consent, would be illegal.[118]

Under DISHA, individuals would have the right to give or refuse consent for every stage of processing including data generation, collection, storage, transmission, access, and disclosure. They would also have the right to withdraw consent for storage or transmission of data about them. Consent would be required repeatedly and at every stage where identifiable data is used. DISHA would also allow consent to be withdrawn for every stage of processing, including generation, collection, storage,

117. **DISHA and the draft Personal Data Protection Bill, 2018: Looking at the future of governance of health data in India,** Ikigai Law, Feb. 25, 2019, available here

118. **Digital Information Security in Healthcare Act (DISHA) [draft] ("DISHA"),** Ministry of Health & Family Welfare, March 18, 2018, available here

transmission and use. Individuals would have the right to withdraw consent and block processing of their data even in the limited cases where consent was not the original legal basis of the processing.[119]

DISHA contains terms that highly restrict processing by any entities other than clinical establishments and Health Information Exchanges. Any other entities could generate, collect, and store health data only:

- To advance delivery of patient-centered medical care;
- To provide information to guide medical decisions; or
- To improve coordination of care among hospitals, labs, etc.

Health data could not be generated, collected, stored, accessed, or disclosed for any purpose not listed by DISHA. And even purposes listed by DISHA would require consent or a legal mandate.[120]

Does secondary use of health data for research require review boards' approval?

All types of biomedical and health research (including clinical, basic science, policy, implementation, epidemiological, behavioral, public health research, etc.) must be reviewed by an Ethics Committee (EC). Before a research project can begin, an EC must review and approve the research proposal to safeguard the dignity, rights, safety, well-being, privacy and confidentiality of research participants. After granting initial approval, ECs have an ongoing responsibility to regularly monitor the research to ensure continued ethical compliance.[121]

Who can access health data for research?

Existing Law

Research data can only be accessed by entities authorized by the informed consent and the research project approved by the EC. Any additional recipients would likely require reconsent of the research participant or, at a minimum, EC approval.

119. **Ibid**
120. **Ibid**
121. **Research Guidelines, op. cit.,** p. 25.

If personal data is transferred to a service provider pursuant to contract, the service provider must still comply with the SPDI Rules, including the scope of the data subject's consent.

Pending Legislation

PDPB would create a possible legal basis for personal data to be processed for secondary use research. The DPA could promulgate regulations to exempt classes of research from any of the PDPBs requirements if the DPA is satisfied that:

a. compliance with PDPB would disproportionately divert resources from the research purpose;
b. the purposes of processing could not be achieved with anonymized data;
c. the data fiduciary has properly de-identified the data and the purpose of processing could be achieved with de-identified data;
d. the personal data would not be used to make any decision specific to or action directed to the data subject; and
e. the personal data would not be processed in a way that creates a risk of significant harm to the data subjects.[122]

Those accessing research data would still have to be entities within the scope of both the informed consent and the EC's approval.

Under DISHA, the National Electronic Health Authority could use health information for certain limited purposes such as public health research, subject to privacy protections. In addition, DISHA would allow the Chief Health Information Executive of a Health Information Exchange to access digital health data, which has raised security concerns.[123]

Can health data be transferred outside the country?

Existing Law

A body corporate may transfer SPDI outside India only if the recipient maintains the same level of data protection as required by the SPDI Rules. In general, consent of the data subject is required for all transfers of SPDI, whether to recipients inside or

122. **The PDP Bill, op. cit.**
123. **India: DISHA: The First Step Towards Securing Patient Health Data in India,** Dr. Milind Antani, Darren Punnen, and Anay Shukla, Mondaq, Aug. 3, 2018, available here

outside India. Consent, however, is not necessary for the transfer if required for the performance of a lawful contract between the corporate entity and the data subject, or as otherwise specified by the IT Act.

Pending Legislation

PDPB would not restrict the transfer outside of India of personal data in general. However, it would place restrictions on sensitive personal data, including health data. Sensitive personal data could be transferred outside of India only if:

A. A copy of the sensitive data were also stored in India, and
B. Either:
    1. The transfer was made pursuant to a contract approved by the Data Protection Authority (DPA), whose office would also be created by PDPB. Such approval would be granted only where effective provisions were made for data subjects' rights and for the data fiduciary to be liable for harm caused by noncompliance; or
    2. The Central Government, after consulting with the DPA, allowed the transfer to a country (or entities within a country) that the government found would afford an "adequate" level of protection for sensitive data in the hands of the transferee, and where the transfer would not prejudicially affect relevant law enforcement.[124]

In addition, the PDPB would introduce a new category of data called "critical personal data," which could not be transferred outside India except where necessary:

- To respond to a medical emergency or severe threat;
- For health services during an epidemic or other public health threat;
- To provide services during a disaster or breakdown of public order; or
- To a country, entity, or international organization, as permitted by the Central Government.[125]

124. **The PDP Bill, op. cit.**
125. **Ibid**

It is possible that health data could be classified as critical personal data after PDPB is enacted. A government data protection privacy committee has recommended that health data, genetic data, biometric data, and unique identification numbers be processed only in India.[126]

PDPB would require those who transfer critical personal data outside India to notify the DPA.

Under DISHA, a clinical establishment or entity could transfer digital health data only with the consent of the data subject, who has been informed of the purposes of collection and their rights under DISHA.[127]

What is the standard for anonymizing or de-identifying health data?

Existing Law

The SPDI Rules define "personal data" as "any information that relates to a natural person, which, either directly or indirectly, in combination with other information available or likely to be available with a body corporate, is capable of identifying such person."[128] Anonymized data that falls outside this definition is not subject to the Act or the SPDI Rules.

With regard to research, however, the definitions regarding identifiability become more complex. The following table from the Research Guidelines lays out the range of identifiability of research datasets and biological samples.[129]

126. **India-Health and Pharma Overview,** Darren Punnen, Dr. Millind Antani, Shreya Shenolikar, OneTrust DataGuidance, available here
127. **DISHA, op. cit.**
128. **The SPDI Rules, op. cit.**
129. **National Guidelines, op. cit.,** p. 129

Anonymous or unidentified	No identifiers are present from the start or if collected, are not maintained. Such samples are received by biobanks without any identifiers and supplied to researchers.	
Anonymized	This involves systematic de-identification, reversible or irreversible: link of samples/data to personal identity is reversibly or irreversibly cut.	
	Coded or reversibly anonymized: There is an indirect link of sample / data to the participant's identity with restricted access. This link could be re-linked if required; therefore, it may also be termed reversible anonymization.	Irreversibly anonymized: Link to the participant's identity is removed and cannot be re-linked.
Identifiable	A direct link of sample / data to the participant's identity exists.	

The Research Guidelines also make it clear that, in general, research must be conducted on the least identifiable data suitable for the project goals. The greater the identifiability, the greater will be the requirements for informed consent, confidentiality, and re-consent. Interestingly, the Guidelines point out that sometimes a degree of identifiability should, or even must, be retained. For example, where data subjects or sample donors are to be permitted to later withdraw consent, their data or samples must be kept in coded form to permit re-identification in order to fulfill their future request to have the data or samples destroyed.[130]

Pending Legislation

The PDPB would draw a distinction between anonymized and de-identified data. Anonymization would be defined as "such irreversible process of transforming or converting personal data to a form in which a data principal cannot be identified,

130. **Ibid.,** p. 128-29

which meets the standards of irreversibility specified by the [Data Protection] Authority." Anonymized data is data that has undergone the process of anonymization.[131] In contrast, de-identified data would mean "the process by which a data fiduciary or data processor may remove, or mask identifiers from personal data, or replace them with such other fictitious name or code that that is unique to an individual but does not, on its own, directly identify the data principal." Except as noted below with regard to the government access right, PDPB would not apply to anonymized data, whereas it would apply to de-identified data.[132]

PDPB contains an unusual provision that would give the Central Government, in consultation with the DPA, the right to require any data fiduciary or data process to provide anonymized data to the government. PDPB states that the purpose of the government's access right to anonymized data is to enable better targeting of delivery of services or formulation of evidence-based policies by the Central Government.[133] Although large-scale anonymized datasets are often valuable trade secrets held closely by the companies or organizations that created them, PDPB does not contain any provision for compensation to the companies or organizations that developed the anonymized datasets that might be turned over to the government.

DISHA would also distinguish between anonymized and de-identified data. Anonymization would mean "the process of permanently deleting all personally identifiable information from an individual's digital health data." De-identification would mean "the process of removing, obscuring, redacting or delinking all personally identifiable information from an individual's digital health data in a manner that eliminates the risk of unintended disclosure of the identity of the owner and such that, if necessary, the data may be linked to the owner again."[134]

Unlike most privacy laws, DISHA does contain certain restrictions on anonymized data. DISHA contains an absolute prohibition on access to digital health data, even in anonymized form, by pharmaceutical companies, insurance companies, employers,

131. **The PDP Bill, op. cit.**
132. **Ibid**
133. **Ibid., sec. 91**
134. **DISHA, op. cit.**

human resource consultants, or other entities specified by the government.[135] Concerns are being raised about the barriers this would impose to medical research and government approval of medications and devices.

Other noteworthy aspects of secondary use of health data for research

In response to the current COVID-19 pandemic, the Indian Council of Medical Research has published an updated version of its National Guidelines for Ethics Committees Reviewing Biomedical and Health Research. These Guidelines stress the importance of maintaining ethical and confidentiality protections for data subjects even amidst the pandemic, while also taking steps to handle the challenges of obtaining informed consent from participants who are ill or physically distant, conducting EC reviews remotely, ensuring biosafety in labs and hospitals, and using telemedicine for research whenever possible. The Guidelines also update the standards for waiver of informed consent, including the possibility of delaying the informed consent process if consent is not possible due to the emergency, if approved by the EC after careful consideration.[136]

135. **Ibid.** Section 29(5) states that "[d]igital health data, whether identifiable or anonymized, shall not be accessed, used or disclosed to any person for a commercial purpose and in no circumstances be accessed, used or disclosed to insurance companies, employers, human resource consultants and pharmaceutical companies, or any other entity as may be specified by the Central Government.

136. **National Guidelines for Ethics Committees Reviewing Biomedical and Health Research During COVID-19 Pandemic,** Indian Council of Medical Research, April 2020, available here

Ireland

What is the legal basis for secondary use of health data for research?

Legal Context

The Data Protection Act 2018 (DPA2018) was enacted to set out limited exemptions to some provisions of the GDPR, and to prescribe safeguards for certain processing activities. The Data Protection Act 2018 (Section 36(2) (Health Research) Regulations 2019 (S.I. 314/2018),[137] as amended by the Data Protection Act 2018 (Section 36(2) (Health Research) (Amendment) Regulations 2019 (S.I. 188/2019)[138] (the Health Regulations), were subsequently introduced under section 36 of DPA2018 by the Minister for Health, to make the implementation of a number of "suitable and specific measures" mandatory in respect of health data research.

The Health Regulations broadly define "health research", such that any of the following types of scientific research concerning human health will fall within the scope of the Health Regulations where they involve the processing of personal data:[139]

- Research with the goal of understanding the normal and abnormal functioning of the human body, at molecular, cellular, organ system and whole body levels;
- Research specifically concerned with developing innovative strategies, products or services to diagnose, treat or prevent disease or injury;
- Research with the goal of improving the diagnosis, treatment (including rehabilitation and palliation) of human disease and injury and of improving the health and quality of life of individuals;
- Research with the goal of improving the efficiency and effectiveness of health professionals and the health care system; and
- Research with the goal of improving the health of the population as a whole or any part of the population through better understanding of social, cultural, environmental, occupational and economic factors on determining health status.

137. **Data Protection Act 2018 (Section 36(2)) (Health Research) Regulations 2018**, available here
138. **Data Protection Act 2018 (Section 36(2)) (Health Research) (Amendment) Regulations 2019**, available here
139. Regulation 3(2), Health Regulations

Additionally, the action of establishing whether a person is suitable for inclusion in research can constitute "health research" in and of itself.

As such any data controller that uses clinical data for research must comply with the Health Regulations, whether that data is sourced directly from the data subject for specific research, or whether it is sourced from another data controller who collected it in an unrelated context (e.g. in the course of a doctor-patient relationship or in respect of a different research project).

The need to comply with the Health Regulations will not apply only where the data is entirely anonymized, or where "health research" is not being conducted. In this regard, it is worth noting that the Health Research Board (HRB) have recognized that service evaluations, clinical audits, and "usual practice" (e.g. investigations of the health of a population and the causes of disease), do not constitute "research".[140]

Explicit Consent

In addition to mandating that researchers comply with a range of safeguards, the Health Regulations have essentially restricted the ability of data controllers to rely on any lawful basis for health research other than explicit consent.

The details of the "suitable and specific measures" that must be adopted to safeguard the rights and freedoms of data subjects are as follows:

- Arrangements so that personal data will be processed as is necessary to achieve the objective of the health research, and not be processed in such a way that damage or distress is, or is likely to be, caused to the data subject;
- Appropriate governance structures for carrying out the health research, including;
  - ethical approval for the health research by a research ethics committee;
  - specification of the data controller involved;
  - compliance with Article 26 of the GDPR in the case of joint controllers;
  - specification of any data processors involved;
  - specification of any person who provides funding for, or otherwise supports, the project;
  - specification of any person other than a joint controller or data processor with whom it is intended to share any personal data (including where it has

140. **Q&A,** Health Research Board (HRB), available here

been pseudonymized or anonymized) and the purpose of such sharing;
  - provision of training in data protection law and practice to individuals who carry out the health research;
- Processes and procedures relating to the management and conduct of the health research, specifically:
  - an assessment of the data protection implications of the health research;
  - a data protection impact assessment where an initial assessment indicates a high risk to the rights and freedoms of individuals;
  - measures that demonstrate compliance with the data minimization principle in Article 5(1)(c) of the GDPR;
  - controls to limit access to the personal data undergoing processing in order to prevent unauthorized consultation, alteration, disclosure or erasure of personal data;
  - controls to log whether and by whom personal data have been consulted, altered, disclosed or erased;
  - measures to protect the security of the personal data;
  - arrangements to anonymize, archive or destroy personal data once the health research has been completed; and
  - other technical and organizational measures designed to ensure that the processing is carried out in accordance with the GDPR, together with processes to test and evaluate the effectiveness of such measures;
- Arrangements to ensure that personal data are processed in a transparent manner; and
- Processing personal data for the purpose of specified health research only where "explicit consent" has been obtained from the data subject in accordance with Article 4 of the GDPR before the research commences.

Although the legislation is clear that the relevant standard of consent is that required by the GDPR, it is worth noting that the Department of Health has published very detailed guidance on "information principles for informed consent for the processing of personal data for health research".[141]

141. **Guidance on Information Principles for informed consent for the processing of personal data for health** research, Department of Health, October 2018, available here

Does secondary use of health data for research require review boards' approval?

The Alternative to Explicit Consent – Consent Declaration

A data controller may apply to the Health Research Consent Declaration Committee (HRCDC) for a declaration that the explicit consent of data subjects is not required (a Declaration), where the public interest in conducting the health research "significantly outweighs" the public interest in obtaining their explicit consent.

Applicants for a Declaration must meet the conditions of Regulation 5, many of which involve demonstrating compliance with the "suitable and specific" measures in Regulation 3. Regulation 5 requires that applicants must:

- conduct a data protection impact assessment in accordance with Article 35(1) of the GDPR;
- obtain ethical approval of the health research from a research ethics committee;
- make an application in writing to the HRCDC and provide written information that clearly identifies:
  - that the controller has a valid and lawful basis for the processing of the personal data;
  - that the controller meets one of the conditions in Article 9(2) of the GDPR;
  - the controller and, where there are joint controllers, the division of responsibilities between them within the meaning of Article 26 of the GDPR;
  - that the health research requires that personal data be obtained and processed rather than anonymized data;
  - that the personal data will not be processed in such a way that damage or distress is, or is likely to be, caused to the data subject;
  - that the collection and use of the personal data will go no further than is necessary for the attainment of the research objective;
  - that there will be no disclosure of the personal data unless that disclosure is required by law or the data subject has given his or her explicit consent to the disclosure;
  - that specified measures in Regulation 3 have been identified and will be put in place before the health research commences;
  - a data protection officer has been appointed in relation to the health research; and
  - ethical approval from a research ethics committee has been received.

- provide the HRCDC with:
  - a copy of the result of the data protection impact assessment that has been carried out, with a particular reference to the possibility of data linkages and details of any consultations undertaken with potential data subjects; and
  - written information demonstrating that the public interest in conducting the health research significantly outweighs the public interest in requiring the explicit consent of the data subject, together with reasons why it is not proposed to seek the consent of the data subject.

The HRCDC has published its own application form, which speaks to these requirements and also poses a range of other questions.[142]

The HRCDC must consider each application as "soon as practicable," and it may consult as appropriate and request further information from the applicant. Before granting a Declaration, the HRCDC must be satisfied that the requirements as set out in the Health Regulations have been met, and that the public interest in carrying out the research significantly outweighs the public interest in requiring the explicit consent of the data subject. The HRCDC can also attach such conditions to a Declaration as it "considers necessary to protect the interests of a data subject likely to be affected by the declaration."

Where the HRCDC refuses to grant a Declaration, attaches conditions to a Declaration, or revokes a Declaration for non-compliance with any conditions, the applicant can appeal the decision to the Minister for Health, provided that they give notice of their intention to appeal within 30 working days of receipt of the HRCDC decision.[143] In such circumstances, the Minister must establish an independent appeal panel within 40 working days, so that the appeal can be heard de novo.

Regulation 11(3)(c) provides that an appeal panel can determine its own procedure, but the Department of Health has issued some guidance on the appeals process which largely repeats the terms of the Health Regulations.[144]

142. HRCDC application form, available here
143. Regulation 11
144. **Health Research Regulations – Appeals Process from Decisions of the Health Research Consent Declaration Committee**, Department of Health, available here

What is the standard for anonymizing or de-identifying health data?

Recital 26 of the GDPR describes "anonymous information" as "information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."

In Ireland if a data controller processes or otherwise shares fully anonymized clinical data with another data controller for the purposes of health research, the need to comply with the GDPR and the Health Regulations does not arise.

Anonymization in and of itself constitutes a processing activity in respect of which a legal basis must be identified. As the HRB have explained, where consent is the legal ground for personal data collection, then consent is required for further anonymization of that data. However, if the data controller has another legal basis (other than consent) and, where relevant, meets at least one of the Article 9(2) GDPR conditions (other than explicit consent), then consent is not required.[145]

In considering whether a given data set has been sufficiently anonymized, it would be necessary to have regard to the Irish Data Protection Commission's guidance on anonymization and pseudonymization, which notes, "data can be considered "anonymised" from a data protection perspective when data subjects are not identified or identifiable, having regard to all methods reasonably likely to be used by the data controller or any other person to identify the data subject, directly or indirectly", including potential identifiability by "singling out, linkability and inference".[146] As the Data Protection Commission's guidance is loosely based on that of the Article 29 Data Protection Working Party's Opinion 05/2014 on Anonymization Techniques,[147] its guidance should be aligned with that of other European regulators.

145. **Health Research Regulations 2018 FAQ**, Health Research Board, available here
146. **Guidance on Anonymisation and Pseudonymisation,** Data Protection Commission, June 2019, available here
147. **Opinion 05/2014 on Anonymisation Techniques**, Article 29 Data Protection Working Party, Adopted on 10 April 2014, available here

Who can access data?

Data controllers are required to specify "with whom it is intended to share any personal data (including whether it has been pseudonymised or anonymised) and the purpose of such sharing)" at the outset, as this goes to the data subject's consent being sufficiently informed. It is also a requirement of the Health Regulations that the relevant data controller implement controls to limit access to the personal data undergoing processing in order to prevent the unauthorized consultation, alteration, disclosure or erasure of personal data. As such, data sharing by the data controller should be on a "need-to-know" basis, and all data should be pseudonymized or anonymized to the greatest extent possible.

Where a Declaration is required, the data controller cannot disclose the personal data "unless that disclosure is required by law or the data subject has given their explicit consent to the disclosure".

If a data controller seeks to share personal data with another data controller for the purposes of health research, such data cannot be shared unless the explicit consent of the data subject was obtained in the first instance. Similarly, if a data controller collects personal data for a given research project on the basis of data subjects' consent, it cannot subsequently use their data for a different research purpose or share it with another data controller unless the data subjects gave their prior, freely given, explicit, specific and informed consent to same.

Can data be transferred abroad?

Personal data can be transferred abroad for the purposes of health research subject to compliance with the relevant provisions of the GDPR.

Other noteworthy aspects of secondary use of health data for research

- Although the Health Regulations are silent on their territorial scope, the HRCDC has indicated that "the collection of personal data for health research is bound by the legislation of the country in which it was collected," and that the Health Regulations "govern all processing of personal data for health research purposes conducted within the Republic of Ireland.";[148]

148. HRCDC, **Guidance** available here

- To the extent that public bodies may need to share personal data with each other in a healthcare context, it is notable that the Data Sharing and Governance Act 2019 (the 2019 Act) aims to regulate such sharing and expand the lawful bases upon which certain public bodies (including the Health Service Executive) may conduct such sharing. However, a number of key provisions have not yet been commenced. These include section 13 (for purposes that do not relate to the administration of a public service pension scheme), which sets out the circumstances in which a public body can disclose personal data to another public body. Further, the parts of the 2019 Act that apply to data sharing will not apply to special categories of personal data, meaning that its applicability to data sharing in a healthcare context may be limited;[149]

- For a short period of time, a temporary National Research Ethics Committee for COVID-19 was established with the aim of granting all necessary approvals within 7 days of confirmation of a validated application. Applicants were required to answer mandatory questions on compliance with the Regulations, and if they needed a consent declaration, they were required to complete an additional section of the form for forwarding to the HRCDC to enable contemporaneous evaluation.[150]

149. Section 5, 2019 Act
150. **NREC Covid-19 overview,** available here

Israel

Introduction

Israel has a universal public health system managed by the Ministry of Health (MoH). Israeli citizens have a unique identification number, allowing for effective connectivity between databases (i.e., cross-referencing and linkage of data about individuals from different databases). Israel's diversity further contributes to the exceptional research quality of its healthcare databases.[151] These unique features of Israel's health system and population have a significant bearing on the local data environment and make Israeli-based research and innovation particularly attractive.

The Israeli public not only funds the health care system through taxation but is also the source of information originally collected during and for primary (health care) purposes; then being repurposed for secondary (mainly research) uses. This means that it has a vested interest in the health information silos held by health organizations (including health maintenance organizations (HMOs), hospitals and medical centers). Health organizations' collection and secondary use of patients' health information is deemed consistent with the public interest in promoting healthcare, inter alia, through research (insofar as the public's interest in protecting its privacy and autonomy with respect to the secondary uses of health information, is upheld).

Israel's existing and draft reform regulatory regimes concerning the secondary use of big health data for research purposes are described below.

151. Within a rather modest population of about 9 million Israeli inhabitants, a rich variety of origins is represented: Eastern and western Europe, Asia, Arab states, North America, and North Africa. Consequently, Israel's populace is a mix of Ashkenazi, Sephardic, and Ethiopian Jews; Arabs; Druze; Bedouins; and Circassians, with generally highly inbred populations. See, S. Tamir, **"The Precision Medicine Data Environment in Israel: A Review"**, The Van Leer Jerusalem Institute (2020), available here

What is the legal basis for secondary use of health data for research?

The current regulatory landscape

In Israel, clinical trials in human subjects are regulated by the Public Health Regulations (Clinical Trials in Human Subjects) – 1980[152] (Clinical Trials Regulations). These regulations define medical experimentation on humans and set forth the terms and stages for review and approval of clinical trials in human subjects by institutional review boards (IRBs) and by the Supreme (Helsinki) Committee for Clinical Trials (Supreme Helsinki Committee).

According to the 2006 Director-General (MoH) Circular No. 15/06 – Helsinki Subcommittee for Approval of Research That is Not a Medical Experiment in Humans[153](IRB Subcommittee Circular), research conducted on data collected from medical, nursing, psychological and other records, that are **s**trictly a secondary data analysis without patient involvement or interaction, do not constitute clinical trials in humans.

The circular creates an expedited route for the approval by an IRB's subcommittee of research restricted to data analysis that is deemed to be of minimal risk. This represents the currently applicable legal and policy instrument for research using big health data.

Israel's Ministry of Health has recently proposed comprehensive reform of this regime by draft Patient Rights Regulations, which are set to establish a designated expert ethics committee mechanism for reviewing health data research.

In Israel, the Right to Privacy has been acknowledged as a basic human right in article 7 of the Basic Law: Human Dignity and Liberty[154] of 1992 (Basic Law).

Personal data is more specifically protected under the Protection of Privacy Law, 5741 – 1981[155] (PPL), with individual health data being considered "sensitive information" under said Law, thereby meriting a higher level of privacy protection. Chapter two

152. **Public Health Regulations (Clinical Trials in Human Subjects) – 1980**, available here [Heb.]
153. **Helsinki Subcommittee for Approval of Research That is Not a Medical Experiment in Humans**, Director-General (MoH) Circular No. 15/06, available here [Heb.]
154. **Basic Law: Human Dignity and Liberty of 1992**, available here [Heb.], [Official translation here]
155. **Protection of Privacy Law, 5741 – 1981**, available here [Heb.] [Unofficial English translation here]

of the Protection of Privacy Law sets forth the provisions for the protection of privacy in databases, along with an establishment of an enforcement authority.

In accordance with the data protection environment promoted by the 2018 EU General Data Protection Regulation (GDPR), Israel added a further layer of privacy protection by introducing the Privacy Protection (Data Security) Regulations – 2017[156] (Privacy Protection Regulations). These regulations specify comprehensive data security obligations for databases and apply in a sweeping and binding manner to any activity of processing personal information that is subject to Israeli law, in both the public and private sectors.

The regulations stipulate categories of security levels for databases, depending on the volume and nature of the information they hold. Databases containing medical information, information regarding a person's mental condition, or genetic information are categorized as "subject to a medium security level" (see First Schedule). Databases containing the same type of information regarding 100,000 persons or more, or databases for which the number of persons authorized to access this information exceeds 100, including a database of a public body, are categorized as "subject to a high level of security" (see Second Schedule). Stricter controls and information security measures must be applied to databases in these two categories, and appropriate obligations are imposed upon database controllers to prevent unauthorized use of data held therein, which is considered a "severe security incident".

These regulations came into effect in May 2018. In 2011, Israel was awarded a declaration of adequacy by the European Commission according to the EU Directive that preceded the GDPR.[157] Consequently, transfer of EU persons' data to Israel is expressly permitted. Naturally, this has important bearings on the ability of Israeli researchers to take part in the international exchange of health data. While Israel has not yet fully adapted its privacy legislation to approximate the GDPR, it has expressed its intent of doing so.

---

156. **Privacy Protection (Data Security) Regulations – 2017**, available here [Heb.], [Unofficial English translation here]

157. 2011/61/EU: **Commission Decision of 31 January 2011 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data by the State of Israel with regard to automated processing of personal data.**

The legal framework for sharing of health data – either by granting access thereto, or by permitting delivery thereof – under Israeli law, is based on the following pieces of legislation:

1. The Patient Rights Law – 1996[158] (Patient Rights Law). According to s. 20(a)(7) of the law, a clinician or medical institution may transmit or release medical or health information to another, inter alia for research purposes and for publication in a scientific journal, provided that patient identifying information shall not be disclosed. The delivery of health information shall be subject to data minimization and purpose limitation requirements, as well as an obligation to take utmost care in assuring that the patient (i.e., data subject) shall remain unidentifiable.
2. The Genetic Information Act – 2000[159] (Genetic Information Act). According to s. 23 therein, a person/organization holding genetic information or a genetic database may transmit the information in its possession for purposes of legally approved research, or publication in a scientific journal, on condition that: (1) the genetic information is transmitted without any identifying detail; or (2) the individual data subject has consented in writing to the delivery of genetic information.

Under its national healthcare scheme, Israel provides healthcare insurance and receives health services through one of four health-care providers (HMOs). HMOs in Israel have been collecting health information electronically for more than 20 years. Despite Israel's small size, these HMOs' health information silos contain broad information on a large number of patients, even on an international scale. Two of the HMOs, Clalit and Maccabi (which insure approximately 50% and 25% of the Israeli population, respectively), invest substantial resources in developing the collection and use of the information in their possession and operate their own research institutes. Consequently, each of the HMOs, as health data controllers, perceives its health information silo as an exclusive asset, in which it has invested considerable resources. HMOs seek to use the health information in their possession to promote research and innovation within the organization, as well as to improve their efficiency and the quality of the medical treatment they provide.

Hospitals/Medical Centers in Israel, typically owned by the HMOs or outright by the government, also gather information accumulated in the course of medical

158. **The Patient Rights Law – 1996**, available here [Heb.]
159. **The Genetic Information Act – 2000**, available here [Heb.], [Unofficial English translation here]

treatment and care and during hospitalizations. Some hospitals make extensive use of the information in their possession, promoting innovative research, while others do so on a smaller scale.

The requirement for informed consent for participation in medical experiments is embedded in the Clinical Trials Regulations, which incorporate the WMA Declaration of Helsinki ethical principles. The required components of informed consent are stipulated in the Patient Rights Law (s. 13). Further, more specific provisions of law relating to consent to participation in genetic research, can be found in the Genetic Information Act (s. 11(a)).

Consent is currently the only legal basis that allows for the processing of personal data under the Protection of Privacy Law.

According to s. 20 of the Patient Rights Law, mentioned above, the sharing of health information by a medical practitioner or institution for purposes of medical research without the explicit consent of the patient, is conditioned upon the information being anonymized (s. 20(a)).

According to the IRB Subcommittee Circular, in the case of health data sought for use in research and collected from existing health records, without patient involvement or interaction, the subcommittee may exempt the researcher from acquiring informed consent, provided that the information is "fully anonymized".

A new regulatory regime for using health data for research purposes

Recognizing the enormous potential inherent to big health data and the advantages of sharing health information gathered and held by health organizations with local and international academic and industry research bodies, Israel's MoH has appointed in 2016 a public committee to examine the implications of and provide guidance for secondary uses of big health data. In January 2018, the committee published its recommendations, which sought to balance the public's interest in the use and sharing of health information for research against the rights of individual data subjects to protection of their privacy and autonomy. The committee's recommendations were immediately followed by two MoH Director-General circulars: Director-General (MoH) Circular No. 01/2018 – Secondary Uses of Health Information; and Director-General (MoH) Circular No. 02/2018 – Collaborations Based on Secondary Uses of Health Information.

The committee's recommendations and the pair of circulars, which served as a basis for the reformulation of the regulation of secondary use of health data for research,

were considered and drafted, inter alia, in light of the OECD Recommendation of the Council on Health Data Governance, of January 2017[160] concerning secondary uses of information.

The new draft regulations*

At the end of 2019, a draft of the Patient Rights Regulations (Research Use of Health Information), 2019[161] (the draft regulations), was published for public consultation. The draft regulations, reflecting the Committee's recommendations, set up a designated review mechanism for health data research applications both at the organizational and the national levels.

The explicit aim of the regulations was to improve the quality of medical care and promote medical research, while protecting the privacy of data subjects and the confidentiality of medical information. Various data ethics principles are embedded in the proposed regulations (including: beneficence; autonomy; and privacy protection through purpose limitation, data minimization, limited access and the de-identification of health information).

The draft regulations define categories of data based on identifiability, including identifying information, identifiable information (which does not include identifying information but can potentially with reasonable effort identify a data subject), and de-identified (anonymized) health information (s.1).

The draft regulations permit use of de-identified health information for research purposes (s. 2(e)); also allowing use of identified or identifiable health information subject to prior opt-in consent of a data subject or under an existing statutory exemption (s. 2(f)).

The draft regulations tighten information security obligations vis-à-vis those required under the Privacy Protection Regulations, by mandating compliance with information security requirements applicable to the category of databases subject to a high level of security (s. 4).

160. OECD, **Recommendation of the Council on Health Data Governance**, OECD/LEGAL/0433; available here

161. * These regulations are still in draft form and subject to various revisions.
[Draft] **Patient Rights Regulations (Research Use of Health Information)**, 2019, available here [Heb.]

Unlike the traditional opt-in consent requirement for clinical trials, the draft regulations adopt an opt-out approach for secondary research uses of health information.

Data subjects' autonomy is respected through the availability of a universal opt out (Exit Request) (s. 3). Under the draft regulations, any person is entitled at any time to request that no personal health information about him or her is used for research purposes. An opt out request may be rescinded by the data subject at any time. Parents or legal representatives may opt out on behalf of minors or persons under guardianship. An opt out request shall be submitted through a national online mechanism managed by the Ministry of Health (appropriate guidelines are set to be published) and will apply to all health organizations.

Does secondary use of health data for research require review boards' approval?

The growing prevalence of non-interventional health data research, as well as the understanding that the use, storage, access and delivery of health information require a more nuanced proficiency in privacy and information security, have led to the creation of a more suitable expert-based ethical review mechanism to replace the one set in the IRB Subcommittee Circular.

The suggested composition of the designated institutional organizational and national review boards reflects this recognition. The Organizational Review Board is set to include a privacy protection or information security officer as well as an expert in the field of information analysis (s. 6(d)). The National Review Board will include, among others, a data analysis specialist and experts in the field of privacy protection.

The Organizational Review Board's mandate is to examine the ethics and privacy protection of research applications using health information held by the organization itself or by partner organizations.

The National Review Board mandate includes, inter alia: (a) review of a health organization's request to establish an internal database; (b) review of an information access application for a single research use, jointly submitted by several health organizations; (c) review of an application for the delivery of de-identified health information outside the control of an organization to a non-health organization; and (d) a residual authority to discuss questions of principle at the request of an organizational review board, the MoH Director General, and more.

Who can access health data for research?

Under current law, researchers seeking access to health data must submit a research application to an IRB of the medical institution/HMO where the data is kept, in accordance with the IRB Subcommittee Circular.

The draft regulations, however, distinguish between internal and external researchers. External researchers need to meet stricter requirements as a condition to be granted access to health information by an organizational review board (s. 16), including, inter alia: having sufficient familiarity with health information research; signing a data use agreement; and duly complying with it.

The default arrangement is to grant access to de-identified data to researchers in a secure research environment (e.g., an online research room), controlled by the health organization.

Delivery of health information to external researchers, as distinct from provision of access thereto, requires the approval of both the organizational and national review boards (s. 7(d)(3)), which will be granted solely under exceptional circumstances and subject to the following conditions: (1) The applicant is a resident of Israel or a corporation registered in Israel; (2) Technological or research limitations prevent the research use within the health organization premises and cannot be addressed with reasonable efforts; (3) The study stands to greatly benefit individuals' or public health or make a significant contribution to the advancement of medicine or medical research; (4) The external researcher possesses unique technological or research capabilities without which it is impossible to realize the research objectives; (5) The external investigator maintains information security and privacy measures that meet the requirements set in the regulations; and (6) There is no evidence that the researcher will misuse the de-identified data (s. 17).

Can health data be transferred outside the country?

Data transfers outside of Israel are governed by the Privacy Protection Regulations (Transfer of Data to Databases Outside of Israel), 2001 (the "Transfer Regulations"). The Transfer Regulations apply to both inter- and intra-entity transfers of personal data outside of Israel. They permit transfers to: (i) EU Member States; (ii) other signatories of Council of Europe Convention 108; and (iii) a country "which receives data from Member States of the European Community, under the same terms of acceptance". In the past, this has been interpreted by the Database Registrar to apply to transfers to Safe Harbor participant companies in the US. However, following the European

Court of Justice decisions in the Schrems cases, the Database Registrar revoked the authorization for transfers from Israel to the US under that clause.

Transfers to other countries are permitted: (i) subject to data subject consent; (ii) from an Israeli corporate parent to a foreign subsidiary; or (iii) provided the data importer "enters into a binding agreement with the data exporter to comply with Israeli legal standards concerning the storage and use of data".

Regardless of the basis for an international transfer, data exporters must also obtain the data importer's written undertaking that the data importer implements sufficient safeguards to protect individuals' privacy and promises to refrain from any onward transfer in its own country or any other country.

Under the draft regulations, the transfer of data overseas, which is deemed to entail a significant privacy risk, is subject to the approval of both the organizational and national review boards. The draft regulations also stipulate that de-identified information provided to an external investigator shall not be transferred outside the borders of Israel (s. 17(d)). However, as mentioned in the draft regulations' explanatory notes, this stricter stipulation merits further scrutiny.

What is the standard for anonymizing or de-identifying health data?

The current legal framework does not provide definitions nor practical guidelines for anonymization or de-identification.

The responsibility for de-identifying health information lies within the data controlling organization. According to the draft regulations, de-identification of requested health data shall be performed by a data controller prior to making the data accessible or its delivery to a researcher. The draft regulations require that the de-identification protocols will be determined by the designated (organizational/national) review board on a case-by-case basis, and in accordance with the privacy risk assessment conducted by the board.

Accompanying the draft regulations were the Ministry of Health's draft guidelines for de-identification of health information for research use, designed to guide health organizations in the implementation of appropriate de-identification procedures. The guidelines offer a practical toolbox, including anonymization principles (mandating the use of the lowest level of identification that still enables the achievement of the desired purposes), risk analysis parameters and key principles of an appropriate de-identification model.

The draft regulations are technology specific, requiring, for example, a k-anonymity factor of K=4 - K=8 for any transfers of structured data out of an organization's secure environment, except for data containing clinical information necessary for the research.

Other noteworthy aspects of secondary use of health data for research

On May 7th 2020 the Israeli Ministry of Health published a scheme for establishing a national repository infrastructure for research related to the Covid-19 pandemic (the Corona Research Repository). The Corona Research Repository will gather data collected in the Israeli healthcare system about the disease and allow using de-identified data for research under privacy and security policy measures.

The Repository will be based on various data silos in the health care system and other governmental agencies, holding data on individuals infected with Covid-19 (verified patients).

Furthermore, the Repository will provide data from anonymous questionnaires collected from the general population on symptoms experienced. The Corona Research Repository will also interconnect with the following public repositories holding population data by locality; by socio-demographic statistics and more.

Using Covid-19 health data for research is subject to ethical and legal approvals. Data for research shall be de-identified through processes that reduce the risk of re-identification of the data subjects. Research will be conducted in virtual "research rooms" that comply with information security and privacy protection standards. Authorized researchers will be required to sign a confidentiality agreement including a commitment not to attempt re-identification of data subjects. The data will be used under the supervision and control of the Ministry of Health.

A verified patient who does not want his or her data to be used for research purposes may choose to opt-out. The Ministry of Health will establish a dedicated website and call center for the opt-out process.

The Repository will be cloud-based and hosted in the Ministry of Health cloud, providing several analytics applications and different open source directories. The "research room" will not have access to the internet nor to GitHub.

Requests to export data outside the "research room" environment will undergo information security and de-identification processes by the Ministry of Health. The data exported must be aggregate and adhere to the principle of K anonymity = 15. The Ministry of Health will allow the export of code written during the research (algorithm).

United States

What is the legal basis for secondary use of health data for research?

The US national medical privacy law, the Health Insurance Portability and Accountability Act of 1996 (1996) (HIPAA), governs the uses and disclosures of Protected Health Information (PHI). HIPAA creates a sphere of healthcare activity inside which patient consent is not required, which is commonly referred to as Treatment, Payment, and Healthcare Operations (TPO). Research is not part of TPO. Therefore, a legal basis for the use or disclosure of PHI for research purposes must be established.

HIPAA allows several pathways for PHI to be used or disclosed for research:

1. Pursuant to a patient authorization, which is a detailed form of consent that must contain specified terms;
2. Pursuant to a waiver of consent by an Institutional Review Board or Privacy Board, applying the criteria described below;
3. For limited purposes, "preparatory to research" (e.g., preparing a research protocol or identifying potential research subjects), subject to strict controls, including that the PHI cannot be removed from the original site;
4. As part of a "Limited Data Set" from which 16 specified direct and indirect identifiers have been removed, which may be shared only for specified purposes and must be subject to a Data Use Agreement prohibiting re-identification and attempts to contact data subjects; or
5. Research involving decedents, subject to specific researcher representations.[162]

Does secondary use of health data for research require review boards' approval?

Yes. An Institutional Review Board (IRB) or Privacy Board can waive HIPAA's general requirement for individual authorization for research uses and disclosures of PHI. (Since Privacy Boards are rare, this section mostly refers to IRBs.)

162. **Research**, US Department of Health & Human Services, available here

The criteria for such IRB waivers are that:

1. The use or disclosure of PHI involves no more than minimal risk to the privacy of individuals;
2. The research could not practicably be conducted without the waiver or data alteration; and
3. The research could not practicably be conducted without access to and use of the PHI.[163]

Can health data be transferred outside the country?

Yes, provided the HIPAA Rules are otherwise followed and provided that any special risks involved in an ex-US location are appropriately assessed and managed. When conducting the risk analysis and risk management required by the Security Rule, the transferor should consider ex-US threats and vulnerabilities, especially regarding enforcement of privacy and security controls. In addition, a Business Associate Agreement that includes mandatory elements must be in place before a Business Associate can receive, create, transmit, or maintain PHI, whether inside or outside the US.[164]

What is the standard for anonymizing or de-identifying health data?

Data that is "de-identified" per HIPAA's standard is no longer Protected Health Information and is not subject to HIPAA. Where health information does not identify an individual, and there is no reasonable basis to believe that it can be used to identify an individual, it does not fall within the definition of PHI.[165] HIPAA allows two methods for de-identification:

- The Safe Harbor method, whereby 18 direct and indirect identifiers have been removed, and the organization does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual;[166]
- The Expert Determination method, whereby an expert with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods of de-identification determines that the risk is very small that the

163. **Ibid**
164. **Do the HIPAA Rules allow a covered entity or business associate to use a CSP that stores ePHI on servers outside of the United States?** US Department of Health & Human Services, available here
165. 45 CFR 160.103
166. 45 CFR 164.514(a)(2)

information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify a data subject.[167]

In contrast to de-identified data, data in a Limited Data Set (see definition above) is still Protected Health Information and is still subject to HIPAA. While it may be used and disclosed for research without consent, it must be used and disclosed only pursuant to a Data Use Agreement containing specific re-identification and other prohibitions.

Of the two HIPAA de-identification methods, the Safe Harbor method is easier to implement, although researchers often criticize the method because of how it can limit data quality and utility. Using Safe Harbor requires removing certain fields that are particularly crucial for scientific analysis – in particular, five-digit zip codes and all dates (dates of birth, treatment, symptoms, surgeries, death, etc.)[168] In addition, the Safe Harbor method has been criticized by statisticians for being insufficiently privacy protective in certain circumstances, such as if the data includes unique or unusual data elements not specifically proscribed by the list of 18 disallowed elements.[169] The Expert Determination de-identification method takes a contextual view of the data set and the circumstances relevant to potential re-identification risk, which can create both increased utility of the de-identified data and stronger privacy protections for data subjects.

The US "de-identification" standard is relevant beyond the US. For example, HIPAA de-identification methods, which have been described and analyzed in academic literature, are widely used by health research organizations and commercial organizations in Canada and elsewhere. Also, non-US sites conducting research funded by US agencies have to comply with HIPAA, and international guidelines for public disclosure of clinical trial data rely on the HIPAA de-identification standard. In 2012, following a public workshop, the US Department of Health and Human Services released extensive guidance about de-identification methods and best practices.[170]

---

167. 45 CFR 164.514(b)(1)
168. 45 CFR 164.514(b)(2)
169. El Emam, **Methods for the de-identification of electronic health records for genomic research**, GenomeMedicine, 2011, 3:25
170. **Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule**, US Department of Health & Human Services, available here

Effective 1-1-2020, the California Consumer Privacy Act of 2018 (CCPA) took effect, setting opt-out of sale, access, deletion and additional requirements for consumer data, including health data, but excluding data covered by HIPAA. Health research data when collected by commercial entities will be subject to CCPA and additional new CPRA protections which become effective in CPRA, adding an opt-out of secondary use of sensitive data, data minimization and restrictions to disclosed purposes. Research is exempted from certain obligations when compatible and authorized by consumer consent.

Other noteworthy aspects of secondary use of health data for research

Where research is conducted using any federal funds (broadly construed), the requirements of the Federal Common Rule must also be met. These requirements include IRB approval of the research protocol and other study details, as well as the informed consent of the data subject (a form of consent that is more detailed and broader than the HIPAA authorization). In addition, Food and Drug Administration (FDA) requirements apply to research to be submitted to the FDA regarding a medication or device. Both the Common Rule and FDA requirements include confidentiality mandates.

In addition, state law mandates that exceed or diverge from HIPAA must also be met. The California Consumer Privacy Act of 2018 (CCPA) contains only a very narrow exception for clinical trial data, which comprises only a tiny percentage of all clinical research data. The narrowness of this exception raises troubling questions of how research data that already complies with HIPAA and other federal requirements could also meet the significantly divergent requirements of the CCPA. A California bill to broaden this exemption is pending as of this writing.[171]

171. CA AB 713, June 11, 2020 version, available here