



ETHICAL CONSIDERATIONS SURROUNDING HEALTH-RELATED BIG DATA RESEARCH

CONSIDERACIONES ÉTICAS EN TORNO A LA INVESTIGACIÓN CON DATOS MASIVOS ("BIG DATA") RELACIONADOS CON LA SALUD

POLYCHRONIS VOULTSOS^{1*}

1. *Laboratory of Forensic Medicine & Toxicology (Medical Law and Ethics), School of Medicine, Faculty of Health Sciences, Aristotle University, University Campus, GR-541 24, Thessaloniki, Greece.*

**Corresponding Author (Autor Correspondiente):*

Address: Laboratory of Forensic Medicine & Toxicology (Medical Law and Ethics), School of Medicine, Faculty of Health Sciences, Aristotle University, University Campus, Postal Code 541 24, Thessaloniki, Greece.

Email: pvoultos@auth.gr

Tel.: 0030 2310999285 / 0030 6932727580

ABSTRACT:

Keywords:

Health-related
Big Data research;
Ethics/Ethical Issues;
Artificial Intelligence;
Big Data science; Big
Data analytics.

Recibido: 15/07/2022

Aceptado: 15/01/2023

As health-related big data research (HRBDR) has drastically increased over the last years due to the rapid development of big data analytics, a range of important ethical issues are raised. In this study, a systematic literature review was conducted. Several and interesting results emerged from this review. The term "big data" has not yet been clearly defined. The already existing ethical principles and concepts need to be revisited in the new HRBDR context. Traditional research ethics notions like privacy and informed consent are to be reconsidered. HRBDR creates new ethical issues such those related to trust / trustworthiness and public values such as reciprocity, transparency, inclusivity and common good. The implementation of dynamic consent rather than broad consent is currently highlighted as the more satisfying solution. Ethical review committees in their current form are ill-suited to provide exclusive ethical oversight on HRBDR projects. Expanding Ethical Review Committees' purview and members' expertise, as well as creating novel oversight bodies by promoting a co-governance system including public and all the stakeholders involved are strongly recommended. The mechanism of "social licence", that is, informal permissions granted to researchers by society, can serve as a guideline. High-stakes decisions are often made under uncertainty. Machine learning algorithms are highly complex and in some cases opaque, and may yield biased decisions or discrimination. Improved interdisciplinary dialogue along with considering aspects like auditing, benchmarking, confidence / trust and explainability /interpretability may address concerns about HRBDR ethics. Finally and most importantly, research ethics shifts towards a population-based model of ethics.

RESUMEN:**Palabras clave:**

Investigación de datos masivos relacionada con la salud; Ética/ Cuestiones éticas; Inteligencia artificial; Ciencia de datos masivos; Analítica de datos masivos

Puesto que la investigación de datos masivos relacionada con la salud (HRBDR) ha aumentado drásticamente en los últimos años debido al rápido desarrollo de la analítica de los datos masivos, surgen una serie de cuestiones éticas importantes. En este estudio, se ha realizado una revisión sistemática de la literatura. El término «datos masivos» aún no se ha definido claramente. Los principios y conceptos éticos ya existentes deben ser revisados en el nuevo contexto de la HRBDR. Las nociones tradicionales de la ética de la investigación, como la privacidad y el consentimiento informado, deben reconsiderarse. La HRBDR genera nuevas cuestiones éticas, como las relacionadas con el crédito/la confianza y los valores públicos, como la reciprocidad, la transparencia, la inclusividad y el bien común. Aplicar el consentimiento dinámico antes que el consentimiento amplio se destaca como una solución más satisfactoria. Los comités de revisión ética en su forma actual no son adecuados para proporcionar una supervisión ética exclusiva en los proyectos de HRBDR. Se recomienda vivamente ampliar el ámbito de actuación de los comités de revisión ética y la experiencia de sus miembros, así como crear nuevos órganos de supervisión mediante la promoción de un sistema de cogobernanza, que incluya a todas las partes interesadas y al público en general. El mecanismo de la «licencia social» (permisos informales otorgados a los investigadores por la sociedad) puede servir como guía. Las decisiones de alto riesgo a menudo se toman en situaciones de incertidumbre. Los algoritmos del aprendizaje automático son altamente complejos y en algunos casos opacos, pues pueden producir decisiones sesgadas o discriminación. La mejora del diálogo interdisciplinario, junto con la consideración de aspectos como las auditorías, las evaluaciones comparativas, la confianza y la explicabilidad/interpretabilidad, puede resolver las preocupaciones sobre la ética de la HRBDR. La ética de la investigación cambia hacia un modelo de ética basado en la población.

1. Introduction**1.1. Background**

Health-related big data research (HRBDR) has drastically increased over the last years due to the rapid development of data science and big data analytics. The development of machine learning (ML) and artificial intelligence (AI) technology contributed substantially to the growth of HRBDR. "Big data in healthcare originates from different sources, including biological and social determinants, health records, environmental signals, habits, and behaviors", "telemedicine, electronic health records, wearable, implantable, injectable and ingestible medical devices, health mobile apps, and the application of artificial intelligence (AI) algorithms to health settings"¹. More particularly, "telemedicine and telehealth services expanded significantly in recent

years, particularly during the COVID-19 pandemic. In our everyday life, "the collected information used in Big Data analysis comes from everywhere: grocery shopping, pharmacy purchases, doctor visits, cars, online shopping, and even the robots used to clean our floors. Even more disturbing is that most of this information is collected without any permission"². In this perspective, the philosopher Dennett in an interview published in Financial Times Weekend has expressed his strong concerns about the increasing transparency of our everyday life, and said, "Every human institution, from marriage to the army to the government to the courts to corporations and banks, religions, every system of civilisation is now in jeopardy because of this new transparency." "The "membranes" protecting these institutions have been permeated and we are emerging into a world where it is near-impossible to keep secrets." "People haven't really

1 Cordeiro, J.V. "Digital Technologies and Data Science as Health Enablers: An Outline of Appealing Promises and Compelling Ethical, Legal, and Social Challenges". *Frontiers in Medicine (Lausanne)*. 2021 Jul 8;8:647897.

2 Peirce, A.G., Elie, S., George, A., Gold, M., O'Hara, K., and Rose-Facey W. "Knowledge development, technology and questions of nursing ethics". *Nursing Ethics*. 2020 Feb;27(1):77-87.

come to grips with the fact that it's not just personal privacy that matters, it's also institutional privacy"³.

Furthermore, "AI applications are gaining ground in complementing even the most knowledgeable or skilled professionals"⁴. We should bear in mind that "data shapes the development of Artificial Intelligence (AI)"⁵.

A range of important ethical issues are raised by the growth of HRBDR. Data analytics and artificial intelligence that are transforming health care and research can give rise to significant ethical concerns.

In this study, we attempted a far-reaching literature review of the ethical issues that are raised by the growth of HRBDR, while being as concise as possible in reporting our findings.

1.2. Definition of Big Data

Importantly, the notion "big data" has not yet been clearly defined⁶. Ienca et al. provide a comprehensive definition of what is meant by the term big data: "Big data trends in biomedical and health research enable large-scale and multi-dimensional aggregation and analysis of heterogeneous data sources, which could ultimately result in preventive, diagnostic and therapeutic benefit"⁷. Ristevski and Chen state that "big data analytics is a "promising process of integrating, exploring and analysing of large amount complex heterogeneous data with different nature: biomedical data, experimental data, electronic health records data and social media data"⁸. Several authors include in the definition of big data some terms that start with V (Vs). However, the number of Vs is not the same in the various definitions

of big data. Belani et al. state that big data is defined by "three Vs": volume, velocity and variety⁹. Dereli et al. state that "all characteristics of big data" are "summed up in" "Five Vs": Volume, Velocity, Variety, Veracity, Valorization¹⁰. Baird and Schuller state: "...with *Veracity* (i.e., habitual truthfulness) being one of the 5 Vs (e.g., Velocity, Volume, Value, Variety and Veracity) for defining truly Big Data"¹¹. Ristevski and Chen state, "The term big data is described by the following characteristics: *value, volume, velocity, variety, veracity* and *variability*, denoted as 6 "Vs"¹². Khan et al. state: "There exist various approaches that are addressing issues and challenges of Big Data with the theory of Vs such as 3 V's, 5 V's, 7 V's etc. The objective of this work is to explore and investigate the status of the current Big Data domain. Further, a comprehensive overview of Big Data, its characteristics, opportunities, issues, and challenges have been explored and described with the help of 51 V's"¹³. Furthermore, Ristevski and Chen state, "*Veracity* refers to the data quality, relevance, uncertainty, reliability and predictive value, while *variability* regards about consistency of the data over time. The *value* of the big data refers to their coherent analysis, which should be valuable to the patients and clinicians"¹⁴. Baird and Schuller reasonably argue that "although *true* Big Data is said to need *Veracity*, the reality of this is sometimes different, with large-scale data often showing particular biases toward clustered demographics"¹⁵.

3 Dennett, D. Lunch with the FT: Daniel Dennett. Philosopher Daniel Dennett on AI, robots and religion. Interview by J. Thornhill. *Financial Times Weekend*, 4–5 March 2017, p. 3. Dannett, 2017.

4 Cordeiro, J.V., *op.cit.* 647897.

5 Baird, A. and Schuller, B. "Considerations for a More Ethical Approach to Data in AI: On Data Representation and Infrastructure". *Frontiers in Big Data*. 2020 Sep 2;3:25.

6 Ienca, M., Ferretti, A., Hurst, S., Puhan, M., Lovis, C. and Vayena, E. "Considerations for ethics review of big data health research: A scoping review". *PLoS One*. 2018 Oct 11;13(10):e0204937.; Mooney, S.J. and Pejaver, V. "Big Data in Public Health: Terminology, Machine Learning, and Privacy". *Annual Review of Public Health*. 2018 Apr 1;39:95-112.

7 Ibid.

8 Ristevski, B. and Chen, M. "Big Data Analytics in Medicine and Healthcare". *Journal of Integrative Bioinformatics*. 2018 May 10;15(3):20170030.

9 Belani, S., Tiarks, G.C., Mookerjee, N. and Rajput, V. "'I Agree to Disagree": Comparative Ethical and Legal Analysis of Big Data and Genomics for Privacy, Consent, and Ownership". *Cureus*. 2021 Oct 13;13(10):e18736.

10 Dereli, T., Coşkun, Y., Kolker, E., Güner, O., Ağırbaşı, M. and Özdemir, V. "Big data and ethics review for health systems research in LMICs: understanding risk, uncertainty and ignorance -- and catching the black swans?" *American Journal of Bioethics*. 2014;14(2):48-50.

11 Baird, A. and Schuller, B. *op.cit*

12 Ristevski, B. and Chen, M. *op.cit*.

13 Khan, N., Naim, A., Hussain, M.R., Naveed, Q.N., Ahmad, N. and Qamar, S. "The 51 v's of big data: survey, technologies, characteristics, opportunities, issues and challenges, in COINS' 19" in Proceedings of the International Conference on Omni-Layer Intelligent Systems (Crete: ACM), May 2019, pp: 19–24.

14 Ristevski, B. and Chen, M. *op.cit*.

15 Baird, A. and Schuller, B. *op.cit*

1.3. Big data and personalized medicine

"Personalized medicine means that "one size fits all" is replaced with the "right drug" for the right patient and at the right time"¹⁶. Personalized medicine "focuses on tailoring treatment to suit the patient's unique biological characteristics and genetic makeup"¹⁷. It is a new perception of medicine, according to which, "inter-individual genetic differences help diagnosis, prevention, and treatment of a health-related condition"¹⁸. For instance, "big data and digital health tools can streamline the detection of early signs of cognitive decline"¹⁹. Personalized medicine makes use of "physical, cyber, and social data obtained from a variety of devices, including; wearables, Electronic Medical Records, and the Internet of Things (IoT)"²⁰.

HRBDR is expected to improve the development of the so-called personalized therapeutic and preventive interventions. Cirillo and Valencia highlight that "big Data are radically transforming Personalized Medicine" and "Multi-omics, images, device data, and electronic health records represent the main big data types in biomedical research"²¹. In a study published in 2018 it is stated that while in the last few years the development of health care systems based on the so-called Personalized Medicine has received a great attention, a common and adequate European regulatory framework is lacking²². The regulatory framework for HRBDR and the regulatory framework for Personalized Medicine are conceptually distinct albeit strictly related and overlapping entities.

16 Salari, P. and Larijani, B. "Ethical Issues Surrounding Personalized Medicine: A Literature Review". *Acta Medica Iranica*. 2017 Mar;55(3):209-217.

17 Althobaiti, K. "Surveillance in Next-Generation Personalized Healthcare: Science and Ethics of Data Analytics in Healthcare". *The New Bioethics*. 2021 Dec;27(4):295-319.)

18 Salari, P, and Larijani, B. *op.cit.*

19 Ienca, M., Vayena, E. and Blasimme, A. "Big Data and Dementia: Charting the Route Ahead for Research, Ethics, and Policy". *Frontiers in Medicine (Lausanne)*. 2018 Feb 6;5:13.

20 Althobaiti, K. *op.cit.*

21 Cirillo, D. and Valencia, A. "Big data analytics for personalized medicine". *Current Opinion in Biotechnology*. 2019 Aug;58:161-167.

22 Duardo-Sánchez, A. and De Miguel Beriain, I.. "Personalized Medicine and Medicinal Chemistry: Toward a Legal Framework in the European Union". *Current Topics in Medicinal Chemistry*. 2018;18(25):2165-2173.

Überall and Werner-Felmayer put it best in saying "Moreover, rapid translation into products for the global health market is based on marketable views on health and disease that in turn affect basic research through, for example, funding policies and the research questions being asked. Along with this, biological reductionism is revived fuelling simplified understandings of the genotype phenotype relationship in terms of biology and the human dimension in a broader sense, as well as visions of achieving human perfection through novel biotechnologies". "In current biomedicine, omics technologies drive systems-oriented modes of research to achieve a more holistic and personalized view of health and disease. This shift in scientific approach co-occurs with an era of biocapitalism characterized by markets for biomaterial (e.g., DNA, cells, and tissues) as exploitable resources, high-throughput technologies as tools, and "Big Data" as currency"²³.

It should be noted that pharmacogenomics and the so-called next-generation sequencing (NGS) boosted the so-called "personalized medicine" or "medicine of precision". According to the United States Food and Drug Administration (FDA) personalized medicine is relying on pharmacogenomics²⁴.

1.4. Ethical notions and framework revisited

A range of important ethical issues are raised by the growth of HRBDR. Most importantly, algorithms may cause discrimination or medical errors²⁵. In this context, the already existing ethical principles and concepts need to be revisited in the new HRBDR contexts where the existing regulatory landscape is fragmented. To that effect, several regulatory frameworks have been proposed. Public and individual interests must be properly balanced in the context of health data processing, which

23 , M. and Werner-Felmayer, G. "Integrative Biology and Big-Data-Centrism: Mapping out a Bioscience Ethics Perspective with a S.W.O.T. Matrix". *OMICS, A Journal of Integrative Biology*. 2019 Aug;23(8):371-379.

24 Salari, P, and Larijani, B. *op.cit.*

25 Canales, C., Lee, C. and Cannesson, M. "Science Without Conscience Is but the Ruin of the Soul: The Ethics of Big Data and Artificial Intelligence in Perioperative Medicine". *Anesthesia & Analgesia*. 2020 May;130(5):1234-1243.

is a matter of public interest²⁶. Already in 2017 it is stated in the literature: “ACE Ethics Committee presented a symposium session at the 2016 Epidemiology Congress of the Americas in Miami on the evolving complexities of ethics and epidemiology as it pertains to “big data””²⁷ “Three topic areas were presented: the policy implications of big data and computing, the fallacy of “secondary” data sources, and the duty of citizens to contribute to big data. A balanced perspective is needed that provides safeguards for individuals but also furthers research to improve population health”²⁸.

Big data analytics is a very promising process²⁹. The European Commissioner for Health and Food Safety at the European Parliament states: “Big data has enormous potential to advance medical research, bring about greater innovation in healthcare, and improve the overall performance of health systems”³⁰. Belani et al. regard big data as an ecosystem in the health care context³¹. Importantly, “in the era of precision medicine, the translation of research findings into targeted therapies depends on the availability of big data”³². Precision healthcare is a promising branch, “where the right individual and public intervention is available for the right patient or population at the right time”³³. Importantly, AI can extract health-related information by combining data unrelated to health, e.g. collected through social media or wearable devices. Ferretti, Ienca and Velarde state that this may “challenge traditional research principles such as data privacy, informed consent, scientific validity of research, risk assessment, and distribution of benefits”, as well as “the very notion of human participants which involves physical interaction with research

participants”. Furthermore, the authors state that “new research methods and technological developments” “introduce new epistemic challenges related to the assessment of scientific validity, technological reliability, accountability, fairness, and transparency”³⁴.

It should be highlighted that HRBDR raises ethical issues not only for patients or data subjects, but also for researchers and other members of society³⁵. The rapid development of HRBDR challenges the already existing (and rooted in the predigital era of HRBDR) ethical safeguards for health – related research, such as informed consent, privacy, confidentiality, fair subject selection, and minimal risk³⁶. The already existing safeguards ethical safeguards for health – related research are of limited help and need to be adapted to the new research contexts³⁷. Further, the rapid growth of HRBDR gives rise to a growing number of new ethical issues such those related to trust / trustworthiness and public values such as reciprocity, transparency, inclusivity and common good³⁸. Favaretto and Shaw et al. state that the increasingly developed HRBDR technology not only exacerbates traditional research ethics issues such as informed consent, but also creates new ethical issues e.g. regarding privacy, confidentiality, data security such as protection of data against unauthorized access, protection and anonymiza-

26 Cordeiro, J.V. *op.cit.* 647897.

27 Salerno, J., Knoppers, B.M., Lee, L.M., Hlaing, W.M. and Goodman, K.W. “Ethics, big data and computing in epidemiology and public health”. *Annals of Epidemiology*. 2017 May;27(5):297-301.

28 Ibid.

29 Ristevski, B. and Chen, M. *op.cit.*

30 Andriukaitis, V. “Personalised medicine in Europe”. *Biomedicine Hub*. 2017; 2: 1–40. p. 1

31 Belani, S. et al., *op.cit.*

32 Lang, M., Lemieux, S., Hébert, J., Sauvageau, G. and Zawati, M.H. “Legal and Ethical Considerations for the Design and Use of Web Portals for Researchers, Clinicians, and Patients: Scoping Literature Review”. *Journal of Medical Internet Research*. 2021 Nov 11;23(11): e26450.

33 Cordeiro, J.V. *op.cit.* 647897.

34 Ferretti, A., Ienca, M., Velarde, M.R., Hurst, S. and Vayena, E. “The Challenges of Big Data for Research Ethics Committees: A Qualitative Swiss Study”. *Journal of Empirical Research on Human Research Ethics*. 2022 Feb-Apr;17(1-2):129-143. See also: Friesen, P., Douglas-Jones, R., Marks, M., Pierce, R., Fletcher, K., Mishra, A. et al. “Governing AI-Driven Health Research: Are IRBs Up to the Task?” *Ethics & Human Research*. 2021 Mar;43(2):35-42. doi: 10.1002/eahr.500085. PMID: 33683015.

35 Favaretto, M., Shaw, D., De Clercq, E., Joda, T. and Elger B.S. “Big Data and Digitalization in Dentistry: A Systematic Review of the Ethical Issues”. *International Journal of Environmental Research and Public Health*. 2020 Apr 6;17(7):2495.

36 Ferretti, A., Ienca, M., Hurst, S. and Vayena, E. “Big Data, Biomedical Research, and Ethics Review: New Challenges for IRBs”. *Ethics & Human Research*. 2020 Sep;42(5):17-28. doi: 10.1002/eahr.500065. Erratum in: *Ethics Hum Res*. 2020 Nov;42(6):20.; Ferretti, A., Ienca, M., Sheehan, M., Blasimme, A., Dove, E.S., Farsides, B. et al. “Ethics review of big data research: What should stay and what should be reformed?” *BMC Medical Ethics*. 2021 Apr 30;22(1):51.

37 Parasidis, E., Pike, E. and McGraw, D. “A Belmont report for health data”. *The New England Journal of Medicine* 2019; 380(16): 1493–1495.; Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*; Ferretti, A., Ienca, M., Velarde, M.R. et al., *op.cit.*

38 Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W., Mostert, M., van Delden, J.J.M. “The social licence for data-intensive health research: towards co-creation, public value and trust”. *BMC Medical Ethics*. 2021 Aug 10;22(1):110.

tion, accuracy and accountability in the use algorithms, and discrimination including exacerbation of healthcare inequalities³⁹. "The potential for broad linkage...jeopardizes social rights such as health care, welfare, housing, employment, education, and equal treatment"⁴⁰. HRBDR is not unlike epidemiology research ethics, which calls for respect for traditional research ethics principles and values like 1) autonomy and informed consent, 2) the principle of minimizing harms while maximizing benefits, 3) the principle of justice in terms of fair distribution of burdens and benefits, 4) privacy, 5) ensuring trust and confidentiality, and 6) ensuring scientific robustness⁴¹. Note that HRBDR projects may have high failure rate, for instance due to creating rigid/seamless projects or setting unachievable goals. Besides, it is of great importance that "Big data can be sold for financial gain for commercial enterprise"⁴². However, to the extent that the HRBDR-related ethical issues represent something novel, additional ethical analysis may be required⁴³. For instance, ethical analysis including the risks of discrimination, especially group stigma, and the "challenge of high-stakes decision-making under uncertainty" may be required to be conducted ⁴⁴. The big data era involves not only uncertainties, but also ignorance. It is reasonably argued that "in the big data era, we must be aware of new ways that stored data can cause harm as well as ways they can confer benefit"⁴⁵. Ferratti et al. reasonably argue that it is ethically questionable and should be considered "whether and when big data projects using deidentified data from public databases should require IRB approval, what counts as "public data," what constitutes "minimal-risk" in data-driven projects, and which novel ethical safeguards, if any, are required to ensure ethical big data research"⁴⁶.

In addition, other concerns that are associated with HRBDR involve "difficulties in allowing access to individuals with limited resources" and the fact that "evidence

shows that data query mediation is full of errors and is laborious thus, affecting the reliability of research"⁴⁷. In that regard, it is to be noted that health informational technology can make healthcare safer. For instance, it can facilitate healthcare providers to make better decisions, do the right thing and avoid errors. However, poor health providers' interaction with informational technology can give rise to errors which may be made worse by machine errors⁴⁸.

Ultimately, it should be highlighted that the future of the HRBDR ethics remains still hardly predictable. Car et al. are "realistic that concerns remain about privacy, equity, security, and benefit to all" and "will continue for decades to come"⁴⁹. To that effect, Ienca and Ignatiadis state "Given their transformative nature, it is still largely unclear how AI-driven approaches to the study of the human brain will...affect normative instruments in neuroethics and research ethics"⁵⁰.

2. Methods

2.1. Design

A search according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines was carried out attempting to locate a wide range of literature relevant to the aims of the study.

2.2. Information sources

Relevant articles were retrieved through a systematic searching in electronic database PubMed to identify peer reviewed articles. The final set of search terms used were: ("Big Data"[Majr]) AND ("Ethics"[Mesh] OR "ethics" [Subheading] OR "Ethics, Research"[Mesh] OR "Ethics Committees, Research"[Mesh] OR "Ethics Committees, Clinical"[Mesh] OR "Principle-Based Ethics"[Mesh] OR "Ethics, Clinical"[Mesh] OR "Ethics, Medical"[Mesh] OR "Bioethics"[Mesh] OR "Ethical Theory"[Mesh])

39 Favaretto, M., Shaw, D., De Clercq, E. et al. *op.cit.*

40 Salerno, J. et al. *op.cit.*

41 Ibid.

42 Belani, S. et al., *op.cit.*

43 Salerno, J. et al. *op.cit.*

44 Ibid.

45 Salerno, J. et al. *op.cit.*

46 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* p. 19.

47 Althobaiti, K. *op.cit.*

48 Ibid.

49 Car, J., Sheikh, A., Wicks, P. and Williams, M.S. "Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom". *BMC Medicine*. 2019 Jul 17;17(1):143.

50 Ienca, M. and Ignatiadis, K. "Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges". *AJOB Neuroscience* 2020 Apr-Jun;11(2):77-87.

Additional articles were identified through other sources.

2.3. Eligibility criteria

Papers were included if they were: 1) Published in peer reviewed journals; 2) Written in the English language; or if they 4) Contained data involving essential or important knowledge relevant to the aims of the study.

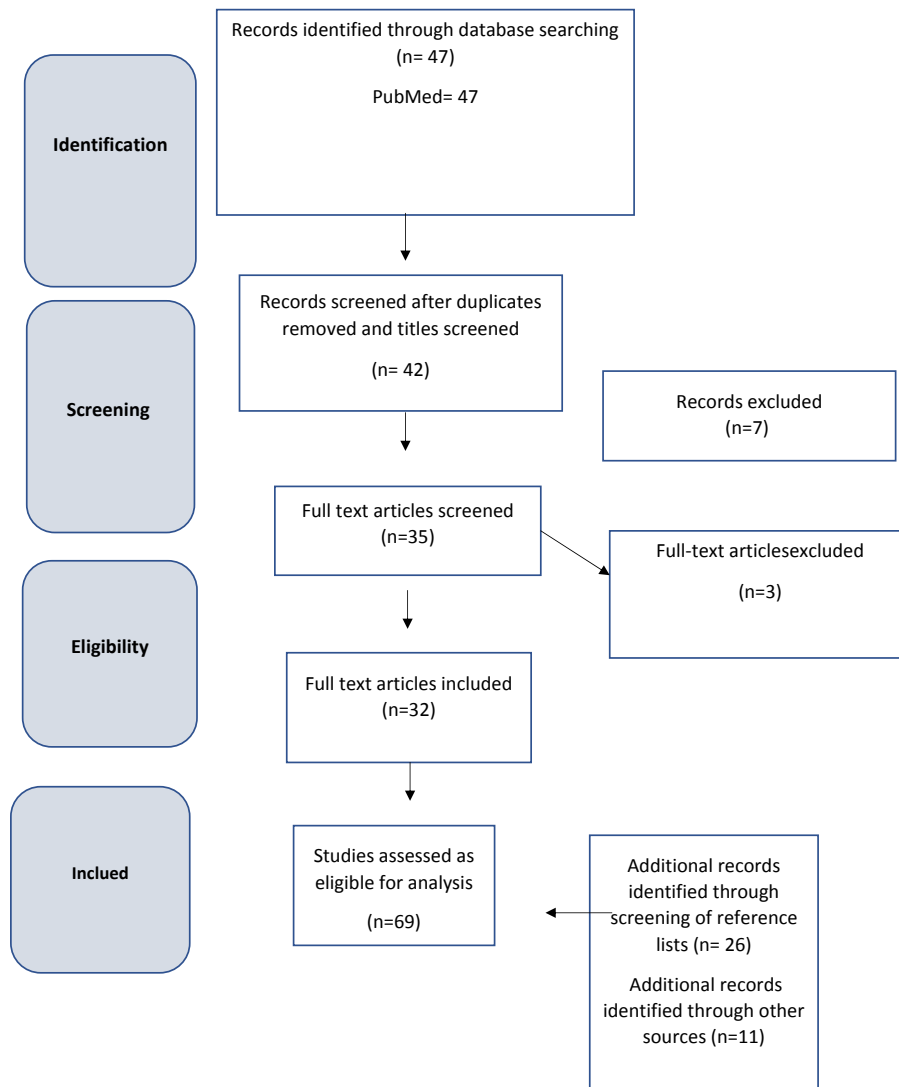
Papers were excluded if data were published in a way that data relevant to the aims of the study could not be deciphered from overall reported data. Furthermore, papers were excluded if they: 1) Reported data

on the topic of interest which, however, do not make substantial contribution to the review; 2) Were focused purely on presenting, analysing and interpreting scientific data, which however were not related directly or indirectly to our research questions.

2.4. Study selection and data extraction

The documents identified through database searching were screened to identify documents that might meet the inclusion criteria. Therefore, the relevant documents were read through carefully. Titles, abstracts and full texts of the records were screened by the author

Fig. 1



of this study who selected those assessed as eligible for further analysis according to the inclusion / exclusion criteria. Reference lists and citations of eligible articles were also screened and reviewed for additional papers. The literature database searching resulted in a total of 47 documents. Forty two records screened after duplicates removed and titles screened. Additional records identified through screening of reference lists (n= 26). Besides, additional records identified through other sources (n=11). At the end 69 articles that assessed as eligible for review according to the inclusion / exclusion criteria were retained for further analysis (see flow chart over systematic review process, Fig. 1).

PRISMA Flow Diagram. *From: Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. "The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." PLoS Med 2009; 6 (6): e1000097. doi: <https://doi.org/10.1371/journal.pmed1000097>*

3. Results

3.1. The concept privacy and privacy protection

Mooney et al. put it best in saying "formal definitions of privacy have been inconsistent, from "the right to be left alone" in 1890 to the late 1960's idea that privacy amounted to control over the information one produces to more recent notions defining non-intrusion, seclusion, limitation, and control as separate categories of privacy"⁵¹. In the era of health-related big data research the privacy landscape has been changed. HRBDR has further complicated the concept of privacy, which is a "broad concept, and difficult to define categorically"⁵². In that connection, Mooney et al. state that emerging technologies such as GPS, drones, or social media may shift the generally accepted ethical standards regarding privacy⁵³. The availability of big data and the rapid development of health-related big data research

have given rise not only to ever-growing, but also diversifying privacy concerns among the public⁵⁴. Indeed, the access and transmission of information has become much easier in the HRBDR context due to advancements in technology⁵⁵. The use of data anonymization, access control, and cryptographic techniques can prevent data privacy breaches⁵⁶. "Anonymization occurs when no one ever knows the individual's identity"⁵⁷ Confidentiality is "information that is only available to authorized individuals" and "can be assured when privacy is attained"⁵⁸.

However, anonymization (de-identification) is not a sufficient firewall to reassure privacy in the context of HRBDR⁵⁹. Sufficiently skilled offenders can re-identify anonymized data⁶⁰. In that regard, it is argued that developing technical methods to better address the privacy concerns, such as de-identification methods that are less inadequate than the current ones, is becoming increasingly challenging in the HRBDR context⁶¹. Furthermore, high risk of disclosure of sensitive personally identifiable information may be due to inadvertent disclosure, e.g. by the use of online tools⁶². Moreover, "big data expands the pool of information, further increasing the vulnerability to re-identification"⁶³. "Deductive disclosure", namely, "merging two datasets that are each successfully anonymized may result in a dataset in which subjects can be personally identified"⁶⁴. When datasets "are combined with internet search queries, social network data, or even facial recognition, anonymity can be easily lost"⁶⁵. Althobaiti states that "cross-referencing of data with other databases" and "long-life span of certain anonymized datasets" may significantly increase the risk of de-identifi-

54 Mooney, S.J. and Pejaver, V. *op.cit.*

55 Belani, S. et al., *op.cit.*

56 Gürsoy, G., Harmanci, A., Tang, H., Ayday, E. and Brenner, S.E. "When Biology Gets Personal: Hidden Challenges of Privacy and Ethics in Biological Big Data". *Pacific Symposium on Biocomputing*. 2019;24:386-390.

57 Belani, S. et al., *op.cit.*

58 *Ibid.*

59 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*; Belani, S. et al., *op.cit.*

60 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

61 Dankar, F.K., Gergely, M. and Dankar, S.K. "Informed Consent in Biomedical Research". *Computational and Structural Biotechnology Journal*. 2019 Mar 25; 17: 463-474.

62 Mooney, S.J. and Pejaver, V. *op.cit.*

63 Belani, S. et al., *op.cit.*

64 Mooney, S.J. and Pejaver, V. *op.cit.*

65 Belani, S. et al., *op.cit.*

51 Mooney, S.J. and Pejaver, V. *op.cit.*

52 McKeown, A., Mourby, M., Harrison, P., Walker, S., Sheehan, M. and Singh, I. "Ethical Issues in Consent for the Reuse of Data in Health Data Platforms". *Science and Engineering Ethics*. 2021 Feb 4;27(1):9.

53 Mooney, S.J. and Pejaver, V. *op.cit.*

cation⁶⁶. Indeed, cross-referencing multiple databases is a new route for breaching privacy⁶⁷. “Deductive disclosure” may be facilitated by high dimensionality of data. For instance, “high dimensionality of the – omics data means, that there have many more dimensions or features than the number of samples”⁶⁸. It is noticeable that even when people do not know things about others, they make inferences about them⁶⁹. This is a subtle issue⁷⁰ However, it is of great importance in the context of HRBDR.

Protecting the privacy and confidentiality of patient data requires the adoption of security safeguards and data confidentiality protections⁷¹. Not surprisingly, there has been a shift “from only assessing privacy to investigating risks”⁷². To address security- and privacy-related ethical challenges in the HRBDR context, “the big data analytics software solutions should use advanced encryption algorithms and pseudo-anonymization of the personal data”⁷³.

To harmonize the regulations and standards that are used to protect privacy and provide more security to the health information that is secured in the electronic health records, there have been proposed various encryption algorithms⁷⁴. Encryption and blockchain are efficient privacy-preserving technical measures to reassure data security and thereby sustain the trust of potential data subjects in science⁷⁵. Blockchain technology helps to overcoming some of the privacy challenges⁷⁶. Moreover, the ideal of controllability needs not only the implementations of blockchain technology, but also the so-called *differential privacy*, namely, providing “algorithms and IT infrastructures that make coarse-grained data available while keeping information on individuals private”⁷⁷.

66 Althobaiti, K. *op.cit.*

67 Gürsoy, G. *op.cit.*

68 Risteovski, B. and Chen, M. *op.cit.*

69 Price, W.N. 2nd, Cohen, I.G. “Privacy in the age of medical big data”. *Nature Medicine*. 2019 Jan;25(1):37-43.

70 *Ibid.*

71 Althobaiti, K. *op.cit.*

72 *Ibid.*

73 Risteovski, B. and Chen, M. *op.cit.*

74 Keshta, I. and Ammar Odeh, A. “Security and privacy of electronic health records: Concerns and challenges”. *Egyptian Informatics Journal* 2021; 22(2): 182.

75 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

76 Belani, S. et al., *op.cit.*

77 Hummel, P. and Braun, M. “Just data? Solidarity and justice in data-driven medicine”. *Life Science Society and Policy*. 2020 Aug 25;16(1): 8.

Given that “big data is commonly, stored in *centralized* infrastructures which limit transparency”, Baird and Schuller state that “democratic, decentralized (i.e., peer-to-peer blockchain-based) approaches” are necessary⁷⁸. Baird and Schuller state that “through social-media (which is in some sense a decentralized network for communication) group morality is developed”. Furthermore, the authors highlight that “a more transparent and open platform makes masking potential network biases a challenge”⁷⁹. Moreover, it is highlighted the “interdisciplinarity in AI research”⁸⁰. From the literature overview that was conducted by Baird and Schuller, emerged that in the AI research context, “incorporating multiple disciplines in the discussion appears to be more prominent with those promoting *decentralized* AI”⁸¹. Interdisciplinarity contributes to tackling “ethical concerns relating to; (i) integration, (ii) *selection-bias*, and (iii) *trust*”⁸².

Various regulations, standards and strategies are used to protect privacy and provide more security to the health information that is secured in the electronic health records. These regulations and standards must be harmonized to avoid inconsistencies between them. To that effect, there have been proposed various encryption algorithms⁸³. “The theme of administrative safeguard is the first safeguard that comprise of relevant techniques like performing audits, employing an officer in charge of information security, and coming up with contingency plans”⁸⁴.

Individual privacy cannot be entirely protected, especially in the HRBDR context. However, it would be helpful if the associated risks are further understood and all the stakeholders involved in a research project are secured that eventual privacy violations will be prosecuted accordingly⁸⁵.

78 Baird, A. and Schuller, B. *op.cit.*

79 *Ibid.*

80 *Ibid.*

81 *Ibid.*

82 *Ibid.*

83 Keshta, I. and Ammar Odeh, A. *op.cit.* p.182.

84 *Ibid.*, p. 180.

85 Lucassen, A., Montgomery, J. and Parker, M. “*Ethics and the social contract for genomics in the NHS*” In: Davies SC, Annual Report of the Chief Medical Officer 2016, Generation Genome. Department of Health, London. 2017, Chap. 16, p. 4.

Furthermore, the degree of privacy-justified secrecy cannot always be easily determined. Price et al. state that “while too little privacy raises concerns it is also true that too much privacy in this area can pose problems”⁸⁶. Not surprisingly, the same holds for the ethical considerations surrounding Personalized Medicine⁸⁷. For instance, “privacy-justified secrecy can erode trust in already-opaque big-data innovations”⁸⁸. “The basic harm of privacy overprotection is the brakes it puts on data-driven innovation”⁸⁹. Privacy protections limit both data aggregation and innovative data use. In addition, it should be highlighted that “...privacy-focused ethical oversight may be insufficient to address other challenges raised by big data”⁹⁰.

Ending up, this section of the study that provides information about the concept of privacy and privacy protection should not come to an end without including some reflections about the new concept of pseudonymization which has been included by the GDPR and changes the paradigm between the formula of anonymization v. new forms of dynamic consent (see below). The General Data Protection Regulation (GDPR), namely, the new EU-wide data protection law with tough provisions, explicitly introduces pseudonymization of personal data as one of the de-identification procedures to lessen the chance that identifiers, either alone or in combination, could lead to identification of a data subject⁹¹. Article 4 of the GDPR provides that the pseudonymization is “the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organizational measures to ensure non-attribution to an identified or identifiable individual”⁹². The processing of personal data in such a manner allows data processors to use personal data

for purposes different from those for which they were originally intended without violating privacy laws. Pseudonymization involves removal of all direct and many indirect identifiers, namely, identifiers that may identify a data subject if combined with additional data points. As some indirect identifiers may remain after removing identifiers, pseudonymization is not anonymization and “is not intended to preclude any other measures of data protection”⁹³ (Recital 28). Although under the GDPR encryption is not mandatory, GDPR repeatedly highlights the importance of using “appropriate technical and organizational measures” of personal data security.

Ultimately and most importantly, the GDPR, in Recital 26 considers whether a procedure of “reidentification is *reasonably likely* to be used...to identify the natural person directly or indirectly” (emphasis added)⁹⁴. In a case of pseudonymization, if the remaining indirect identifiers do not pose high risk of identifying a data subject, reidentification could not be regarded as *reasonably likely* to occur. This indicates that the GDPR provides flexibility to some extent.

Note, however, that it is not always easy for data processors to implement pseudonymization properly and effectively, especially in the context of modern clinical research. To illustrate this point, a study recently published van Gastel et al. is worthy to be mentioned here. The authors state, “In practice, there are a number of constraints on how pseudonimisation can be applied. There are a number of sources these constraints originate from: biological data properties, from standard practices such as how to handle incidental findings, and how bio samples are handled”⁹⁵. This is especially the case of larger studies, such as the called Personalized Parkinson Project, which conducted by the authors, where “participants are monitored during a longer period of time and are (re-)invited for multiple interviews and examinations”⁹⁶. van Gastel, Jacobs and Popma

86 Price, W.N. 2nd and Cohen, I.G. *op.cit.*

87 Salari, P, and Larijani, B. *op.cit.*

88 Price, W.N. 2nd and Cohen, I.G. *op.cit.*

89 Ibid.; Althobaiti, K. *op.cit.*

90 Ferretti, A., Ienca, M., Velarde, M.R. et al. *op.cit.*

91 Regulation (EU) 2016/679. [On line publication] “General Data Protection Regulation (GDPR)”. 2026. Retrieved from <https://gdpr.eu/article-53-conditions-of-supervisory-authority/> [Consulted: 07/11/2022]

92 Ibid.

93 Ibid.

94 Ibid.

95 van Gastel, B.E., Jacobs, B. & Popma, J. “Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson’s Disease Study”. *Journal of Parkinson’s Disease*. 2021;11(s1):S19-S25. doi: 10.3233/JPD-202431. PMID: 34092652; PMCID: PMC8385496.

96 Ibid.

suggested “an advanced form of pseudonymization”, which they called “Polymorphic Encryption and Pseudonymisation” and used “in a large cohort study on Parkinson’s disease, called Personalized Parkinson Project”⁹⁷. According to the authors, the proposed form of pseudonymization, that is, the so-called “Polymorphic Encryption and Pseudonymisation”, is “a stronger form of pseudonymisation based on asymmetric encryption, in such a way that enables sharing of data amongst different researcher groups, while not relying on a third party for pseudonymisation”⁹⁸.

3.2. Data sharing policies

Data sharing policies is a topic that needs to be explored in connection with the issue of privacy and privacy protection. Importantly, sharing data policies are not unproblematic policies. Murtagh et al. “argue that three key structural features are foundational for practising responsible data sharing: independence and transparency; interdisciplinarity; and participant-centric decision-making”⁹⁹. Deverka et al. conducted public deliberations in the U.S. recruiting deliberants from diverse geographic areas of the country. Deliberants “expected robust data security and the protection of privacy as a fundamental condition for sharing data”¹⁰⁰. Ensuring a robust, independent and transparent data access process especially for the most complex and sensitive research resources serves as a safeguard to ensure the right to privacy. The role of independent research ethics committees is highlighted¹⁰¹. Note, however, that there are not homogenous share decision policies. A review conducted by Barbui revealed a split between top medical journals’ policies regarding data sharing. There were identified journals that “do not require any statement to be published along with the study report on the pos-

sibility to access the raw data”, journals that “encourage data sharing and require a formal statement describing the conditions under which raw data are accessible”, and journals that require “full availability of all data underlying the findings described in published study reports”¹⁰².

Ultimately, it is noticeable that the problematic of sharing data in the HRBDR context touches upon the eternal philosophical debate on the fundamental concepts of good and right. Hummel and Braun highlight that promoting the *good* through data sharing practices, that means, making health data available, can considerably compromise the *right*. The authors notice that while contributing to HRBDR shows solidarity and promotes the common good, it can also “compromise privacy, social equality, and fairness”. Furthermore, they state that “solidarity and justice share certain features, but also differ in crucial respects”¹⁰³. As data sharing is made at the expense of the right, there is a need for aligning the good and the right in the HRBDR context, i.e. by reducing the so-called “algorithmic injustice”¹⁰⁴.

3.3. The concept of ownership

Ownership can be described as, “... both the possession and responsibility for information”¹⁰⁵. It is arguably stated that the ownership of genetic data “is not a straightforward issue” in the HRBDR context¹⁰⁶. Indeed, the notion of ownership of information has become increasingly vague and questionable in the context of HRBDR¹⁰⁷. Furthermore, it should be noted that in the context of HRBDR more than one person are involved in the data creation process¹⁰⁸.

Already in 2017 Salerno et al. stated that “individuals in a health care system have a reciprocal duty not

97 Ibid.

98 Ibid.

99 Murtagh, M.J., Blell, M.T., Butters, O.W., Cowley, L., Dove, E.S., Goodman, A., et al. “Better governance, better access: practising responsible data sharing in the METADAC governance infrastructure”. *Human Genomics*. 2018 Apr 26;12(1):24.

100 Deverka, P.A., Gilmore, D., Richmond, J., Smith, Z., Mangrum, R., Koenig, B.A., et al. “Hopeful and Concerned: Public Input on Building a Trustworthy Medical Information Commons”. *Journal of Law, Medicine & Ethics*. 2019 Mar;47(1):70-87.

101 Murtagh M.J. et al. *op.cit.*

102 Barbui, C. “Sharing all types of clinical data and harmonizing journal standards”. *BMC Medicine*. 2016; 14:63.

103 Hummel, P. and Braun, M. *op.cit.*

104 Ibid.

105 Belani, S. et al., *op.cit.*

106 Stoeklé, H.C., Mamzer-Bruneel, M.F., Frouart, C.H., Le Tourneau, C., Laurent-Puig, P., Vogt, G., and Hervé, C. “Molecular Tumor Boards: Ethical Issues in the New Era of Data Medicine”. *Science and Engineering Ethics*. 2018 Feb;24(1):307-322.

107 Salari, P. and Larjani, B. *op.cit.*

108 Mooney, S.J. and Pejaver, V. *op.cit.* Therefore, these persons have some right to the data, Mooney, S.J. and Pejaver, V. *op.cit.*

to oppose, and perhaps even to facilitate the credible collection and analysis of their data and information for their own treatment and that of others who follow them¹⁰⁹. Note, however, that intense feelings of social solidarity and the concept of collective altruism exert moral pressure on people for contributing data. This may be ethically questionable¹¹⁰. Furthermore, and most importantly, data subjects' insufficient control over their highly sensible health-related information is a disincentive for people to contribute their data¹¹¹. At any rate, to get people contribute data, they may be given something in return for it. While this coercive offer is in line with the principle of reciprocity, it remains questionable whether it is ethically justified or unacceptable practice¹¹². Some years ago Prainsack et al. proposed "an approach to governance that recognises people's willingness to participate in a public research biobank...moving beyond overly restrictive and burdensome, exclusively autonomy-based governance towards governance that is reflective of people's willingness to accept costs to assist others"¹¹³. In a similar vein, Dankar et al. argue for a solidarity/community-based governance in which solidarity urges people to be willing to engage in risky activities such as accepting high levels of risk and uncertainty for the benefit of others¹¹⁴. Responsible data governance that includes protection of privacy and confidentiality, and respect for the "agreed and anticipated parameters" enhances the trustworthiness of the research in the eyes of the public¹¹⁵. Not surprisingly, trust is reported as a significant predictor of people's willingness to contribute their data¹¹⁶. Public trust is particularly crucial for ensuring support for HRBDR, especially for longitudinal studies¹¹⁷. Note, however, that the notion

of trust has been reconceptualized in the personalized medicine context. It has emphasized the role of "trust facilitators" such as: technological innovations to minimize risk of re-identifying data, governance practices that ensure transparency, respect for autonomy, and participant engagement¹¹⁸. Patients' understanding of HRBD projects, electronic informed consent (e-IC), or dynamic consent system, and effective communication where "patients feel that they can communicate freely, anywhere, at any time and in complete safety" increase their trust in researchers¹¹⁹.

It is argued that "data ownership is at the heart of the ethical debate behind informed consent"¹²⁰. "While ownership is often discussed as the right of the subjects to control their data, it can also refer to owning intellectual property produced from using the data"¹²¹. Althobaiti highlights the HRBDR-related concern about the "intellectual property ownership obtained from the analysis of data sets"¹²². "Right of data control is defined as the subjects' (unlimited) right to track all data uses and restrict undesired uses. Naturally, this includes understanding who holds what data, and for what purpose"¹²³. "Data control is also sometimes interpreted to include the right to data guardianship, in other words, individuals hold on to their data, and decide to grant access (to family, health authorities, or research facilities) as they see fit"¹²⁴.

3.4. The concept of "minimal risk"

For the US jurisprudence, the ethical standard "minimal risk" is meant as harm or discomfort not greater than what is inherent in routine clinical practice¹²⁵. Other countries' regulations define it in a more stringent manner, thus making "broad consent" difficult to be applied¹²⁶. Not only normative, but also conceptual ad-

109 Salerno, J. et al. *op.cit.*

110 Middleton, A. "Society and personal genome data". *Human Molecular Genetics*. 2018 May 1;27(R1):R8-R13.

111 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

112 Middleton, A. *op.cit.*

113 Prainsack, B. and Buyx, A. "A solidarity-based approach to the governance of research biobanks". *Medical Law Review*. 2013 Winter;21(1):71-91.

114 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

115 Murtagh M.J. et al. *op.cit.*

116 Deverka, P.A. et al. *op.cit.*

117 Murtagh M.J. et al. *op.cit.*

118 Deverka, P.A. et al. *op.cit.*

119 Stoeklé, H.C. et al. *op.cit.*

120 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

121 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

122 Althobaiti, K. *op.cit.*

123 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

124 Ibid.

125 Monach, P.A. and Branch-Elliman, W. "Reconsidering 'minimal risk' to expand the repertoire of trials with waiver of informed consent for research". *BMJ Open*. 2021 Sep 14;11(9):e048534.

126 Ibid.

justments are required for making the standard “minimal risk” applicable to HRBDR ethics¹²⁷.

At any rate, the concept of minimal risk is difficult to be defined in the HRBDR context. Importantly, in studies involving “seemingly innocuous” data, for instance data which are collected from publicly available platforms, e.g. social media, the “risk of reidentification is potentially greater than minimal risk”¹²⁸. Furthermore, “...Hyder and colleagues (2014) in their target article suggest that risk assessment should be, part of HSR ethics review, but we respectfully emphasize, given the very dynamic “five V” attributes of big data, uncertainty, ignorance, and at times black swan events will and ought to be considered in the course of a sound big data HSR ethics review”¹²⁹. “*Black swans* are rare outlier events with massive impacts on society, which cannot be predicted a priori, especially because they have no precedence, and consequently fall outside usual cognitive imaginative capacities and expectations. For instance, the earthquake in Japan on March 11, 2011, for example, can be considered a “black swan” event”¹³⁰.

3.5. The concept of “data subject”

Moreover, HRBDR has blurred the concept of “human subject”, which has been shifted towards the concept of data subject in the new research context¹³¹. HRBDR has changed the dynamics between researchers and participants. “Big data, together with modern analytic methods used to interpret the data, creates novel dynamics between researchers and participants”¹³². In line with US law and research tradition, projects that do not directly involve human subjects, e.g. projects that involve non-identifiable data or publicly available data, are regarded as being exempted from the purview of Research Ethics Review Committees (ERCs)¹³³.

127 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* p.25.
 128 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*
 129 Dereli, T. et al. *op.cit.*
 130 Ibid.
 131 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*
 132 Ibid.
 133 Ibid.

3.6. New informed consent

Furthermore, in the HRBDR context there is an important consent-related ethical challenge. Data subjects are not cognizant of their participation in new studies with goals which were unknown at the time of the original data donation. Indeed, in studies involving data from million subjects, specific informed consent is impractical to obtain and re-obtain¹³⁴. In the HRBDR context it has become unfeasible to pre-define “an array of future uses on the dataset limits creativity”¹³⁵. Future research is surrounding by uncertainty. As a consequence, the data subjects become particularly vulnerable to potential risks related to HRBDR. This consent-related challenge leaves an ethical gap and heightens the already existing uncertainty¹³⁶. Besides, research ethics shifts towards a population-based model of ethics. Ballandyne is correct in stating that “Shifting from a research ethics perspective to a public health lens brings a different set of issues into view”, including the question “who has legitimate decision-making capacity?”¹³⁷. Furthermore, as mentioned above, in studies involving “seemingly innocuous” data the “risk of reidentification is potentially greater than minimal risk”¹³⁸. Hibbin et al. in their study cite number of scholars who do not espouse the opinion that social media data fulfill the definition of human participation research¹³⁹. Hibbin et al. state that this remains “an ongoing live issue in the literature”¹⁴⁰. Furthermore, Hibbin et al. state, “However, different views on the publicly available nature of SM [social media] platforms and the perceived privacy of SM users present an ethically gray area for research ethics committees (RECs) charged with ensuring that research using SM data is conducted responsibly and ethically”¹⁴¹.

134 Ibid.

135 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

136 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*

137 Ballandyne, A. “Adjusting the focus: a public health ethics approach to data research”. *Bioethics*. 2019;33(3):357–366.

138 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*

139 Hibbin, R.A., Samuel, G. and Derrick, G.E. “From “a Fair Game” to “a Form of Covert Research”: Research Ethics Committee Members’ Differing Notions of Consent and Potential Risk to Participants Within Social Media Research”. *Journal of Empirical Research on Human Research Ethics*. 2018 Apr;13(2):149-159.

140 Ibid.

141 Ibid.

Conventional informed consent models in their current shape appear “ill suited” for large-scale HRBDR projects that involve large-scale health-related data, structured or unstructured.¹⁴²

Tiered informed consent is one of the new forms of consent that have been proposed as alternative to the traditional specific informed consent. “Tiered consent procedures allow potential participants to at least partly tailor their consent preferences around general categories”¹⁴³. Compared to the model of broad consent (see below), in tiered consent “the participant is able to be more specific about the uses that she is consenting to”¹⁴⁴. “Tiered informed consent can address many...challenges by providing sufficient information about intended specimen/data use, storage, and opportunity to withdraw from the study; and ensuring that participants can individually select a level of specimen and/or data sharing with which they are comfortable by responding to a series of specific questions that address different levels of use and onward specimen/data sharing”¹⁴⁵. Tiffin writes that “tiered consent provides individuals with the autonomy to participate in a study at a risk level with which they are comfortable, whereas broad consent demands exposure to maximum risk in order to participate”¹⁴⁶. The author observes that “tiered consent is variously described as too difficult and time-consuming to explain to participants—especially where translation is required, too difficult to capture and store electronically, and too difficult to query when assembling data sets for secondary use. Further, the author comments: “The reluctance to promote tiered consent seems to entrench the ideology that ‘one size fits all’”¹⁴⁷. To implement tiered consent, participants should “have the heterogeneity, individuality or imagination that might lead one participant to choose to share their

data and biospecimens in all future research”¹⁴⁸. Tiered consent seems to have not gained great acceptance in the research context.

To address this problem, scholars developed the alternative form of broad consent, which gained acceptance in the research context. Broad consent, as understood by Mikkelsen et al., “is consent obtained at the time of enrolment in the biobank against a background of assurances about the overall scope and aims of the biobank as well as its governance”¹⁴⁹. In this point of view, it is widely accepted that data which are collected from publicly available platforms, e.g. social media, can be used and re-used by researchers relying on a form of broad consent or consent waivers, without IRB approval¹⁵⁰. Broad consent differs from blanket consent “as data subjects do not give permission for any use of their data but rather define in broad terms the purposes of use”¹⁵¹. At any rate, it is important to bear in mind that as the future research, namely, the secondary research is unspecified, broad consent cannot be entirely informed.

“Ethical approval for studies using broad consent includes mechanisms to ensure that those consents are respected and the expectations inherent in them are maintained, for example, explicitly stating which bodies can approve data and sample access”¹⁵². In that regard, it is important to bear in mind that in the AI context “Informed consent can be complex given uncertainties, fears, or even overconfidence about uses of AI”¹⁵³. More precisely, while broad consent increases transparency, investigators need to consider whether adopting broad consent “may discourage research participation or engender unreasonable expectations regarding future return of results.” “IRBs and their institutions face additional challenges”¹⁵⁴.

142 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

143 Mikkelsen, R.B., Gjerris, M., Waldemar, G. and Sandøe, P. “Broad consent for biobanks is best - provided it is also deep”. *BMC Medical Ethics*. 2019 Oct 15;20(1):71.

144 Ibid.

145 Nembaware, V., Johnston, K., Diallo, A.A., Kotze, M.J., Matimba, A., Moodley, K. et al. “A framework for tiered informed consent for health genomic research in Africa”. *Nature Genetics*. 2019 Nov;51(11):1566-1571.

146 Tiffin, N. “Tiered informed consent: respecting autonomy, agency and individuality in Africa”. *BMJ Global Health*. 2018 Dec 17;3(6):e001249.

147 Ibid.

148 Ibid.

149 Mikkelsen, R.B. et al. *op.cit.*

150 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*

151 Geneviève, L.D., Martani, A., Wangmo, T., Paolotti, D., Koppeschaar, C., Kjelsø, C. et al. “Participatory Disease Surveillance Systems: Ethical Framework”. *Journal of Medical Internet Research*. 2019 May 23;21(5):e12273..

152 Murtagh M.J. et al. *op.cit.*

153 Schiff, D. and Borenstein, J. “How Should Clinicians Communicate With Patients About the Roles of Artificially Intelligent Team Members?” *AMA Journal of Ethics*. 2019 Feb 1;21(2):E138-145.

154 Fisher, C.B. and Layman, D.M. “Genomics, Big Data, and Broad Consent: a New Ethics Frontier for Prevention Science”. *Pre-*

Ferretti et al. state that “the literature reports an increasing number of authors who are against the idea of a new consent form for big data studies”¹⁵⁵. Given the issues raised by broad consent and the fact that it requires initial face-to-face contact, seeking consent electronically could be an ethically satisfying alternative to traditional informed consent. eConsent is a flexible dynamic form of consent.

Dynamic consent is another approach to consent that claims to be consistent with the three pillars of informed consent. It is based on digital interface and “ensures that face-to-face contact is not necessary for each renewal of consent”¹⁵⁶. “Dynamic consents are personalized online consent and communication platforms, and use modern IT to provide a communication channel between researchers and participants of a project. This allows consent documents to be tied to events in real-time, as they occur in the data life cycle. Such framework enables alerting individuals of new research opportunities, and allows participants to accept or decline participation in a research project if they match the research’s specific profile, moreover, it enables sponsors to conform their protocols to the dynamic privacy laws”¹⁵⁷. Importantly, Budin-Ljøsne et al. state that whilst “Dynamic Consent offers opportunities for ongoing communication between researchers and research participants” and “it is relatively easy to set up and maintain, its implementation will require that researchers re-consider their relationship with research participants....”¹⁵⁸. Furthermore, the authors argue that “With Dynamic Consent, informed consent is not restricted to a functional or legal instrument, but also becomes a social agreement between researchers and research participants”¹⁵⁹.

Dynamic / electronic consent protects the participant’s autonomy better than the well-discussed broad

consent¹⁶⁰. Deverka et al. conducted public deliberations in the U.S. recruiting deliberants from diverse geographic areas of the country. Compared to the consent model of broad consent, consent models such as dynamic consent received more support from deliberants¹⁶¹. Hummel and Braun state, referring to data sharing, that the ideal of controllability is in tension with the so-called “broad consent” and “motivates *dynamic* consent procedures which allow individuals to exercise continuous control”¹⁶².

Nevertheless, dynamic / e-consent has some inherent problems, mainly related to the absence of personal interaction, namely, the absence of face-to-face contact. Participants may give up their consent without having fully understood and may be less likely to consent by signing out from the electronic platforms¹⁶³. Further, it is not easy to ensure that participants are the authentic to give their consent legitimately. Note that this can be addressed by using biometric identification technologies, which however, raise privacy-related ethical and legal issues¹⁶⁴.

eConsent conveys adequate information related to research and asks for documented informed consent, through electronic devices. eConsent improves the participant’s comprehension and decentralizes/democratizes the research project. In that connection, it is to be noted that studies have shown that “participants have discrepancies in their understanding of consent information”¹⁶⁵. eConsent enhances participant acceptance rate, reduces study drop-out, and enhances inclusion and diversity of study participants While it is broadly implemented in USA, in Europe its implementation lags behind due to regulatory barriers, fragmentation and uncertainty. It is a flexible and dynamic consent, which can protect participants’ rights without compromising on the potential of research.

Ivanova and Katsaounis have recently proposed a novel tool: The dynamic real-time e-consent. “The e-consent system operates under GDPR and HIPAA (Health In-

vention Science. 2018 Oct;19(7):871-879.

155 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*

156 Mikkelsen, R.B. et al. *op.cit.*

157 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

158 Budin-Ljøsne, I., Teare, H.J., Kaye, J., Beck, S., Bentzen, H.B., Caenazzo, L. et al. “Dynamic Consent: a potential solution to some of the challenges of modern biomedical research” *BMC Medical Ethics*. 2017 Jan 25;18(1):4.

159 Ibid.

160 Geneviève, L.D. et al. *op.cit.*; Althobaiti, K. *op.cit.*

161 Deverka, P.A. et al. *op.cit.*

162 Hummel, P. and Braun M. *op.cit.*

163 Geneviève, L.D. et al. *op.cit.*

164 Ibid.

165 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

insurance Portability and Accountability Act) compliance". "The implementation of dynamic e-consent tools can improve the consent process and transparent accessibility to relevant clinical information"¹⁶⁶. "This dynamic real-time (DRT) e-consent provides an evolution of patients' engagement in research and the ability to obtain real-time longitudinal data, including clinical secondary data"¹⁶⁷.

At any rate, McKeown et al. are correct in stating, "the consent for reuse issue is significant for three reasons in both the general context of health research and the specific context of where this is carried out via a data platform. First, understanding the aetiology of diseases requires their study longitudinally. Second, the apparent power of big data analytics derives from its ability to make novel predictive inferences across datasets about the interactions of disparate risk factors. Third, this iterative novelty limits what can be communicated to participants about the purposes for which their data may be used"¹⁶⁸.

Ultimately and most importantly, it should be highlighted that scholars propose novel mechanisms to be considered, such as "data portability rights", or "electronic consent management mechanisms and participatory forms of data governance" to enable more effective data control on the part of data subject¹⁶⁹. Alternatives to the specific informed consent have been proposed through novel governance models in the HRBDR context, such as ownership-based governance, solidarity/community-based governance, and technical mechanisms that "attempt to perform data analysis in a privacy preserving manner without the need for consent"¹⁷⁰. Furthermore, advanced directives might safeguard the autonomy of data subjects in longitudinal HRBDR studies¹⁷¹. Moreover, it is to be noted that human decision making is complex and affected by context and cognitive biases, combining

emotion and reason. Health-related big data processing can promote individual autonomy by expanding knowledge and provide it to both health professionals and patients. However, health data misuse can facilitate misinformation, discrimination or stigmatization. "Lack of common platforms and cross-disciplinary languages to deal with increasing technical complexity are significant challenges". "Also, can technology, data and analytical models alone capture human vulnerability, suffering, fears, hopes and potential? Evidently not"¹⁷².

At any rate, it should be highlighted that "consent revocation is a necessary motivator for research participation"¹⁷³. Note, however, that "as current data is reused and shared with multiple research organizations over indefinite periods of time, it is complicating the issue of revocation significantly"¹⁷⁴. To that effect, it is argued that withdrawal only applies to future and additional usage of data and not to data involved in studies already underway¹⁷⁵.

3.7. Principle of justice - discrimination

Besides, it is reasonably argued that "the misuse of big data has demonstrably resulted in various forms of ethnic, gender and class discrimination"¹⁷⁶. Ferretti et al. state that "correlations arising from health-related big data analytics can be abused by various actors for unethical purposes such as discriminating against applicants to health insurance services or jobs based on health risk indicators"¹⁷⁷. Price et al. state that "A recent survey of clinical trial participants on the sharing of participant-level clinical trial data beyond genomic information found that "6.6% were "very concerned" and 14.9% were "somewhat concern" that "I could be discriminated against if the information was linked back to me"¹⁷⁸. More precisely, in the machine learning / AI context there may be the so-called "algorithm-

166 Ivanova, D. and Katsaounis, P. "Real-time dynamic tiered e-consent: a novel tool for patients' engagement and common ontology system for the management of medical data". *Innovations in Digital Health, Diagnostics, and Biomarkers*. 2021; 1:45-49.

167 Ibid.

168 McKeown, A., Mourby, M., Harrison, P. et al. *op.cit.*

169 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

170 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

171 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

172 Cordeiro, J.V. *op.cit.* 647897.

173 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

174 Ibid.

175 Ibid.

176 Ienca, M., Ferretti, A., Hurst, S. et al. *op.cit.*

177 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.*

178 Price, W.N. 2nd and Cohen, I.G. *op.cit.*

mic discrimination”¹⁷⁹. Algorithms can mirror or even amplify or exacerbate the health care inequalities and biases that already are existing across the global population¹⁸⁰. This can be addressed with making algorithms more inclusive by testing them in various context¹⁸¹.

Algorithms’ potential to make biased decisions is a subject of controversy. Many people, including physicians, are of the opinion that algorithmic decisions are less biased than human decisions¹⁸². This assumption is based on the belief that algorithms, “unlike humans, decontextualize decision-making” and “are more likely than humans to treat everyone equally”. However, this perception may lead “people to endorse stereotypical beliefs that fuel discrimination and reduces their willingness to act against potentially discriminatory outcomes”¹⁸³. We should bear in mind that algorithmic decision-making may yield biased decisions or discrimination if the underlying data used by the algorithm are not representative. Furthermore, AI can lead to discrimination, e.g. racial biased decisions, inadvertently or by using stigmatizing terms such as those flagging members of a race, e.g. blacks, or other social groups such as LGBTIQ+ communities, as more likely to do something negative, for instance to offend. Furthermore, it is to be noted that “unlike traditional software, the machine learning algorithms and the models AI uses to make decisions are highly complex and, in some cases, opaque....This lack of transparency is referred to as the ‘black box’ of AI”¹⁸⁴.

Data science should serve the public interest and promote social justice. To this end, data science should reduce access asymmetries, discrimination and stigmatization, and improve the availability and quality of medical services. Therefore, inclusive and democratic deliberation is required. “Normative orders such as law and eth-

ics should act as beneficial limit-setters and promoters of just, creative and innovative realities”¹⁸⁵.

Furthermore, HRBDR raises justice-related ethical concerns. “In international collaboration involving low- and middle-income (LMI) countries, big data raises questions of justice in terms of allocation of resources in the health sector”¹⁸⁶. Indeed, while in the LMI countries health resources are scarce, data collection and use may set limits to other health services, which however, may be a higher priority. On the other hand, the use of data can measure deficiencies and injustices in the healthcare system, thus increasing visibility of inequity in healthcare on a global scale¹⁸⁷.

It is stated that if “participation is free, open, voluntary, and nondiscriminative of any resident” of a country, this may ensure “a fair distribution of risks and benefits to all research participants and the public at large”, and might address many concerns about justice including the fair distribution of burdens and benefits¹⁸⁸.

3.8. Liability

Given that artificial intelligence systems become more and more autonomous and given the lack of transparency in artificial intelligence systems, it is very difficult to foresee undesirable consequences and assign liability to that system¹⁸⁹. However, as medicine is much more than application of complex algorithms, human physicians should be thought of as being part of artificial intelligence teams¹⁹⁰. To that effect, it is argued that there may be physicians with a key role in an artificially intelligent team that might be regarded as responsible for an error¹⁹¹. Schiff and Borenstein state that “physicians should be responsible for acquiring basic understanding of the AI devices they use and the types and likelihood

179 Metcalf, J. and Crawford, K. “Where are human subjects in Big Data research? The emerging ethics divide”. *Big Data & Society*. 2016;3(1):2053951716650211.

180 Canales, C., Lee, C. and Cannesson, M. *op.cit.*

181 Ibid.

182 Ibid.; Bonezzi, A. and Ostinelli, M. “Can algorithms legitimize discrimination?” *Journal of Experimental Psychology: Applied*. 2021 Jun;27(2):447-459.

183 Bonezzi, A. and Ostinelli, M. *op.cit.*

184 Lysaght, T., Lim, H.Y., Xafis, V., Ngiam, K.Y. “AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research”. *Asian Bioethics Review*. 2019 Sep 12;11(3):299-314.

185 Cordeiro, J.V. *op.cit.* 647897.

186 Manrique de Lara, A. and Peláez-Ballestas, I. “Big data and data processing in rheumatology: bioethical perspectives”. *Clinical Rheumatology*. 2020 Apr;39(4):1007-1014.

187 Ibid.

188 Geneviève, L.D. et al. *op.cit.*; Althobaiti, K. *op.cit.*

189 Bathaee, Y. “The artificial intelligence black box and the failure of intent and causation”. *Harvard Journal of Law and Technology*. 2018; 31:890-934.

190 Canales, C., Lee, C. and Cannesson, M. *op.cit.*

191 Schiff, D. and Borenstein, J. *op.cit.*

of errors across subgroups, insofar as this information is available. Physicians should also be responsible for communicating relevant information to patients and health care teams and for adhering to use standards provided by device companies¹⁹². At any rate, currently the responsibility rests with the human physicians who must decide to follow or not a “decision” made from an algorithm¹⁹³. Access decisions for potentially sensitive data, especially those involving complex societal values should not be made based on algorithms. To ensuring “reasoned and responsible” decisions, a human-mediated decision-making process is then required, with study participants being central to this decision-making, especially when it comes to longitudinal studies¹⁹⁴. Ongoing consent dialogue is required¹⁹⁵. However, we need to bear in mind the physicians’ propensity to favor suggestions from automated decision-making systems, namely, the risk of “automation bias”. That is to say, they tend to trust algorithms or machines more than themselves¹⁹⁶.

3.9. Blurred and porous lines of distinction between: research and practice, facts and values, public and individual healthcare

HRBDR is increasingly blurring the lines of distinction between clinical care and research¹⁹⁷. Metcalf and Crawford are correct in stating that “the iterative nature of algorithmically driven data analytics blurs the line between research and practice”¹⁹⁸. Dankar et al. arguably state that “in this highly dynamic context, the lines between clinical and research practice are fading. Research in translational genomics is increasingly calling for the return of individual results back to participants and their physicians, thus challenging the traditional approach to consent even further”¹⁹⁹. Furthermore, the authors state, “Multiple studies have deduced that the

majority of participants would like to learn more about their genetic results (than is recommended). Further, they would like to decide what results to be returned. This complicates the ‘information’... ”²⁰⁰.

The sharp line of distinction between scientific facts and human values has become highly porous in the context of HRBDR²⁰¹. HRBDR involves the use of AI. In the context of AI ethics, it is argued that “it is necessary to build tangible bridges between abstract values and technical implementations, as long as these bridges can be reasonably constructed”²⁰². Furthermore, “AI ethics, conversely, turns away from the description of purely technological phenomena in order to focus more strongly on genuinely social and personality-related aspects”²⁰³. In that regard, it is to be noted that in the context of HRBDR “there is a need to move beyond “what the big data represents” to the “why it is so””²⁰⁴.

Althobaiti arguably states that “the provision of informed consent is one of the major challenges associated with public health surveillance. To strike a balance between protecting individual rights and pursuing societal welfare, consent from participants is essential”²⁰⁵. Furthermore, the author states that “there are several conditions under which conducting surveillance without obtaining explicit patient consent is ethically justifiable. However, this creates an ethical dilemma, especially at the intersection of public and clinical health ethics principles”²⁰⁶. While the clinical health ethics focuses on the four fundamental principles of ethics, namely, the principles of autonomy, beneficence, nonmaleficence, and justice, public health ethics focuses on population and community, namely, on public values and common good²⁰⁷.

192 Ibid.

193 Canales, C., Lee, C. and Cannesson, M. *op.cit.*

194 Murtagh M.J. et al. *op.cit.*

195 Ibid.

196 Canales, C., Lee, C. and Cannesson, M. *op.cit.*

197 Beier, K., Schweda, M. and Schicktanz, S. “Taking patient involvement seriously: a critical ethical analysis of participatory approaches in data-intensive medical research”. *BMC Medical Informatics and Decision Making*. 2019 Apr 25;19(1):90.

198 Metcalf, J. and Crawford, K. *op.cit.*

199 Dankar, F.K., Gergely, M. and Dankar, S.K. *op.cit.*

200 Ibid.

201 Dereli, T. et al. *op.cit.*

202 Hagendorff, T. [On line publication] “The Ethics of AI Ethics—An Evaluation of Guidelines”. ArXiv:1903.03425 [Cs, Stat]. 2019. Retrieved from <http://arxiv.org/abs/1903.03425> [Consulted: 06/11/2022]

203 Ibid.

204 Kar, A.K. and Dwivedi, Y.K. “Theory building with big data-driven research – Moving away from the “What” towards the “Why””. *International Journal of Information Management* 2020; 54(102205): 1-10

205 Althobaiti, K. *op.cit.*

206 Ibid.

207 Ibid.

3.10. The notion of “public interest”

Allen et al. arguably state that “there’s not really a good set of guidelines around what constitutes public interest”²⁰⁸. The currently unclear determination of the term “public interest” grows the financialization of HRBDR and public distrust in it. “In a political-economic environment where a clear determination of ‘public interest’ is lacking, growing commercial interests are increasingly structured around proprietorial control of such data assets and the knowledge produced for the financialisation of biomedicine”²⁰⁹. “Consequently, public distrust and resistance to national big data programmes in healthcare grows”²¹⁰. Allen et al. conducted a qualitative research and found that while “some data custodians believed that it was their role to determine the public good of research projects”, “others regarded it as the role of the research ethics committee”²¹¹. Ultimately and most importantly, the authors put it best in saying “we do not see data custodians as having sole responsibility for establishing and maintaining social licence and in particular sole responsibility for determining the public interest. Even though some legislation gives the decision-making responsibility about public interest to the data custodian, social licence is more likely to be maintained if advice is taken from an ethics committee”²¹².

3.11. Addressing concerns relating to the use of AI

To tackle concerns relating to the ethics of AI involved in HRBDR, Baird and Schuller introduce four key aspects/considerations, which have an “inherent relation to data infrastructures”: “auditing, benchmarking, confidence and trust and explainability and interpretability”²¹³. The authors argue that considering these aspects along with “improved interdisciplinary discussion” can

mitigate “data-based AI ethical concerns”²¹⁴. To make clear the aspects *interpretability* and *explainability* the authors state that “a distinction can be made, *interpretability* being methods for better understanding a machine learning architecture or data source (i.e., the *how*), and *explainability* being methods for understanding *why* particular decision were made”²¹⁵. Furthermore, the authors state that “communicating AI to the general public may also see an improvement, which in turn will help to build trust”²¹⁶. To that effect, the authors emphasize the role of various experts with various backgrounds. They state that “seamless **integration** of AI is necessary for its success and adoption by the general public. Aspects including cultural and environmental impact need to be considered, and various experts should provide knowledge on the target area”. Furthermore, the authors observed “that knowledge of **selection-bias** often requires contributions from experts with non-technical backgrounds”²¹⁷.

3.12. Ethical governance of HRBDR

In addition, ethical governance of HRBDR is problematic. Importantly, formal regulation lacks, not suffices or is not applicable in the particular ecosystem HRBDR. Legal framework, guidelines and ethical oversight practices differ from country to country²¹⁸. To reduce uncertainty, various stakeholders, such as scientific community, have developed nonbinding best-practice guidelines²¹⁹. These are “soft-law” regulations. Soft regulations developed “in accordance with the requirements of an advanced democratic society” has been an appropriate regulatory framework for HRBDR²²⁰. Ferretti et al. “conducted a scoping review of soft-law documents and guidelines with the aim of assessing ongoing normative

208 Allen, J., Adams, C. and Flack, F. “The role of data custodians in establishing and maintaining social licence for health research”. *Bioethics*. 2019 May;33(4):502-510

209 Vezyridis, P. and Timmons, S. “Resisting big data exploitations in public healthcare: free riding or distributive justice?” *Sociology of Health & Illness*. 2019 Nov;41(8):1585-1599.

210 Ibid.

211 Allen, J., Adams, C. and Flack, F. *op.cit.*

212 Ibid.

213 Baird, A. and Schuller, B. *op.cit.*

214 Ibid.

215 Ibid.

216 Ibid.

217 Ibid.

218 Vayena, E. “Value from health data: European opportunity to catalyse progress in digital health”. *The Lancet* 2021; 397 (10275): 652–653; Ferretti, A., Ienca, M., Velarde, M.R. et al. *op.cit.*

219 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* p.19.

220 Gruson, D., Helleputte, T., Rousseau, P. and Gruson, D. “Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation”. *Clinical Biochemistry*. 2019 Jul;69:1-7

efforts that are proliferating” in the domain of HRBDR²²¹. The authors highlight that due to “fragmentation and heterogeneity of the current landscape of guidance documents” the uncertainty about the ethics review of HRBD studies remains height despite the recommendations provided by the soft-law documents²²². In that regard, it should be highlighted that HRBDR is a “unique” context where data come from “multiple sources, which are differently protected by the law”²²³. The soft-law documents recommended making efforts “to improve the ethics review process and formalize a coherent ethical review framework for the evaluation of big data projects”²²⁴. Moreover, a literature review conducted by Barbui revealed the “fragmentation and heterogeneity” of the current guidance landscape. The review revealed that the data sharing policies of top biomedical journals “remain largely heterogenous”²²⁵.

Furthermore, as ERCs’ traditional mandate is “deeply rooted” in the “pre-digital era of biomedical research”, oversight mechanisms such as IRBs or ERCs in their current form are ill-suited to provide exclusive ethical oversight on HRBDR projects²²⁶. “ERCs might still be faced with uncertainty when reviewing health-related big data studies”²²⁷. To that effect, *ad hoc* criteria for ethically evaluating HRBDR projects are urgently proposed²²⁸.

Moreover, the trust should be reassured, given that it is “an essential moral mechanism intrinsic to any ethical governance” of HRBDR²²⁹. “Ethically robust guidelines for the collection and sharing of personal health data would facilitate big data research while maintaining public trust and protecting data subjects”²³⁰. “In AI data...having *confidence* in the data results in deep-

er *trust*”²³¹. “In this context, *trust* is a qualitative term, and...the term *confidence* typically refers to a quantifiable measure to base *trust* on”²³². Already in 2017 Salerno et al. state that “collecting and analyzing data and information should be performed with the purpose of improving population health”, and that “communities should trust the scientists who collect, analyze, and offer advice based on it”. Researchers “should work with communities to create new norms that include an expectation on the part of communities”. Besides, “protecting data from disclosure will require that we raise professional expectations of all parties who “touch” data”²³³. To that effect, the authors highlight the importance of “ensuring data security, preventing public disclosure of personal information and, generally, practicing responsible stewardship”²³⁴. Public trust to HRBDR is essential for data contribution. Interestingly, Middleton et al. state, “Willingness to donate one’s DNA and health data for research is relatively low, and trust in the process of data’s being shared with multiple users (e.g., doctors, researchers, governments) is also low. Participants were most willing to donate DNA or health information for research when the recipient was specified as a medical doctor and least willing to donate when the recipient was a for-profit researcher. Those who were familiar with genetics and who were trusting of the users asking for data were more likely to be willing to donate”²³⁵. In that regard, scholars provide some recommendations for future research governance.

3.13. Expanding ERCs’ purview

Ferretti et al. “conducted a scoping review of soft-law documents and guidelines with the aim of assessing ongoing normative efforts that are proliferating” in the domain of HRBDR. Documents recommended that the ERCs’ purview should be expanded to review additional

221 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* p.17.

222 Ibid.

223 Bassan, S. and Harel, O. “The Ethics in Synthetics: Statistics in the Service of Ethics and Law in Health-Related Research in Big Data from Multiple Sources”. *Journal of Law and Health*. 2018;31(1):87-117

224 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* p.25.

225 Barbui C. *op.cit.*

226 Ienca, M., Ferretti, A., Hurst, S. et al. *op.cit.*; Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

227 Ienca, M., Ferretti, A., Hurst, S. et al. *op.cit.*

228 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

229 Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W. et al. *op.cit.*

230 Ienca, M., Vayena, E. and Blasimme, A. *op.cit.*

231 Baird, A. and Schuller, B. *op.cit.*

232 Baird, A. and Schuller, B. *op.cit.*

233 Salerno, J. et al. *op.cit.*

234 Ibid.

235 Middleton, A., Milne, R., Almarri, M.A., Answer, S., Atutornu, J., Baranova, E.E. et al. “Global Public Perceptions of Genomic Data Sharing: What Shapes the Willingness to Donate DNA and Health Data?” *The American Journal of Human Genetics*. 2020 Oct 1;107(4):743-752

kinds of studies in addition to those involving human subjects, such as studies that involve de-identified or publicly available data²³⁶.

3.14. Expanding ERCs members' expertise

Furthermore, Ferretti et al. state that ERCs "are often composed of stakeholders (such as lawyers, physicians, nurses, and laypeople) who rarely have received formal training in computer or data science", ERC members' expertise should be further expanded "to account for the computational and ethical complexity of big data studies"²³⁷. This may become even more important as HRBDR gets vastly more complicated, e.g. involving semi-supervised AI or hybrid intelligence, namely, a combination of human and machine intellect boosting each other. To that effect, various stakeholders, such as scientific community, have developed educational activities such as providing research ethics training to data scientists²³⁸. In this respect it is to be noted that from a qualitative research conducted by Favaretto et al. emerged "absence of appropriate expertise among members of the boards, and lack of harmonized evaluation criteria between committees"²³⁹. Favaretto et al. "argue for updating the expertise of board members and the institution of a consultancy model between researchers and ECs"²⁴⁰.

3.15. Creating novel oversight bodies

Complementary governance mechanisms such as data boards, data security committees, corporations or allied bodies may be used to "expand the bandwidth and sensitivity of ethical oversight" with the aim of securing a better oversight and governance of the HRBDR²⁴¹. "The creation of novel oversight bodies such as data boards was also proposed as an adaptive governance solution to

the big data ethics conundrum"²⁴². "Possible strategies include engaging research subjects and communities in the decision-making process or promoting a co-governance system"²⁴³. Rauter et al. emphasize that patients should be encouraged to participate in HRBDR projects at all steps of the process²⁴⁴. Muller et al. reasonably argue that "actively integrating stakeholders into all the stages of data governance" can facilitate "co-creating" what is considered as trustworthy"²⁴⁵. Importantly, "building capacity and expanding competencies" of oversight bodies can "improve the credibility of IRBs from the point of view of researchers and will increase researchers' willingness to undergo ethics review"²⁴⁶.

Terms such as 'involvement' or 'participation' may receive various interpretations, ranging from simple opt-in consent in participation to co-decision making²⁴⁷. Indeed, "significant differences exist regarding the participatory dimension and degree of influence"²⁴⁸. Beier et al. examined the crucial role of participatory concepts in the context of HBDR and identified three roles of patients/subjects in the HBDR context: the role of provider "of biomaterials and data", the role of administrator "of their own research participation" and the role of "(co-) principal investigator"²⁴⁹. Murtagh et al. is correct in stating, "in order to make responsible decisions, participant-centredness takes three forms: (1) respecting study participant expectations, (2) involving study participants in decision-making roles and (3) communicating the results of access decisions to participants (and others) in a format that is clear and accessible, for example, in plain language summaries"²⁵⁰. Interestingly, in some countries, participants have become active stakeholders. Rauter et al. report that German Patient Organizations have become increasingly active stakeholders²⁵¹ Beier et al.

236 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* pp. 18.,25.

237 Ibid.

238 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*

239 Favaretto, M., De Clercq, E., Briel, M. and Elger, B.S. "Working Through Ethics Review of Big Data Research Projects: An Investigation into the Experiences of Swiss and American Researchers". *Journal of Empirical Research on Human Research Ethics*. 2020 Oct;15(4):339-354

240 Ibid.

241 Ienca, M., Ferretti, A., Hurst, S. et al. *op.cit.*

242 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* p.25.

243 Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*

244 Rauter, C.M., Wöhlke, S. and Schicktanz, S. "My Data, My Choice? - German Patient Organizations' Attitudes towards Big Data-Driven Approaches in Personalized Medicine. An Empirical-Ethical Study". *Journal of Medical Systems*. 2021 Feb 22;45(4):43

245 Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W. et al. *op.cit.*

246 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.* p.22.

247 Rauter, C.M., Wöhlke, S. and Schicktanz, S. *op.cit.*

248 Beier, K., Schweda, M. and Schicktanz, S. *op.cit.*

249 Ibid.

250 Murtagh M.J. et al. *op.cit.*

251 Rauter, C.M., Wöhlke, S. and Schicktanz, S. *op.cit.*

state that “an illustrative example of participants’ role as co-decision-makers is the recent Swiss initiative *MI-DATA*”²⁵². Furthermore, it is noticeable that while Beier et al. highlight the “current boom of appeals to ‘public engagement’” in the HRBDR context, they observed that “references to patient or citizen participation and involvement in HBDR...are often vague, merely rhetorical, or oblivious to normative implications”²⁵³. Last, it is to be noted that while more patient involvement is ethically required, it does not make HRBDR projects “ethically acceptable”²⁵⁴.

To be effective, the integrated (hybrid) mechanisms of oversight, merging traditional with more innovative research review models, would review all stages of research and actively integrate all relevant stakeholders into all stages of data governance and ethical evaluation of HRBDR projects²⁵⁵. In that regard, Dankar et al. reasonably argue that “the oversight mechanism should maintain ethical controls throughout the continuum of data activities (data collection and data use)”²⁵⁶.

Ultimately, there are some requirements for participants’ involvement in research governance to be considered effective. More precisely, the authors place considerable emphasis on “socio-empirical research exploring lay persons’ moral views and attitudes regarding ethical issues”²⁵⁷. Note, however, that “research initiatives need to be transparent about the aim, extent and benefits of participation in order to avoid ‘participatory misconception’ resulting in misleading expectations of participants regarding their power”²⁵⁸. “Involvement of study participants in decision-making roles requires active work to ensure it is meaningful”²⁵⁹. For instance, a learning process that deepens their knowledge and understanding of governance issues is required²⁶⁰. To that effect, it is to be noted that “researchers often do not communi-

cate their work in ways which are clear or accessible to non-expert audiences”²⁶¹. Furthermore, participant-centeredness means considering them more than research *subjects*, and brings not only study participants closer to the research, but also research and researchers closer to participants²⁶². Deverka et al. conducted public deliberations in the U.S. recruiting deliberants from diverse geographic areas of the country. Deliberants sought transparent procedures for selecting community representatives to structure public participation in governance²⁶³.

Furthermore, it is to be noted that not only Research Ethics Committees, but also research teams and governance need to be interdisciplinary. The management and governance of the complex and sensitive HRBD sources, e.g. those including next-generation sequencing, and other ‘omics’ data, need to be interdisciplinary. A range of expertise from across different disciplines including research participants, with multiple subject positions, is required²⁶⁴. Robinson concludes that “the inclusion of a qualified, senior Health Information Manager in research teams and on institutional Human Research Ethics Committees would help to prevent potential problems”²⁶⁵.

3.16. *The mechanism of “social licence”*

Most importantly, the mechanism of “social licence” can serve as a guideline for ethical governance of HRBDR²⁶⁶. Social licence “describes whether a given data use is accepted by stakeholders”²⁶⁷. Furthermore, it is argued that “Social licence refers to the informal permissions granted to institutions...by members of the public to carry out a particular set of activities”²⁶⁸, which “go

261 Ibid.

262 Ibid.

263 Deverka, P.A. et al. *op.cit.*

264 Murtagh M.J. et al. *op.cit.*

265 Robinson, K. “A false promise of COVID-19 ‘big’ health data? Health data integrity and the ethics and realities of Australia’s health information management practice”. *Health Information Management Journal*. 2021 Jan-May;50(1-2):9-12.

266 Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W. et al. *op.cit.*

267 Ballantyne, A. and Stewart, C. Big Data and Public-Private Partnerships in Healthcare and Research: The Application of an Ethics Framework for Big Data in Health and Research. *Asian Bioethics Review*. 2019 Sep 30;11(3):315-326.

268 Shaw, J.A., Sethi, N. and Cassel, C.K. Social license for the use of big data in the COVID-19 era. *NPJ Digital Medicine*. 2020 Oct 2;3:128

252 Beier, K., Schweda, M. and Schicktanz, S. *op.cit.*

253 Ibid.

254 Rauter, C.M., Wöhlke, S. and Schicktanz, S. *op.cit.*

255 Ferretti, A., Ienca, M., Hurst, S. et al. *op.cit.*; Ferretti, A., Ienca, M., Sheehan, M. et al. *op.cit.*; Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W. et al. *op.cit.*

256 Dankar, F.K., Gergely, M., Dankar, S.K. *op.cit.*

257 Beier, K., Schweda, M. and Schicktanz, S. *op.cit.*

258 Ibid.

259 Murtagh M.J. et al. *op.cit.*

260 Ibid.

further than the requirements of formal regulation, towards voluntary adherence to social codes of trustworthy and responsible behaviour"²⁶⁹. Allen et al. state "The concept of social licence is used to describe a privilege granted to an occupation or profession to do things other members of society are not allowed to do and which may not be morally acceptable in the wider society"²⁷⁰. Data custodians rely on social licence in order to support deviation from the usual norms of confidentiality²⁷¹.

Muller et al. argue that social licence may secure an alignment between all the stakeholders involved in HRBDR projects, e.g. patients, non-medical stakeholders, medical stakeholders, or public²⁷².

The use of the mechanism of social licence is of great ethical importance. Muller et al. state that "social licence grants moral liberties to researchers based on society's trust in their activities. Those liberties in turn demand trustworthiness, coming with duties to act in ethically acceptable ways. This is defined as behaviour that addresses concerns about data privacy and risk, and that promotes public values such as transparency, reciprocity, inclusivity, and the common good"²⁷³. That is to say that social licence may serve as a mechanism contributing to address HRBDR-related ethical issues, both rooted in the pre-digital era of biomedical research and newly appeared. The ethical issues that are rooted in the pre-digital era of biomedical research are adapted to the new HPBDR contexts.

Creating and maintaining "social licence" is an issue of great importance in the HRBDR context. Research governance play a crucial role "in providing moral legitimacy and securing social licence for medical research"²⁷⁴. Allen et al. highlight the role of data custodians in establishing and maintaining social licence²⁷⁵. Interestingly, Carter, Laurie and Dixon-Woods conducted a qualitative

research and identified three conditions for establishing social licence – reciprocity, non-exploitation and the public good. More precisely, the authors state, "Necessary—and hopefully sufficient—conditions for social licence include: (i) *reciprocity*, which must begin with sound two-way communication, (ii) *non-exploitation*, which must exclude the spectre of disempowerment, and (iii) *service of the public good*, which need not exclude a wealth agenda so long as there is confidence that research governance and information governance systems can hold researchers, and others with custodial responsibility for medical information, to account"²⁷⁶.

In that connection, it is to be noted that Muller et al. state, "There are several merits to using the concept of social licence as a guideline for ethical governance. Firstly, it fits the novel scale of data-related risks; secondly, it focuses attention on trustworthiness; and finally, it offers co-creation as a way forward. Greater trust can be achieved in the governance of data-intensive health research by highlighting strategic dialogue with both patients contributing the data, and the public in general"²⁷⁷. Indeed, it is of great importance that social licence involves strategic dialogue. "Strategic dialogue with both patients contributing the data, and the public in general" can build more trust in the governance of HRBDR²⁷⁸. To that effect, "we must follow and contribute to the societal discussion of privacy norms"²⁷⁹. As mentioned above, Baird and Schuller argue that improved interdisciplinary discussion along with considering "*auditing, benchmarking, confidence and trust and explainability and interpretability*" can mitigate "data-based AI ethical concerns"²⁸⁰.

4. Conclusions

The development of machine learning and artificial intelligence technology contributed substantially to the growth of health-related big data research, which may

269 Ford E, Boyd A, Bowles JKF, Havard A, Aldridge RW, Curcin V et al. Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond. *Learning Health Systems*. 2019 Mar 25;3(3):e10191

270 Allen, J., Adams, C. and Flack, F. *op.cit.*

271 Ibid.

272 Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W. et al. *op.cit.*

273 Ibid.

274 Allen, J., Adams, C. and Flack, F. *op.cit.*

275 Ibid.

276 Carter, P., Laurie, G.T. and Dixon-Woods, M. "The social licence for research: why care.data ran into trouble". *Journal of Medical Ethics*. 2015 May;41(5):404-409.

277 Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W. et al. *op.cit.*

278 Ibid.

279 Mooney, S.J. and Pejaver, V. *op.cit.*

280 Baird, A. and Schuller, B. *op.cit.*

show promising preventive results in prevention, diagnosis and therapy, and is expected to improve the development of the so-called personalized medicine. The term “big data” has not yet been clearly defined. Volume, velocity and variety are mentioned among the most common characteristics for defining “big data”. Big data research constitutes a novel and particular ecosystem in the health care and a new context in the health care ethics. The already existing ethical principles and concepts need to be revisited in the new health-related big data research context. From publicly available data or data that are not health related may result, wittingly or unwittingly in the big data research process, health-related and/or sensitive personally identifiable information. Importantly, even when people do not know things about others, they make inferences about them. To that effect, various encryption algorithms have been proposed, among other things. Given that big data is usually stored in centralized infrastructures, decentralization using blockchain techniques is strongly recommended to increase transparency and democratization. Traditional research ethics notions such as privacy/privacy protection, ownership, minimal risk, participant/data subject, confidentiality, fair subject selection and informed consent are to be reconsidered. Broad consent has been widely discussed as alternative to traditional specific informed consent. Nevertheless, forms of dynamic consent are reasonably thought as being a more satisfying solution. Importantly, health-related big data research not only exacerbates traditional research ethics issues, but also creates new ethical issues such those related to trust / trustworthiness and public values such as reciprocity, transparency, inclusivity and common good. In this point of view, ethical governance of health-related big data research is problematic. Formal regulation lacks, not suffices or is not applicable in the particular ecosystem HRBDR. Legal framework, guidelines and ethical oversight practices differ from country to country. In the available literature it is highlighted the role of the so-called “soft-law” in ethical governance of health-related big data research. Ethical review committees in their current form are ill-suited to provide exclusive ethical oversight on health-related big data projects. Expanding ethical

review committees’ purview and members’ expertise, as well as creating novel oversight bodies by promoting a co-governance system including public and all the stakeholders involved and reviewing all the stages of research are strongly recommended. Co-governance can improve trustworthiness among other things. Ensuring trust and confidentiality, and especially the trust of potential data subjects in science, is highlighted in the health-related big data research ethics. Most importantly, the mechanism of “social licence” can serve as a guideline for ethical governance of health-related big data research. Social licence refers to the informal permissions granted to researchers by society to carry out a particular set of research. Lines of distinction between research and practice, facts and values, public and individual health-care have become blurred and porous in the context of health-related big data research. In the HRBDR context there is a lot of ignorance and uncertainty. High-stakes decisions are often made under uncertainty. Algorithmic decision-making may yield biased decisions or discrimination, for instance if the underlying data used by the algorithm are not representative, inadvertently or by using stigmatizing terms. In that regard, it should be highlighted that the machine learning algorithms and the artificial intelligence models that are used in decision making are highly complex and, in some cases, opaque. To address many concerns associated with justice, a fair distribution of risks and benefits to all research participants and the public at large is required, especially when it comes to big data-driven research involving low- and middle-income countries. Interdisciplinarity is highlighted in the health-related big data research context. Improved interdisciplinary dialogue along with considering aspects like auditing, benchmarking, confidence and trust and explainability and interpretability may address concerns related to the use of AI in health-related big data research. Research ethics shifts towards a population-based model of ethics.

Availability of data and materials

The authors declare that the data analyzed during the current study and supporting the findings of this

study are available from the corresponding author on reasonable request.

Abbreviations

AI = Artificial Intelligence
 ERC = Ethics Review Committee
 REC= Research Ethics Committee
 HRBDR = Health-related Big Data Research
 LM = Learning Machine
 PM = Personalized Medicine
 REC = Research Ethics Committee
 SL = Social Licence
 SM = Social Media
 LMI countries = low- and middle-income countries

Funding source

There are no sources of funding to be declared.

Conflicts of interest

The authors declare that they have no competing interests.

References

- Allen, J., Adams, C. & Flack, F. "The role of data custodians in establishing and maintaining social licence for health research". *Bioethics*. 2019 May;33(4):502-510. doi: 10.1111/bioe.12549. Epub 2019 Jan 18. PMID: 30657596.
- Althobaiti, K. "Surveillance in Next-Generation Personalized Healthcare: Science and Ethics of Data Analytics in Healthcare". *The New Bioethics*. 2021 Dec;27(4):295-319. doi: 10.1080/20502877.2021.1993055. Epub 2021 Oct 31. PMID: 34720071.
- Andriukaitis, V. "Personalised Medicine in Europe". *Biomedicine Hub*. 2017 Nov 21;2(Suppl 1):1-2. doi: 10.1159/000479406. PMID: 31988922; PMCID: PMC6945890.
- Baird, A. & Schuller, B. "Considerations for a More Ethical Approach to Data in AI: On Data Representation and Infrastructure". *Frontiers in Big Data*. 2020 Sep 2;3:25. doi: 10.3389/fdata.2020.00025. PMID: 33693398; PMCID: PMC7931893.
- Ballantyne, A. Adjusting the focus: A public health ethics approach to data research. *Bioethics*. 2019 Mar;33(3):357-366. doi: 10.1111/bioe.12551. Epub 2019 Jan 22. PMID: 30667080.
- Ballantyne, A. & Stewart, C. "Big Data and Public-Private Partnerships in Healthcare and Research: The Application of an Ethics Framework for Big Data in Health and Research". *Asian Bioethics Review*. 2019 Sep 30;11(3):315-326. doi: 10.1007/s41649-019-00100-7. PMID: 33717319; PMCID: PMC7747238.
- Barbui, C. Sharing all types of clinical data and harmonizing journal standards. *BMC Medicine*. 2016 Apr 3;14:63. doi: 10.1186/s12916-016-0612-8. PMID: 27038634; PMCID: PMC4818917.
- Bassan, S. & Harel, O. "The Ethics in Synthetics: Statistics in the Service of Ethics and Law in Health-Related Research in Big Data from Multiple Sources". *Journal of Law and Health*. 2018;31(1):87-117. PMID: 30889335.
- Bathae, Y. "The artificial intelligence black box and the failure of intent and causation". *Harvard Journal of Law and Technology*. 2018; 31:890-934.
- Beier, K., Schweda, M. and Schicktanz, S. "Taking patient involvement seriously: a critical ethical analysis of participatory approaches in data-intensive medical research". *BMC Medical Informatics and Decision Making*. 2019 Apr 25;19(1):90. doi: 10.1186/s12911-019-0799-7. PMID: 31023321; PMCID: PMC6482526.
- Belani, S., Tiarks, G.C., Mookerjee, N. & Rajput, V. "'I Agree to Disagree': Comparative Ethical and Legal Analysis of Big Data and Genomics for Privacy, Consent, and Ownership". *Cureus*. 2021 Oct 13;13(10):e18736. doi: 10.7759/cureus.18736. PMID: 34796049; PMCID: PMC8589338
- Bonezzi, A. & Ostinelli, M. "Can algorithms legitimize discrimination?" *Journal of Experimental Psychology: Applied*. 2021 Jun;27(2):447-459. doi: 10.1037/xap0000294. Epub 2021 Mar 22. PMID: 33749300.
- Budin-Ljøsne, I., Teare, H.J., Kaye, J., Beck, S., Bentzen, H.B., Caenazzo, L. Collett, C., D'Abramo, F., Felzmann, H., Finlay, T., Javaid, M.K., Jones, E., Katić,

- V., Simpson, A. & Mascalzoni, D. "Dynamic Consent: a potential solution to some of the challenges of modern biomedical research". *BMC Medical Ethics*. 2017 Jan 25;18(1):4. doi: 10.1186/s12910-016-0162-9. PMID: 28122615; PMCID: PMC5264333.
- Canales, C., Lee, C. & Cannesson, M. "Science Without Conscience Is but the Ruin of the Soul: The Ethics of Big Data and Artificial Intelligence in Perioperative Medicine". *Anesthesia & Analgesia*. 2020 May;130(5):1234-1243. doi: 10.1213/ANE.0000000000004728. PMID: 32287130; PMCID: PMC7519874.
- Car, J., Sheikh, A., Wicks, P. & Williams, M.S. "Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom". *BMC Medicine*. 2019 Jul 17;17(1):143. doi: 10.1186/s12916-019-1382-x. PMID: 31311603; PMCID: PMC6636050.
- Carter, P., Laurie, G.T. & Dixon-Woods, M. "The social licence for research: why care.data ran into trouble". *Journal of Medical Ethics*. 2015 May;41(5):404-409. doi: 10.1136/medethics-2014-102374. Epub 2015 Jan 23. PMID: 25617016; PMCID: PMC4431337.
- Cirillo, D. & Valencia, A. "Big data analytics for personalized medicine". *Current Opinion in Biotechnology*. 2019 Aug;58:161-167. doi: 10.1016/j.copbio.2019.03.004. Epub 2019 Apr 6. PMID: 30965188.
- Cordeiro, J.V. "Digital Technologies and Data Science as Health Enablers: An Outline of Appealing Promises and Compelling Ethical, Legal, and Social Challenges". *Frontiers in Medicine (Lausanne)*. 2021 Jul 8;8:647897. doi: 10.3389/fmed.2021.647897. PMID: 34307394; PMCID: PMC8295525.
- Dankar, F.K., Gergely, M. & Dankar, S.K. "Informed Consent in Biomedical Research". *Computational and Structural Biotechnology Journal*. 2019 Mar 25;17:463-474. doi: 10.1016/j.csbj.2019.03.010. PMID: 31007872; PMCID: PMC6458444.
- Dennett, D. Lunch with the FT: Daniel Dennett. Philosopher Daniel Dennett on AI, robots and religion. Interview by J. Thornhill. *Financial Times Weekend*, 4-5 March 2017, p. 3.
- Dereli, T., Coşkun, Y., Kolker, E., Güner, O., Ağırbaşı, M. & Ozdemir, V. "Big data and ethics review for health systems research in LMICs: understanding risk, uncertainty and ignorance -- and catching the black swans?" *American Journal of Bioethics*. 2014;14(2):48-50. doi: 10.1080/15265161.2013.868955. PMID: 24521341.
- Deverka, P.A., Gilmore, D., Richmond, J., Smith, Z., Mangrum, R., Koenig, B.A., Cook-Deegan, R., Villanueva, A.G., Majumder, M.A. & McGuire, A.L. "Hopeful and Concerned: Public Input on Building a Trustworthy Medical Information Commons". *Journal of Law, Medicine & Ethics*. 2019 Mar;47(1):70-87. doi: 10.1177/1073110519840486. PMID: 30994071; PMCID: PMC6730638.
- Duardo-Sánchez, A. & De Miguel Beriain, I. "Personalized Medicine and Medicinal Chemistry: Toward a Legal Framework in the European Union". *Current Topics in Medicinal Chemistry*. 2018;18(25):2165-2173. doi: 10.2174/1568026619666181120144213. PMID: 30465506.
- Favaretto, M., De Clercq, E., Briel, M. & Elger, B.S. "Working Through Ethics Review of Big Data Research Projects: An Investigation into the Experiences of Swiss and American Researchers". *Journal of Empirical Research on Human Research Ethics*. 2020 Oct;15(4):339-354. doi: 10.1177/1556264620935223. Epub 2020 Jun 19. PMID: 32552544.
- Favaretto, M., Shaw, D., De Clercq, E., Joda, T. & Elger, B.S. "Big Data and Digitalization in Dentistry: A Systematic Review of the Ethical Issues". *International Journal of Environmental Research and Public Health*. 2020 Apr 6;17(7):2495. doi: 10.3390/ijerph17072495. PMID: 32268509; PMCID: PMC7177351.
- Ferretti, A., Ienca, M., Hurst, S. & Vayena, E. "Big Data, Biomedical Research, and Ethics Review: New Challenges for IRBs". *Ethics & Human Research*. 2020 Sep;42(5):17-28. doi: 10.1002/eahr.500065. Erratum in: *Ethics Hum Res*. 2020 Nov;42(6):20. PMID: 32937036; PMCID: PMC7814666.
- Ferretti, A., Ienca, M., Sheehan, M., Blasimme, A., Dove, E.S., Farsides, B., Friesen, P., Kahn, J., Karlen, W., Kleist, P., Liao, S.M., Nebeker, C., Samuel, G., Shabani, M., Rivas Velarde, M. & Vayena, E. "Ethics review of big data research: What should stay and what should be reformed?" *BMC Medical Ethics*. 2021 Apr

- 30;22(1):51. doi: 10.1186/s12910-021-00616-4. PMID: 33931049; PMCID: PMC8085804.
- Ferretti, A., Ienca, M., Velarde, M.R., Hurst, S. & Vayena E. "The Challenges of Big Data for Research Ethics Committees: A Qualitative Swiss Study". *Journal of Empirical Research on Human Research Ethics*. 2022 Feb-Apr;17(1-2):129-143. doi: 10.1177/15562646211053538. Epub 2021 Nov 15. PMID: 34779661; PMCID: PMC8721531.
- Fisher, C.B. & Layman, D.M. "Genomics, Big Data, and Broad Consent: a New Ethics Frontier for Prevention Science". *Prevention Science*. 2018 Oct;19(7):871-879. doi: 10.1007/s11121-018-0944-z. PMID: 30145751; PMCID: PMC6182378.
- Ford, E., Boyd, A., Bowles, J.K.F., Havard, A., Aldridge, R.W., Curcin, V., Greiver, M., Harron, K., Katikireddi, V., Rodgers, S.E. & Sperrin, M. "Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond". *Learning Health Systems*. 2019 Mar 25;3(3):e10191. doi: 10.1002/lrh2.10191. PMID: 31317072; PMCID: PMC6628981.
- Friesen, P., Douglas-Jones, R., Marks, M., Pierce, R., Fletcher, K., Mishra, A., Lorimer, J., Véliz, C., Hallowell, N., Graham, M., Chan, M.S., Davies, H. & Sallamuddin, T. "Governing AI-Driven Health Research: Are IRBs Up to the Task?" *Ethics & Human Research*. 2021 Mar;43(2):35-42. doi: 10.1002/eahr.500085. PMID: 33683015.
- Geneviève, L.D., Martani, A., Wangmo, T., Paolotti, D., Koppeschaar, C., Kjelsø, C. Guerrisi, C., Hirsch, M., Woolley-Meza, O., Lukowicz, P., Flahault, A. & Elger, B.S. "Participatory Disease Surveillance Systems: Ethical Framework". *Journal of Medical Internet Research*. 2019 May 23;21(5):e12273. doi: 10.2196/12273. PMID: 31124466; PMCID: PMC6660191.
- Gruson, D., Helleputte, T., Rousseau, P. & Gruson, D. "Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation". *Clinical Biochemistry*. 2019 Jul;69:1-7. doi: 10.1016/j.clinbiochem.2019.04.013. Epub 2019 Apr 22. PMID: 31022391.
- Gürsoy, G., Harmanci, A., Tang, H., Ayday, E. & Brenner, S.E. "When Biology Gets Personal: Hidden Challenges of Privacy and Ethics in Biological Big Data". *Pacific Symposium on Biocomputing*. 2019; 24:386-390. PMID: 30864339; PMCID: PMC7577606.
- Hagendorff, T. [On line publication] "The Ethics of AI Ethics—An Evaluation of Guidelines". ArXiv:1903.03425 [Cs, Stat]. 2019. Retrieved from <http://arxiv.org/abs/1903.03425> [Consulted: 06/11/2022]
- Hibbin, R.A., Samuel, G. & Derrick, G.E. "From "a Fair Game" to "a Form of Covert Research": Research Ethics Committee Members' Differing Notions of Consent and Potential Risk to Participants Within Social Media Research". *Journal of Empirical Research on Human Research Ethics*. 2018 Apr;13(2):149-159. doi: 10.1177/1556264617751510. Epub 2018 Jan 19. PMID: 29350083; PMCID: PMC5873513.
- Hummel, P. & Braun, M. "Just data? Solidarity and justice in data-driven medicine". *Life Science Society and Policy*. 2020 Aug 25;16(1):8. doi: 10.1186/s40504-020-00101-7. PMID: 32839878; PMCID: PMC7445015.
- Ienca, M., Ferretti, A., Hurst, S., Puhon, M., Lovis, C. & Vayena E. "Considerations for ethics review of big data health research: A scoping review". *PLoS One*. 2018 Oct 11;13(10):e0204937. doi: 10.1371/journal.pone.0204937. PMID: 30308031; PMCID: PMC6181558.
- Ienca, M. & Ignatiadis, K. "Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges". *AJOB Neuroscience (American Journal of Bioethics Neuroscience)*. 2020 Apr-Jun;11(2):77-87. doi: 10.1080/21507740.2020.1740352. PMID: 32228387.
- Ienca, M., Vayena, E. & Blasimme, A. "Big Data and Dementia: Charting the Route Ahead for Research, Ethics, and Policy". *Frontiers in Medicine (Lausanne)*. 2018 Feb 6;5:13. doi: 10.3389/fmed.2018.00013. PMID: 29468161; PMCID: PMC5808247.
- Ivanova, D. & Katsaounis, P. "Real-time dynamic tiered e-consent: a novel tool for patients' engagement and common ontology system for the management of medical data". *Innovations in Digital Health, Diagnostics, and Biomarkers*. 2021; 1:45-49. DOI: 10.36401/IDDB-21-01.

- Kar, A.K. & Dwivedi, Y.K. "Theory building with big data-driven research – Moving away from the "What" towards the "Why"". *International Journal of Information Management* 2020; 54(102205): 1-10
- Keshta, I. & Ammar Odeh, A. "Security and privacy of electronic health records: Concerns and challenges". *Egyptian Informatics Journal* 2021; 22(2): 177-183.
- Khan, N., Naim, A., Hussain, M.R., Naveed, Q.N., Ahmad, N. & Qamar, S. "The 51 v's of big data: survey, technologies, characteristics, opportunities, issues and challenges, in COINS' 19" in Proceedings of the International Conference on Omni-Layer Intelligent Systems (Crete: ACM), May 2019, pp: 19–24. 10.1145/3312614.3312623
- Lang, M., Lemieux, S., Hébert, J., Sauvageau, G. & Zawati, M.H. "Legal and Ethical Considerations for the Design and Use of Web Portals for Researchers, Clinicians, and Patients: Scoping Literature Review". *Journal of Medical Internet Research*. 2021 Nov 11;23(11):e26450. doi: 10.2196/26450. PMID: 34762055; PMCID: PMC8663501.
- Lucassen, A., Montgomery, J. & Parker, M. "Ethics and the social contract for genomics in the NHS" In: Davies SC, Annual Report of the Chief Medical Officer 2016, Generation Genome. Department of Health, London. 2017.
- Lysaght, T., Lim, H.Y., Xafis, V. & Ngiam, K.Y. "AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research". *Asian Bioethics Review*. 2019 Sep 12;11(3):299-314. doi: 10.1007/s41649-019-00096-0. PMID: 33717318; PMCID: PMC7747260.
- Manrique de Lara, A. & Peláez-Ballestas, I. "Big data and data processing in rheumatology: bioethical perspectives". *Clinical Rheumatology*. 2020 Apr;39(4):1007-1014. doi: 10.1007/s10067-020-04969-w. Epub 2020 Feb 15. PMID: 32062767.
- McKeown, A., Mourby, M., Harrison, P., Walker, S., Sheehan, M. & Singh, I. "Ethical Issues in Consent for the Reuse of Data in Health Data Platforms". *Science and Engineering Ethics*. 2021 Feb 4;27(1):9. doi: 10.1007/s11948-021-00282-0. PMID: 33538942; PMCID: PMC7862505.
- Metcalf, J. & Crawford, K. "Where are human subjects in Big Data research? The emerging ethics divide". *Big Data & Society*. 2016;3(1):2053951716650211. doi: 10.1177/2053951716650211.
- Middleton, A. "Society and personal genome data". *Human Molecular Genetics*. 2018 May 1;27(R1):R8-R13. doi: 10.1093/hmg/ddy084. PMID: 29522190; PMCID: PMC5946868.
- Middleton, A., Milne, R., Almarri, M.A., Answer, S., Atutornu, J., Baranova, E.E., Bevan, P., Cerezo, M., Cong, Y., Critchley, C., Fernow, J., Goodhand, P., Hasan, Q., Hibino, A., Houeland, G., Howard, H.C., Hussain, S.Z., Malmgren, C.I., Izhevskaya, V.L., Jędrzejak, A., Jinhong, C., Kimura, M., Kleiderman, E., Leach, B., Liu, K., Mascalonzi, D., Mendes, Á., Minari, J., Wang, N., Nicol, D., Niemiec, E, Patch, C., Pollard, J., Prainsack, B., Rivière, M., Robarts, L., Roberts, J., Romano, V., Sheerah, H.A., Smith, J., Soulier, A., Steed, C., Stefànsdóttir, V., Tandre, C., Thorogood, A., Voigt, T.H., West, A.V., Yoshizawa, G. & Morley, K.I. "Global Public Perceptions of Genomic Data Sharing: What Shapes the Willingness to Donate DNA and Health Data?" *The American Journal of Human Genetics*. 2020 Oct 1;107(4):743-752. doi: 10.1016/j.ajhg.2020.08.023. Epub 2020 Sep 17. PMID: 32946764; PMCID: PMC7536612.
- Mikkelsen, R.B., Gjerris, M., Waldemar, G. & Sandøe, P. "Broad consent for biobanks is best - provided it is also deep". *BMC Medical Ethics*. 2019 Oct 15;20(1):71. doi: 10.1186/s12910-019-0414-6. PMID: 31615491; PMCID: PMC6794864.
- Monach, P.A. & Branch-Elliman, W. "Reconsidering 'minimal risk' to expand the repertoire of trials with waiver of informed consent for research". *BMJ Open (British Medical Journal Open)*. 2021 Sep 14;11(9):e048534. doi: 10.1136/bmjopen-2020-048534. PMID: 34521663; PMCID: PMC8442055.
- Mooney, S.J. & Pejaver, V. "Big Data in Public Health: Terminology, Machine Learning, and Privacy". *Annual Review of Public Health*. 2018 Apr 1;39:95-112. doi:

- 10.1146/annurev-publhealth-040617-014208. Epub 2017 Dec 20. PMID: 29261408; PMCID: PMC6394411.
- Muller, S.H.A., Kalkman, S., van Thiel, G.J.M.W., Mostert, M. & van Delden, J.J.M. "The social licence for data-intensive health research: towards co-creation, public value and trust". *BMC Medical Ethics*. 2021 Aug 10;22(1):110. doi: 10.1186/s12910-021-00677-5. PMID: 34376204; PMCID: PMC8353823.
- Murtagh, M.J., Blell, M.T., Butters, O.W., Cowley, L., Dove, E.S., Goodman, A., Griggs, R.L., Hall, A., Hal-lowell, N., Kumari, M., Mangino, M., Maughan, B., Mills, M.C., Minion, J.T., Murphy, T., Prior, G., Suderman, M., Ring, S.M., Rogers, N.T., Roberts, S.J., Van der Straeten, C., Viney, W., Wiltshire, D., Wong, A., Walker, N. & Burton, P.R. "Better governance, better access: practising responsible data sharing in the METADAC governance infrastructure". *Human Genomics*. 2018 Apr 26;12(1):24. doi: 10.1186/s40246-018-0154-6. PMID: 29695297; PMCID: PMC5918902.
- Nembawar, V., Johnston, K., Diallo, A.A., Kotze, M.J., Matimba, A., Moodley, K., Tangwa, G.B., Torrorey-Sawe, R. & Tiffin, N. "A framework for tiered informed consent for health genomic research in Africa". *Nature Genetics*. 2019 Nov;51(11):1566-1571. doi: 10.1038/s41588-019-0520-x. PMID: 31659323; PMCID: PMC7258308.
- Parasidis, E., Pike, E. & McGraw, D. "A Belmont Report for Health Data". *The New England Journal of Medicine*. 2019 Apr 18;380(16):1493-1495. doi: 10.1056/NEJMp1816373. PMID: 30995370.
- Peirce, A.G., Elie, S., George, A., Gold, M., O'Hara, K. & Rose-Facey, W. "Knowledge development, technology and questions of nursing ethics". *Nursing Ethics*. 2020 Feb;27(1):77-87. doi: 10.1177/0969733019840752. Epub 2019 Apr 28. PMID: 31032700.
- Prainsack, B. & Buyx, A. "A solidarity-based approach to the governance of research biobanks". *Medical Law Review*. 2013 Winter;21(1):71-91. doi: 10.1093/medlaw/fws040. Epub 2013 Jan 16. PMID: 23325780.
- Price, W.N. 2nd, & Cohen, I.G. "Privacy in the age of medical big data". *Nature Medicine*. 2019 Jan;25(1):37-43. doi: 10.1038/s41591-018-0272-7. Epub 2019 Jan 7. PMID: 30617331; PMCID: PMC6376961.
- Rauter, C.M., Wöhlke, S. & Schicktanz, S. "My Data, My Choice? - German Patient Organizations' Attitudes towards Big Data-Driven Approaches in Personalized Medicine. An Empirical-Ethical Study". *Journal of Medical Systems*. 2021 Feb 22;45(4):43. doi: 10.1007/s10916-020-01702-7. PMID: 33616768; PMCID: PMC7900081.
- Regulation (EU) 2016/679. [On line publication] "General Data Protection Regulation (GDPR)". 2026. Retrieved from <https://gdpr.eu/article-53-conditions-of-supervisory-authority/> [Consulted: 07/11/2022]
- Ristevski, B. & Chen, M. "Big Data Analytics in Medicine and Healthcare". *Journal of Integrative Bioinformatics*. 2018 May 10;15(3):20170030. doi: 10.1515/jib-2017-0030. PMID: 29746254; PMCID: PMC6340124.
- Robinson, K. "A false promise of COVID-19 'big' health data? Health data integrity and the ethics and realities of Australia's health information management practice". *Health Information Management Journal*. 2021 Jan-May;50(1-2):9-12. doi: 10.1177/1833358320941190. Epub 2020 Jul 17. PMID: 32680440.
- Salari, P. & Larijani, B. "Ethical Issues Surrounding Personalized Medicine: A Literature Review". *Acta Medica Iranica*. 2017 Mar;55(3):209-217. PMID: 28282721.
- Salerno, J., Knoppers, B.M., Lee, L.M., Hlaing, W.M. & Goodman, K.W. "Ethics, big data and computing in epidemiology and public health". *Annals of Epidemiology*. 2017 May;27(5):297-301. doi: 10.1016/j.annepidem.2017.05.002. Epub 2017 May 10. PMID: 28595734.
- Schiff, D. & Borenstein, J. "How Should Clinicians Communicate With Patients About the Roles of Artificially Intelligent Team Members?" *AMA Journal of Ethics* (American Medical Association Journal of Ethics). 2019 Feb 1;21(2):E138-145. doi: 10.1001/amajethics.2019.138. PMID: 30794123.
- Shaw, J.A., Sethi, N. & Cassel, C.K. "Social license for the use of big data in the COVID-19 era". *NPJ Digital Medicine*. 2020 Oct 2;3:128. doi: 10.1038/s41746-020-00342-y. PMID: 33083563; PMCID: PMC7532144.

- Stoeklé, H.C., Mamzer-Bruneel, M.F., Frouart, C.H., Le Tourneau, C., Laurent-Puig, P., Vogt, G. & Hervé, C. "Molecular Tumor Boards: Ethical Issues in the New Era of Data Medicine". *Science and Engineering Ethics*. 2018 Feb;24(1):307-322. doi: 10.1007/s11948-017-9880-8. Epub 2017 Mar 9. PMID: 28281147; PMCID: PMC5799317.
- Tiffin, N. "Tiered informed consent: respecting autonomy, agency and individuality in Africa". *BMJ Global Health*. (British Medical Journal Global Health) 2018 Dec 17;3(6):e001249. doi: 10.1136/bmjgh-2018-001249. PMID: 30613430; PMCID: PMC6304098.
- Überall, M. & Werner-Felmayer, G. "Integrative Biology and Big-Data-Centrism: Mapping out a Bioscience Ethics Perspective with a S.W.O.T. Matrix". *OMICS, A Journal of Integrative Biology*. 2019 Aug;23(8):371-379. doi: 10.1089/omi.2019.0043. Epub 2019 Jun 28. PMID: 31259670.
- van Gastel, B.E., Jacobs, B. & Popma, J. "Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study". *Journal of Parkinson's Disease*. 2021;11(s1):S19-S25. doi: 10.3233/JPD-202431. PMID: 34092652; PMCID: PMC8385496.
- Vayena, E. "Value from health data: European opportunity to catalyse progress in digital health". *The Lancet*. 2021 Feb 20;397(10275):652-653. doi: 10.1016/S0140-6736(21)00203-8. Epub 2021 Feb 8. PMID: 33571452.
- Vezyridis, P. & Timmons, S. "Resisting big data exploitations in public healthcare: free riding or distributive justice?" *Sociology of Health & Illness*. 2019 Nov;41(8):1585-1599. doi: 10.1111/1467-9566.12969. Epub 2019 Aug 18. PMID: 31423602.