

Record linkage as a vital key player for the COVID-19 syndemic – The call for legal harmonization to overcome research challenges

Julia Nadine Doetsch^{1,*}, Eero Kajantie^{2,3,4,5}, Vasco Dias^{2,3,4,5}, Marit S. Indredavik⁴, Randi Kallar Devold⁷, Raquel Teixeira¹, Jarkko Reittu^{8,9}, and Henrique Barros¹⁰

Submission History

Submitted:	30/11/2022
Accepted:	03/05/2023
Published:	13/06/2023

¹EPIUnit – Instituto de Saúde Pública, Universidade do Porto, Rua das Taipas, nº135, 4050-600 Porto, Portugal

²Institute for Health and Welfare, Public Health Promotion Unit, Helsinki and Oulu, Finland

³PEDEGO Research Unit, MRC Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland

⁴Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

⁵Children's Hospital, Helsinki University Hospital and University of Helsinki, Finland

⁶INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

⁷Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

⁸Finnish Institute for Health and Welfare, Legal Services, Helsinki, Finland

⁹University of Helsinki, Faculty of Law, Finland

¹⁰Departamento de Ciências da Saúde Pública e Forenses e Educação Médica, Faculdade de Medicina, Universidade do Porto (FMUP), Porto, Portugal

Key messages

1. Chronicity and social context influence COVID-19 risk highlighting its syndemic dimension
2. Record Linkage advances knowledge on COVID-19, associated chronic diseases, and social indicators
3. Further harmonization of data protection requirements for scientific research may create multilevel public health measures
4. As a multidimensional tool, it optimizes integrated strategies and fosters solidarity on Health in All Policies (HiAP)

Keywords

COVID-19; record linkage; routinely collected data, population-based cohorts; data privacy and protection

Severe failures in public health response? – The COVID-19 syndemic

The initial public health response to the COVID-19 pandemic aimed to prevent exponential dissemination and circumvent drastic collapses of healthcare systems [1]. Containment measures and required isolation promoted sedentary behaviours and stressful responses, which, as major determinants of chronic diseases, exacerbated prevalent co-morbidities. Patients with underlying chronic health conditions, older age, and less favourable social contexts have a threefold disadvantage: developing the disease with a higher risk, suffering a more severe course, and experiencing a fatal outcome [2]. Hence, the COVID-19 pandemic has a syndemic dimension [3], aggregating epidemics in a population, with social and complex biological interactions, which aggravate the burden of disease and challenge population-level forecasting.

Therefore, a better understanding of the association between physical and mental chronic diseases, socioeconomic status, and risk of COVID-19 adverse outcomes could have a transformative effect on controlling long-term consequences. Although multiple tools and data collection methods have been used to stimulate research on COVID-19, these population data, collected either routinely (e.g., electronic health records, prescription claims), or through population-based observational cohorts, are collected in separate data systems, so that yet too few COVID-19 trials use medical databases that have been previously linked.

Hence, a pressing demand to refine treatment requires a joint call: Record linkage – defined as the merging of data from an individual or an incident, not existing in a distinct record, into a combined dataset [4].

*Corresponding Author:

Email Address: jndoetsch@ispup.up.pt (Julia Nadine Doetsch)

Record linkage: A vital key player – “a call to action”

The COVID-19 pandemic urged health research to rapidly respond to pressing threats in a timely and coordinated manner, where strong connections through record linkage would serve as an essential asset. Against this background, Paprica et al (2020) have discussed the benefits of prospective record linkage to facilitate COVID-19 trials [5]. Combining expertise from clinical and healthcare services research can improve the comprehension of the importance of comorbidities, long-term outcomes, and demographic factors, and allow the investigation of rare outcomes, and prior healthcare system utilization by delivering robust data on the impact of COVID-19 [5].

Alongside routinely collected datasets, cohort data present high validity, accuracy, and effectiveness by providing estimates for incidence and the magnitude of disease determinants or health events over time. Linking cohorts with routinely collected data permits integration of individual information across different datasets and examination of the association between multiple chronic conditions when comparing individuals, health, and socio-economic status as well as changes at different time points. Thus, it enhances knowledge on COVID-19 and chronic diseases from a life course perspective enabling decelerating disease dissemination [6].

Furthermore, by contrast to setting up a new data collection or comparing individual data sets, it is a cost-effective and time-saving upgrade resulting in an efficient, powerful, and vital data collection tool by enabling big data handling, continuous data collection on cross-sectoral services, fast-paced data circulation, convenient observation of patient's health status, and comprehensive follow-up. Its potential opportunity was demonstrated in the WOSCOPS 20-year follow-up study in Scotland and its success was proven in the UK RECOVERY trial for COVID-19 [5].

However, as Paprica et al (2020) have argued, there are caveats as well: requirements of understanding of data quality limitations, excellent knowledge of database holdings, case validation work, and public and trial participants' support in data usage [5]. Aside from technical and methodological record linkage challenges, and as well as the need for data quality, accuracy and representativeness across disciplines and countries, or the sustainability of data infrastructures for data harmonization, the legal structure in implementing the GDPR across the European Economic Area (EEA) depicts a caveat in the objective of legal harmonization, which we would like to discuss in the following.

GDPR and legal dilemmas across EU/EEA – A research challenge

Secondary use and linkage of data collected directly from cohort participants based on individual consent, which is sufficient to facilitate linkage for these participants, is a major challenge due to data protection and privacy rights of data subjects. In 2018, the European General Data Protection Regulation (GDPR) was implemented as an

overarching, robust, and inclusive legal framework across the EEA.

Freedom of member states to implement GDPR clauses on health data processing, either for the administration of the healthcare systems or for reasons of public health and research purposes [Articles 9.2.h, 9.2.i, 9.2.j, 9.4.] is partially responsible for the existing limitations to the continuation of cross-national research. At the national level, the aspiration to provide a high degree of data protection and an emphasis on consent may jeopardise and place constraints on scientific research processes to a substantial extent and amplifies the complexity of record linkage, within and across member states.

While intended to promote the free flow of data within the European Union (EU), data transfers to countries outside the EU/EEA or international organizations [Articles 44–50] are only permitted under alternative conditions: Having an adequacy decision issued by the European Commission [Article 45], providing appropriate safeguards, including binding corporate rules and complex contractual arrangements with standard privacy clauses [Article 46; 47], or, as an exceptional and temporary measure, through “specific derogations [Article 49] including public interest and explicit consent [Article 9/2/a)].

According to the European Data Protection Board, such derogations should not be used for repetitive transfers of long research projects. Yet, given the statements that “The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right” [Article 4]; how can impeding research barriers be overcome? Privacy-Preserving Record Linkage techniques aiming to undergo record linkage without revealing actual values of personal identifying attributes due to data privacy concerns, offer a possibility for research to address some privacy requirements where research purposes can be fulfilled through pseudonymisation. Notwithstanding, the required level of pseudonymisation to lawfully process data continues to be highly debated due to difficulties that derive from the massive amount of data, multiple data sources, and ‘dirty’ data [7].

Next steps forward for the research agenda – the call for harmonization

Indeed, the balance between data protection and the availability of information for research for the public good has not been struck yet. Researchers from member states face challenges to overcome variances in the national implementation of the GDPR. In effect, legal discrepancies have been proving detrimental to research in member states, including those which already had established a margin for research for the common good without explicit consent [8].

In Portugal, health-related scientific research essentially relies on consent as legal grounds for personal data processing. Even though, the obligation to collect informed consent for the participation in non-interventional clinical studies can exceptionally be derogated by the determination of the Competent Ethics Commission, consent for the processing of personal data is still required in those cases, as the Portuguese data protection act clarified, before the GDPR [9]. Following

the GDPR approval, the Portuguese new data protection act [Law n° 58/2019, 8 August] timidly touched upon the subject of scientific research, save from the possibility of giving consent to “certain areas of research” (inspired by recital 33, GDPR). Portugal’s legal system, so far, seems to have privileged informational self-determination over other individual rights and collective interests, such as access to information, freedom of research and the advance of science.

In other member states, such as Finland, the legal system puts a strong emphasis on the public good, making the linking of cohort data with routine administrative data or registries easier, especially since the entrance of the national legislation further implementing the GDPR [10].¹ Moreover, the usage of unique personal identification numbers for research without explicit consent for the majority of register-based research in Finland allows for linking research data, expanding the data available to individuals, detecting overlap between data collections, and facilitating the reproduction of research results [10]. The differences between several member states have been addressed in the country-comparative article by Doetsch JN, Dias V, Indredavik MS *et al.* (2021) [10].

If the intention is to study data across more than one cohort or population, not only the linkage of data but also the harmonization of data is needed. Harmonization is defined as enhancing consistency in the use of data elements in terms of their meaning and presentation format [11]. Harmonization of data helps to surpass national obstacles that can hinder health research that contribute to the public good, to generate comparable data across different data sources, and to facilitate record linkage of cross-national data exchange for multinational projects, leading to unique opportunities for health research across member states.

In the discussion of federated data analysis and legal compliance, data harmonization is an interrelated process and requirement. Harmonization of data across multiple jurisdictions might substantially simplify the implementation of privacy enhancing technologies, namely enabling distributed analysis (“federated learning”) without data leaving the jurisdictions in which they are located or simply having access to non-personal data such as data catalogues or statistical outputs. In effect, one of the possible advantages is that they allow federated database analyses and the extraction of aggregated anonymised data through a joint platform. Another decisive advantage over conventional data models is the guarantee of legal certainty. Moreover, federated learning has already successfully been applied in some European projects, such as the RECAP preterm project [11].

Given current discussions, we argue in line with Davies, Jones and Conolly (2018), that an important point to consider in order to increase the likelihood of giving consent toward broader record linkage and harmonization is the public attitude and expectation. Thereby, the four main points are i) the importance of organizational trust and legitimacy² that leads to a societal benefit; ii) continuous request for consent

¹‘Assessment of the EU Member States’ rules on health data in the light of GDPR’ Specific Contract No SC 2019 70 02 in the context of the Single Framework Contract Chafea/2018/Health/03 https://ec.europa.eu/health/system/files/2021-02/ms_rules_health-data_en_0.pdf.

²Increasing transparency and trust in the digital economy and European space has been one of the main objectives of the GDPR (GDPR, Recital 6 and 7)

as decisions may change with time (e.g. dynamic consent model [9]; iii) high transparency of data usage; iv) data linkage communication on the usage of data (e.g. written notifications, by mail or email) [12].

Therefore, we support that further harmonization of data protection requirements for scientific research activities in the EU/EEA should be pursued, focusing in particular on health-related research [13]. Furthermore, such harmonization efforts should not ignore and should be committed to the goal of taking full advantage of the flexibilities provided by the GDPR for scientific research, without prejudice to ensure a high level of protection of the rights and freedoms of data subjects.

Science, Solutions & Solidarity – fostering health in all policies in light of research

Thus, in line with Paprica *et al.* (2020), we recommend that data assets on COVID-19 should be linked to amplify their scientific value and impact on society. We call for collaboration between study participants, data managers, and research funders to make prospective linkage of routinely collected data with cohort data the norm, beginning with COVID-19 trials [5]. We argue that research funded by taxpayers calls for a wide range of possibilities, i.e., linking cohort data and routinely collected data which should be explored to their full potential. In the following, we exemplify three main considerations “Science, Solutions and Solidarity”, in line with the World Health Organization (WHO).

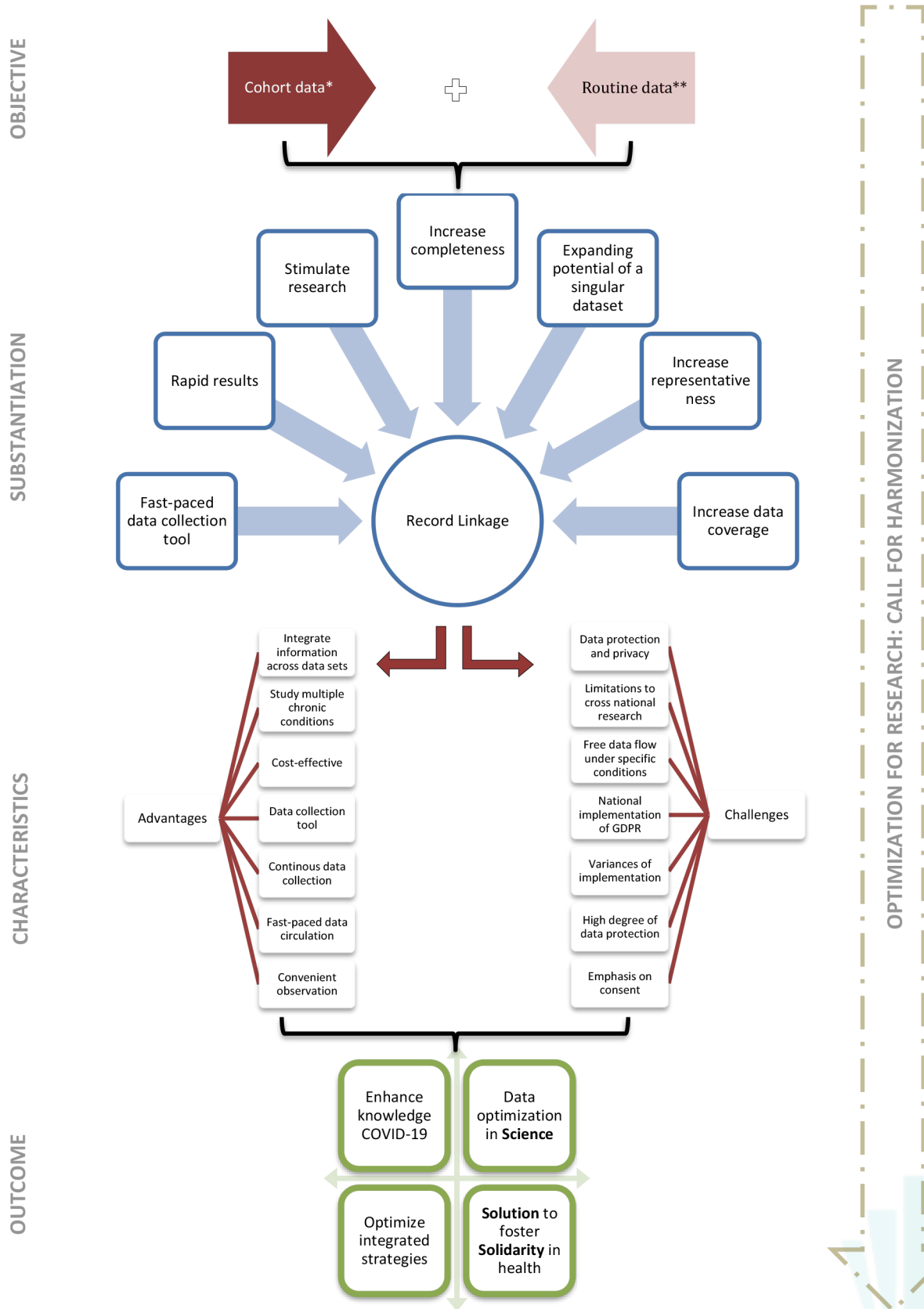
Linking cohort data and routinely collected data facilitates the manifold demand for research optimization in science. In that sense, the WHO communicated “Science, Solution, and Solidarity” asserting togetherness in managing the COVID-19 pandemic. This can promote equity in healthcare with promising assets advancing knowledge in understanding the multiplicity of chronic diseases and identifying the association with COVID-19. Congruently, the recent proposal of the European Commission for the creation of a European Health Data Space, aimed at, among other goals, providing a consistent, trustworthy and efficient set-up for the secondary use of health data for research, is a very welcomed step forward, although requiring a coherent articulation with the existing data protection landscape [14].

Hence, record linkage as a multidimensional tool may ultimately enable defining and optimizing integrated strategies. We summarized the main points of this commentary in a framework: the advances of record linkage for research optimization on the COVID-19 syndemic, its challenges embodied with a legal focus, and the proposed solution (Figure 1). Further harmonization of data protection requirements for scientific research may create multilevel public health measures as a solution to foster solidarity on health in all policies.

In a nutshell – future recommendations

Concluding, we would like to highlight the commentaries’ four key messages. Firstly, chronicity and social context influence COVID-19 risk highlighting its syndemic dimension that

Figure 1: Record Linkage—A multidimensional tool for research optimisation: a call for harmonisation



Legend: *data from a group of individuals sharing a statistical factor in common in a demographic study
 **data collected for purposes without specific a priori research objectives before collection

demands to refine treatment through record linkage. Secondly, record linkage of routinely collected data and data collected through observational population-based cohorts advances

knowledge on COVID-19, associated chronic diseases, and social indicators. Thirdly, further legal harmonization of data protection requirements for scientific research may enhance

multilevel public health measures where legal challenges in record linkage for health research across EU/EEA countries would be easier to overcome with the help of proper policies and suitable technical and methodological tools. An example thereby is federated data analysis, or other privacy enhancing solutions, which on the other hand, rely on other technical aspects like data harmonization and the sustainability of data curation infrastructures etc. The intersection between these two layers (legal and technical) should not be forgotten. Fourthly, record linkage is a multidimensional tool optimizing integrated strategies for health policy and fostering solidarity on Health in All Policies (HiAP) based on WHO's key aims "science, solution and solidarity". In a proposed summarising framework, we showed how linking data is vital for research optimisation due to its multidimensional possibilities.

Aside the harmonisation goals addressed in the commentary, future guidance should include consistent adherence to data standards, data quality assurance, fostering a collaborative environment across data controllers towards common solutions, and pursuing representation in data to ensure equity.

Competing interest statement

All authors have completed the Unified Competing Interest form and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

Transparency declaration

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Ethics Statement

Ethical approval was not required because the manuscript is a commentary. No participants or data were involved.

Funding and study sponsors

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this study was supported national funding from the Foundation for Science and Technology—FCT (Portuguese Ministry of Science, Technology and Higher Education), under the Unidade de Investigação em Epidemiologia—Instituto de Saúde Pública da Universidade do Porto (EPIUnit) [UIDB/04750/2020]; and by the RECAP preterm project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 733280.

Statement of independence of researchers from funders

All authors confirm their independence as researchers from funders. All authors, external and internal, had full access to the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Declaration of conflicting Interests

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

Contributor and guarantor information

HB conceived the idea for the study. JD, EK, MI, RD, RT, VD, JR, HB were involved in the design of the study. JD wrote the Commentary under the close supervision of HB, who participated in the drafting, interpretation, intellectual content and revision. EK, MI, RD, RT, VD, JR provided critically important comments for the intellectual content and suggested alterations to the text. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

We would like to acknowledge Evert-Ben van Veen who provided critically important comments on the text.

References

1. Marleen Bekker, Damir Ivankovic OB. Early lessons from COVID-19 response and shifts in authority: public trust, policy legitimacy and political inclusion. *Eur J Public Health* 2020;39:2019–20. <https://doi.org/https://doi.org/10.1093/eurpub/ckaa181>.
2. Tisminetzky M, Delude C, Hebert T, Carr C, Goldberg RJ, Gurwitz JH. Age, Multiple Chronic Conditions, and COVID-19: A Literature Review. *Journals Gerontol Ser A* 2020;XX:1–7. <https://doi.org/10.1093/gerona/glaa320>
3. Horton R. Offline: COVID-19 is not a pandemic. *Lancet* 2020;396:874. [https://doi.org/10.1016/S0140-6736\(20\)32000-6](https://doi.org/10.1016/S0140-6736(20)32000-6)
4. European Commission. Commission and Germany's Presidency of the Council of the EU underline importance of the European Health Data Space 2020. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2049 (accessed February 23, 2021).

5. Paprica PA, Sydes MR, McGrail KM, Morris AD, Schull MJ, Walker R. Prospective data linkage to facilitate COVID-19 trials - A call to action. *Int J Popul Data Sci* 2020;5:1–2. <https://doi.org/10.23889/IJPDS.V5I2.1383>
6. Tingay KS, Bandyopadhyay A, Griffiths L, Akbari A, Brophy S, Bedford H, et al. Record linkage to enhance consented cohort and routinely collected health data from a UK birth cohort. *Int J Popul Data Sci* 2019;0.
7. Han S, Shen D, Nie T, Kou Y, Yu G. An enhanced privacy-preserving record linkage approach for multiple databases. *Cluster Comput* 2022;25.
8. Skovgaard LL, Wadmann S, Hoeyer K. A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good. *Health Policy (New York)* 2019;123:564–71. <https://doi.org/10.1016/j.healthpol.2019.03.012>
9. Doetsch JN, Dias V, Redinha R, Barros H. Record linkage of routine and cohort data of children in Portugal: challenges and opportunities when using record linkage as a tool for scientific research 2022:1–25. <https://doi.org/10.1093/medlaw/fwac040>
10. Doetsch JN, Dias V, Indredavik MS, Reittu J, Devold RK, Teixeira R, et al. Record linkage of population-based cohort data from minors with national register data: a scoping review and comparative legal analysis of four European countries. *Open Res Eur* 2021;1:58. <https://doi.org/10.12688/openreseurope.13689.2>
11. Zeitlin J, Sentenac M, Morgan AS, Ancel PY, Barros H, Cuttini M, et al. Priorities for collaborative research using very preterm birth cohorts. *Arch Dis Child Fetal Neonatal Ed* 2020;105:538–44. <https://doi.org/10.1136/archdischild-2019-317991>
12. Davies, Malen; Jones, Hollie; Conolly A. Public Attitudes to Data Linkage A report prepared for University College London by. 2018.
13. Townend D. Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? *Hum Genet* 2018;137:657–64. <https://doi.org/10.1007/s00439-018-1924-x>
14. de Bienassis K, Fujisawa R, Cravo Oliveira Hashiguchi T, Klazinga N, Oderkirk J. Health data and governance developments in relation to COVID-19: How OECD countries are adjusting health data systems for the new normal. 2022.

