International Journal of Law and Psychiatry 60 (2018) 40–44

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

Privacy in Big Data psychiatric and behavioural research: A multiple-case study

M. Mostert *, B.M. Koomen, J.J.M. van Delden, A.L. Bredenoord

Department of Medical Humanities, Julius Center for Health and Primary Care, University Medical Center, Utrecht, the Netherlands

ARTICLE INFO

Article history:
Received 31 January 2018
Received in revised form 3 July 2018
Accepted 6 July 2018
Available online xxxx

Keywords:
Big Data
Psychiatry
Research
Privacy
Qualitative research

ABSTRACT

In Big Data health research, concerns have risen about privacy and data protection. While the ethical and legal discussion about these issues is ongoing, so is research practice. The aim of this qualitative case study is to gain more insight into how these concerns are currently dealt with in practice. For this multiple-case study, the YOUth cohort, a longitudinal cohort focusing on psychosocial development, and Big Data Psychiatry, a pilot study in Big Data analytics on psychiatric health data, were selected. A broad range of relevant documents were collected and semi-structured interviews with stakeholders were conducted. Data were coded, studied and divided into themes during an iterative analytical process. Three themes emerged: abandoning anonymisation, reconfiguring participant control, and the search for guidance and expertise. Overall, the findings show that it takes considerable effort to take privacy and data protection norms into account in a Big Data health research initiative, especially when individual participant level data need to be linked or enriched. By embracing the complexity of the law in an early phase, setbacks could be prevented, the existing flexibility within the law could be utilised, and systems or organisations could be designed and constructed to take relevant rules into account. Our paper illustrates that a close collaboration of experts with different backgrounds within the initiative may be necessary to be able to successfully navigate this process.

1. Introduction

Big Data is finding its way into health research. Some believe that this will provide unprecedented opportunities for psychiatry (Monteith et al., 2015). A broad range of issues, however, need to be dealt with. One of the key areas of concern in Big Data health research is related to privacy and data protection (Mittelstadt and Floridi, 2016), especially when psychiatric or other sensitive health-related data are collected, re-used, linked and analysed.

The rise of such data-intensive health research initiatives has sparked a lively debate about how the use of data should be governed by principles and rules, especially during the adoption of the General Data Protection Regulation (GDPR) in the EU (Mostert et al., 2016; Ploem et al., 2013; Sethi, 2015). Although this debate on normative issues is ongoing, researchers and other stakeholders already need to deal with challenges related to privacy and data protection on a daily basis. They cannot wait until the normative framework is sufficiently crystallized. They are confronted with a level of normative complexity and uncertainty which could have a negative impact, both on achieving scientific goals and on the protection of relevant rights and interests. In the UK, for example, a study has shown that the confusing nature of the regulatory landscape resulted in a culture of caution and (overly) conservative approaches to data sharing (Sethi and Laurie, 2013).

Against this background, some health research initiatives have attempted to engage with and utilise the potential of Big Data, while at the same time ensuring privacy and data protection. To our knowledge, no qualitative research has been published about how this challenge is dealt with by relevant stakeholders in the specific context of such groundbreaking initiatives. By mapping the relevant challenges faced and solutions sought by those involved in the organisation of such initiatives, valuable lessons can be learned. In this qualitative case study, we analyse two real-world examples of data-intensive psychiatric and/or behavioural research. The study is designed to provide insight into challenges related to privacy and data protection in data-intensive health research, and aims to contribute to a better understanding of how rules and interests can be taken into account in a specific initiative or context.

2. Methods

A qualitative multiple-case study has been conducted. The case study is a commonly used empirical research methodology, which allows the researcher to investigate a phenomenon in depth and within its real-world context (Baxter and Jack, 2008; Yin, 2014). Information

* Corresponding author at: Julius Center, room no. STR 5.133, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands.
E-mail address: m.mostert-2@umcutrecht.nl (M. Mostert).

https://doi.org/10.1016/j.ijlp.2018.07.002
0160-2527/© 2018 Elsevier Ltd. All rights reserved.

was gathered about the Big Data Psychiatry pilot project (hereafter: BDP) and the YOUth cohort (hereafter: YOUth). This multiple-case study has been evaluated and exempted from further ethical scrutiny by the Research Ethics Committee of the University Medical Center Utrecht. Explicit informed consent has been obtained from all respondents and the management of both initiatives.

2.1. Case selection and background

The cases have been selected because of their approaches to different aspects of Big Data research. BDP employs a Big Data approach to its analytical methods, in particular for aiding in hypothesis generation. In YOUth, another aspect of Big Data is reflected in its comprehensive data collection, which is continuously being supplemented and updated. Although no clear and widely accepted definition of Big Data exists, such innovative ways in which data are analysed or captured are considered to be core building blocks of a Big Data approach (Mayer-Schönberger and Ingelsson, 2018).

The first case, BDP, aims to explore the potential of Big Data analytics in gaining new insights in the complex psychiatric phenotype. The ultimate goal in BDP is to develop a Big Data analytics instrument that will support health care professionals in their daily practice, for instance by predicting the chance of side effects of medication on the basis of individual patient profiles (Scheepers et al., 2018). A relatively limited set of databases, related to a group of psychiatric patients in Utrecht, was used in the pilot phase of BDP. As a proof of concept, the Cross Industry Standard Process for Interactive Data Mining (CRISP-IDM) was performed on these databases. This resulted in a number of hypotheses and findings, including those related to the themes of aggression during hospitalisation and the effects of medication (Menger et al., 2016). Four working groups have been formed in BDP, and one of these working groups is committed to exploring the theme of privacy and confidentiality. This multi-disciplinary working group focuses on how to safeguard the privacy of participants in the pilot phase and the future programme.

The second case, YOUth, is a longitudinal cohort. YOUth aims to explain why some children develop well and others fail to thrive in society by examining how neurocognitive development mediates the influence of biological, child-related and environmental determinants on behavioural development. The cohort study focuses on psychosocial development, ranging from normal development to deviant behaviour and psychiatric disorders. In order to do so, a great variety of health-related data are continuously collected. These data vary from an array of behavioural and cognitive test results to data about environmental, general child and biological factors (including results from EEG and MRI examinations). The YOUth data being collected will also be linked to other data sources for a broad range of future studies, all in the field of behavioural and psychiatric research.

2.2. Data collection

During our data collection phase, both factual information and the views of different stakeholders from the two cases were collected. The factual information includes internal reports of meetings and discussions, research protocols and other documentation, files related to the application for ethical approval, and text on public websites. Our data collection in YOUth took place between February and April 2017, and in BDP between November 2015 and January 2016. The stakeholders were selected on the basis of their variation in backgrounds and involvement in dealing with privacy and data protection related issues related to the cases. Among the stakeholders, the following areas of expertise or backgrounds are represented: management, lead researcher, research staff, privacy and health law, information technology, consultancy, data management, and patient representation. We conducted 14 semi-structured qualitative interviews in total to collect the views of the stakeholders in both cases. The stakeholders were asked questions related to the challenges they experienced regarding privacy and data protection, and how these challenges were dealt with or should be dealt with according to their views.

2.3. Data analysis

After collecting data, our research group developed codes and identified themes. The full transcripts and other relevant collected data were coded using NVivo. Mostert and Koomen coded the gathered data. Mostert and/or Bredenoord read the coded data and checked the codes for consistency. During the process of analysis, the codes were adjusted through constant comparison across the transcripts and other relevant data and through discussion within the research group. After reaching consensus on the coding, the themes mentioned below were identified by analysing the data. All interviews were conducted in Dutch and the quotes in the results section have been translated idiomatically. The results were presented to respondents to be checked for accuracy.

3. Results

During the process of analysis, it became clear that all respondents encountered challenges or issues related to privacy and data protection. After analysis of the interviews and the other information, three main themes emerged: abandoning anonymisation; reconfiguring participant control, and; the search for guidance and expertise.

3.1. Abandoning anonymisation

The first theme concerns the move away from anonymisation as a strategy to prevent the applicability of data protection law. During the first meetings of the working group on privacy and confidentiality in BDP, some of the respondents adjusted their view on what data could be regarded as anonymous. In this phase, the importance of distinguishing between pseudonymous and anonymous data became clear, but the difficulties in making this distinction were also acknowledged:

"(..) the difference between anonymous and pseudonymous data is hard to understand by layman, and it turned out that it is incredibly difficult for jurists to explain what this difference is. Only after this difference has been made clear, you are able to proceed (..)."(R1BDP).

Afterwards, it became clear to all respondents in BDP that irreversible anonymisation according to the standards as set out in the forthcoming GDPR would severely limit the use of data. Another way to proceed had to be found. BDP chose to integrate a Trusted Third Party (TTP) in the data warehouse architecture of BDP. A TTP aims to facilitate the data linkage process on behalf of multiple data holders in a secure way. Only data that are relevant to a certain research question are extracted from local data sources by the TTP. Afterwards, the different personal data sources are linked by the TTP and a unique pseudonym is assigned to the linked data to prevent future data linkage or enrichment on the individual participant level. The TTP was not considered to be a viable solution in YOUth as it would hinder a permanent enrichment of the cohort with external data sources:

"Or a sort of Trusted Third Party, that is always complicated… because than you need to link data for every single research question and that is a barrier to this kind of cohorts. (..) sometimes I just want to enrich my whole dataset (..)." (R11YOUth).

Furthermore, respondents in both cases regarded de-identifying or pseudonymising data as a challenge, especially when it pertained to unstructured or rich data sources, such as open text fields or imaging data. One of the respondents emphasised the difficulties in de-identifying such data as follows:

"Once you start working with big data, (..) you could potentially link data sources to enrich the profile of people in such a way that identification may become very easy. (..). With a limited number of variables you could

already get such unique information that someone could be identified." (R7BDP).

To deal with the above-mentioned challenges, organisational and technical measures were suggested or implemented in both cases. In YOUth, a data access committee was being installed to ensure control over which data would be released, under what conditions and to whom. An important task of this committee would be to determine whether data could be shared without a risk of re-identification. This, however, was considered to be a difficult and time-consuming task. A research data platform was being developed in which this process would be partially automated, while promoting reproducibility and transparency. The value of such a platform was described as follows:

"At the moment, we are building a research data platform (..) in which we will combine our expertise to store the data prepared for release and which allows the data manager to easily assess whether a combination of data and variables may lead to re-identification. And the system will (..) prevent that this combination of data will be released, (..) And the second advantage is that you will register everything that is done with the data." (R12YOUth).

An alternative to sharing the data itself, mentioned by respondents from both cases, is to only release data analyses. In this way, the data would remain local and only analyses, which would bear no risk of re-identification, would be shared with third parties. Respondents noted that such a system solves many privacy-related problems, but also limits potential data use. One respondent in YOUth, for instance, suggested that this approach would exclude the permanent enrichment of one cohort with the data from another cohort, because this would require access to the raw data.

3.2. Reconfiguring participant control

The second and most discussed theme concerns the challenge of allowing participants to control the use of their data. In BDP, most respondents agreed that it is important to obtain informed consent from participants for the re-use and linkage of their personal data. At the same time, some of these respondents recognised the disadvantages of this approach, such as a possible lack of inclusion of data from patients who are already underrepresented in health research and the risk of consent bias. In meetings of the working group on privacy and confidentiality in BDP, participants discussed the fact that it is not always required to obtain informed consent for research on personal data, according to Dutch law, and that there could be other ways to allow participants to exercise control. Furthermore, the importance of making participants aware of the use of their psychiatric data was stressed by some of the respondents in BDP, as the following quote illustrates:

"*There is something special about psychiatric data. (..) some people may not agree, but a diagnosis is not always in your best interest. (..) it is in particular sensitive data because it is about your mental well-being, your mental state, with all kinds of possibilities and impossibilities in the work sphere. It therefore is, in short, very private and sensitive information. When you will link such data on the individual level, it becomes relevant whether the participant is aware of this.*" (R5BDP).

A question raised in both cases was how to obtain informed consent from participants when their personal data would be used for a broad range of purposes and would be linked to other data sources. In YOUth, multiple respondents described that concerns among the participants arose because they were explicitly asked for consent to request their data from several other databases. These respondents found it hard to eliminate these concerns among their participants related to future data linkages with external data sources.

Another topic of discussion was what should be done to allow participants to control the use of their data in the long-term. In YOUth, researchers especially struggled with the question whether re-consent should be obtained from children once they reached adulthood. The Research Ethics Committee (REC) recommended actively offering children the possibility of opting-out as soon as they would become legally competent. A suggestion of multiple respondents in BDP was that, in the future, all patients should be enabled to be informed and adjust their preferences in an online environment, such as a patient portal. One of the respondents underlined the importance of utilising digital tools as follows:

"I think that such a tool is a beautiful way to not take away the control from participants. That you will not ask for consent once and open the floodgates. And I think that it is a good way to ensure the trust of people in such a whole project, because that is what it is all about." (R5BDP).

Other respondents in BPD, however, pointed out that while developments like dynamic informed consent procedures and patient empowerment are laudable, they could also threaten research, mainly because of the risk of (consent) bias. Moreover, some respondents stressed that the availability or accessibility of appropriate tools to implement dynamic consent procedures was lacking.

Interesting in this context is the fact that multiple respondents emphasised that safeguards and measures other than informed consent are just as important when it comes to ensuring the trust of participants and/or the public. The active participation of participants was regarded as an important component of these measures. In both cases, they were still in the phase of exploring ways to implement patient participation in their project. In YOUth, a parent representative panel was already established and embedded in the organisation. One of the respondents emphasised the importance of such measures as follows:

"Constantly involving them [participants] *in what we are doing. Is this possible when you view it from a patient perspective? Is it right? Is there support for doing it this way?"* (R6BDP).

Other suggested measures include the clear designation of responsibilities and tasks, implementing accountability and oversight mechanisms, and certification of the security measures taken. A measure that received particular attention in YOUth was the implementation of policy on the return of clinically relevant incidental findings, after the REC urged YOUth to elaborate on what to do with this kind of findings.

3.3. The search for guidance and expertise

The third theme concerns the search for guidance and expertise, which was regarded as an issue by the majority of respondents in both cases. Respondents described that they often struggled with uncertainties, mainly about legal norms, when setting up the project or cohort. To deal with uncertainties in the field of privacy and confidentiality, a working group was established in BDP in an early phase. Related to the feedback and advice of this working group, a respondent noted:

"It still really is a matter of pioneering and finding out… It really surprises me that all those members of the privacy group say: you are really front-runners and it is very praiseworthy that you are carefully addressing this. And at the same time, it is by far not good enough. Then I think, wow, if this is the case, than it tells you something about the state of affairs, also in other areas. (…).

And we try to be the most virtuous of them all (..). *As a result, we literally have been delayed several times."* (R1BDP).

To avoid such delays or other setbacks, some respondents underlined that it is essential to take into account the legal and ethical aspects at the very beginning of the project. The need for collaboration between different areas of expertise, such as legal, security and ICT was also emphasised. Furthermore, many respondents expressed the need for more specific, uniform and up to date guidelines and best practices for data linkage, on the national and/or international level. Some respondents suggested that a national advisory body, where all the needed expertise and guidance is concentrated, would be of great value. In YOUth, this need for more specific and uniform guidance was especially focussed on how to obtain consent for future data linkage, as this quote illustrates:

"Well, a dilemma I encounter is that there are or were few good examples, and I still wonder if we are doing it right, how to obtain consent for that linkage. (..). *Also because I think that every institution has different requirements and that is what makes it so complicated, because you really want one golden standard for how it should be asked."* (R11YOUth).

Some respondents explained why it was so hard to provide clear and uniform guidance on privacy aspects. A first factor is that care and research are closely related, and that they are getting more and more intertwined. Secondly, respondents pointed out that there are many (upcoming) changes in laws and regulations. In addition, the existing national code of conduct related to the use of personal data in health research was not updated for years. The consequences of these developments, and the intention of the legislative bodies, were understood as follows by one of the respondents:

"It meant that everybody had to guess what the best approach would be. And what we saw is that the legislator perhaps consciously created a grey area to be sure that people would keep thinking and to prevent fixed rules. This means it is not that strange that if you talk to people, who do research or are busy with patient care, that they do not really understand." (R7BDP).

4. Discussion

By studying the two cases, we aim to provide insight in how challenges related to privacy and data protection are dealt with in real world examples of data-intensive psychiatric and behavioural health research. A first insight is that anonymisation was regarded largely impracticable in both cases, especially when data sources needed to be linked or enriched. Organisational and technical measures have been implemented with the aim of complying with the law and mitigating risks, which most notably resulted in the use of a TTP and the development of a research data platform to access and link data. Secondly, it becomes clear that the search for meaningful and proportionate ways to allow individuals to control and be aware of the use of their data is ongoing. This aspect is considered of great importance by some, especially when it comes to linkage of personal data related to psychiatric disorders. Improvements of the (one-off) broad informed consent procedure are being considered, in particular by means of utilising digital tools to ensure a more ongoing engagement with participants. Thirdly, uncertainty about how to comply with the law is perceived as having negative impact on the initiatives. This uncertainty is mainly attributed by respondents to a lack of easily accessible expertise and guidance, but also to the recent changes in data protection law.

An issue that connects all the described themes and many of the findings is the struggle with legal complexity. Respondents reported a broad range of negative consequences related to legal complexity, such as uncertainty, delays and other setbacks. This critique of the law does not seem to be unique to these cases. Laws that govern the use of data for research, and in particular data protection laws, are often reported to be confusing, open to varying interpretation or burdensome (Sethi and Lauri, 2013; Koops, 2011). However, what needs to be taken into account is that broad or open norms also have advantages. Without open norms, the law would be static and inflexible. This would result in major problems, since the multifaceted and continuously evolving data-intensive health research landscape requires a considerable degree of flexibility. Paradoxically, it has been pointed out that *because* of the complex regulatory landscape, the existing flexibility within the legal framework to address some of the regulatory hurdles is often overlooked in practice (Sethi and Laurie, 2013).

In the cases, multiple ways forward were suggested that could help mitigating legal complexity. One of these suggestions is the drawing up of context-specific guidelines for data linkage. Currently, some initiatives like BBMRI-ERIC are already working towards official approval under the GDPR of an international code of conduct for personal data processing in health research (BBMRI-ERIC, 2017). Although a context-specific code of conduct could indeed help reduce legal uncertainty, it is unrealistic to expect that this will mitigate the need to deal with legal complexity on the level of the initiative. Open and often fluid norms will remain, and these norms will need to be interpreted and translated into concrete, effective and proportionate rules and measures on the level of each single data initiative.

From the cases, we can learn how this challenge of taking the complex interplay of norms and practical requirements into account could be approached. A first key element of the approach is to identify and address privacy-related issues in an early phase. In BDP, the identification of privacy-related issues has in particular been done by establishing a multi-disciplinary working group on privacy and confidentiality during the beginning of the pilot phase. The meetings of this working group inspired multiple important decisions made and measures taken in BDP, including the move away from anonymisation as the main compliance strategy. Afterwards, discussions within BDP focussed on finding another way to link multiple personal data sources while respecting relevant rules on privacy and data protection. In YOUth, the search for a way to link personal data sources to enrich the cohort in a responsible and effective way also emerged during the start-up phase. This brings us to a second key element of the approach in both cases, which can be labelled as *privacy by design*. The privacy by design approach is a form of value-sensitive design, which is characterised by the embedding of privacy-enhancing and preserving measures directly into the design and operation of systems, processes and organisations (Cavoukian, 2010). The use of a TTP in BDP and the development of a research data platform in both cases are measures that resemble this approach. Other examples of research data platforms have been described in the literature, with the initiative called DataSHIELD serving as a prime example. The DataSHIELD platform demonstrates how a value-sensitive design could contribute to linking and utilising multiple data sources for health research purposes, while respecting relevant rights and values (Budin-Ljøsne et al., 2015; Wallace et al., 2014; Wolfson et al., 2010). Another incentive to embrace a value-sensitive design approach is provided by the GDPR. In the GDPR, a novel legal obligation emerged under the title 'Data protection by design and by default'. This obligation is subject to high fines and considered to be among the most innovative and ambitious norms of the GDPR (Bygrave, 2017). Nevertheless, it has proven very difficult to encode data protection principles and rules in systems for data processing in health research. The idea of integrating legal norms in information processing systems during the design phase is by some regarded to be "at odds with the dynamic and fluid nature of many legal norms" (Koops and Leenes, 2014). It therefore seems that both the potential and the limits of such value-sensitive design approaches need to be recognised in practice and deserve further study.

Finally, some of the measures taken in the cases illustrate that compliance with the law was not regarded sufficient to guarantee that all relevant rights, interests and values would be taken into account. In YOUth, for instance, the presence of a policy on the return of clinically relevant unsolicited findings was considered to be an essential measure to respect participant's interests. Another example was the ambition in both cases to involve and engage with participants in a meaningful way, and on an ongoing basis. These (and other) measures go beyond what is required by the law, but may be necessary to meet the expectations of participants and society regarding the conduct and activities in data-intensive health research. What the consequences of a failure to meet these expectations might be is illustrated by the disturbing developments in the care.data program (Carter et al., 2015). It should therefore be prevented that the focus on compliance with the law completely overshadows the protection of other relevant rights and interests and values, just because they are not that well operationalised in the legal system.

This study has some limitations. First of all, only two cases were analysed and a limited number of interviews took place, so that the findings do not necessarily represent the most common or relevant privacy and data protection issues. However, we do feel that within these two cases a sufficiently complete image was extracted due to the fact that most of the individuals who actually dealt with relevant issues were interviewed and a broad range of available documents were studied.

5. Concluding remarks

Overall, the findings show that it takes considerable effort to take privacy and data protection norms into account in a data-intensive health research initiative, especially when individual level data need to be linked or enriched. By embracing the complexity of the law in the initiative in an early phase, setbacks can be prevented, the existing flexibility within the law can be utilised where appropriate, and systems or organisations can be designed and developed so that they take relevant rules into account. The cases and discussion illustrate that a close collaboration of experts with different backgrounds within the initiative may be necessary to be able to successfully navigate this process.

Conflict of interest

Declarations of interest: none.

Acknowledgements

The authors thank the respondents and the management of YOUth and Big Data Psychiatry for their participation and cooperation.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

References

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, *13*, 544–559.

BBMRI-ERIC (2017). A Code of Conduct for Health Research. http://code-of-conduct-for-health-research.eu Accessed 18 January 2018.

Budin-Ljøsne, I., Burton, P., Isaeva, J., Gaye, A., Turner, A., Murtagh, M. J., ... Harris, J. R. (2015). DataSHIELD: An ethically robust solution to multiple-site individual-level data analysis. *Public Health Genomics*, *18*, 87–96.

Bygrave, L. A. (2017). Data protection by design and by default: Deciphering the EU's legislative requirements. *Oslo Law Review*, *4*, 105–120.

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: Why care.Data ran into trouble. *Journal of Medical Ethics*, *41*, 404–409.

Cavoukian, A. (2010). Privacy by design: The definitive workshop. A foreword by Ann Cavoukian, Ph.D. *Identity in the Information Society*, *3*, 247–251.

Koops, B. J. (2011). The evolution of privacy law and policy in the Netherlands. *Journal of Comparative Policy Analysis*, *13*, 165–179.

Koops, B. J., & Leenes, R. E. (2014). Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law. *International Review of Law, Computers & Technology*, *28*, 159–171.

Mayer-Schönberger, V., & Ingelsson, E. (2018). Big data and medicine: A big deal? *Journal of Internal Medicine*, *283*, 418–429.

Menger, V., Spruit, M., Hagoort, K., Scheepers, F. E. (2016). Transitioning to a data driven mental health practice: Collaborative expert sessions for knowledge and hypothesis finding. Computational and Mathematical Methods in Medicine, 2016, 9089321.

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, *22*, 303–341.

Monteith, S., Glenn, T., Geddes, J., & Bauer, M. (2015). Big data are coming to psychiatry: A general introduction. *Int J Bipolar Disord*, *3*, 21.

Mostert, M., Bredenoord, A. L., Biesaart, M. C. I. H., & Van Delden, J. J. M. (2016). Big data in medical research and EU data protection law: Challenges to the consent or anonymise approach. *European Journal of Human Genetics*, *24*, 956–960.

Ploem, M. C., Essink-Bot, M. L., & Stronks, K. (2013). Proposed EU data protection regulation is a threat to medical research. *BMJ*, *346*, f3534.

Scheepers, F. E., Menger, V., & Hagoort, K. (2018). Data science in psychiatry. *Tijdschrift voor Psychiatrie*, *60*, 205–209.

Sethi, N. (2015). Reimagining regulatory approaches: On the essential role of principles in health research. *SCRIPTed*, *12*, 91–116.

Sethi, N., & Laurie, G. T. (2013). Delivering proportionate governance in the era of eHealth. *Med Law Int*, *13*, 168–204.

Wallace, S. E., Gaye, A., Shoush, O., & Burton, P. R. (2014). Protecting personal data in epidemiological research: DataSHIELD and UK law. *Public Health Genomics*, *17*, 149–157.

Wolfson, M., Wallace, S. E., Masca, N., Rowe, G., Sheehan, N. A., Ferretti, V., ... Burton, P. R. (2010). DataSHIELD: Resolving a conflict in contemporary bioscience—Performing a pooled analysis of individual-level data without sharing the data. *International Journal of Epidemiology*, *39*, 1372–1382.

Yin, R. K. (2014). Getting started: How to know whether and when to use the case study as a research method. In R. K. Yin (Ed.), *Case study research: Design and methods* (pp. 3–25). CA, USA: Thousand Oaks.