Original Manuscript

Privacy of Clinical Research Subjects: An Integrative Literature Review

Journal of Empirical Research on Human Research Ethics
1–16

DOI: 10.1177/1556264618805643
journals.sagepub.com/home/jre
SAGE

Sanna-Maria Nurmi[1], Mari Kangasniemi[1], Arja Halkoaho[2,3], and Anna-Maija Pietilä[1,4]

Abstract

With changes in clinical research practice, the importance of a study-subject's privacy and the confidentiality of their personal data is growing. However, the body of research is fragmented, and a synthesis of work in this area is lacking. Accordingly, an integrative review was performed, guided by Whittemore and Knafl's work. Data from PubMed, Scopus, and CINAHL searches from January 2012 to February 2017 were analyzed via the constant comparison method. From 16 empirical and theoretical studies, six topical aspects were identified: the evolving nature of health data in clinical research, sharing of health data, the challenges of anonymizing data, collaboration among stakeholders, the complexity of regulation, and ethics-related tension between social benefits and privacy. Study subjects' privacy is an increasingly important ethics principle for clinical research, and privacy protection is rendered even more challenging by changing research practice.

Keywords

confidentiality, clinical research, integrative literature review, privacy, research ethics

Introduction

Privacy is among the core principles of ethics in clinical research (Beauchamp & Childress, 2013; Emanuel & Wendler, 2008), and it has only gained importance with the changing nature of research practice (Brall, Maeckelberghe, Porz, Makhoul, & Schröder-Bäck, 2017). The central changes are related to the growing complexity of clinical research (Brall et al., 2017; Scott, McConnell, Lewis, & Lewis, 2012), the increased involvement of various stakeholders in the research process (Brall et al., 2017; Nurmi, Halkoaho, Kangasniemi, & Pietilä, 2017), the more pronounced role of information technology, and the increased use of electronic health data for research purposes (Bélanger & Crossler, 2011; Cohen, 2013; Hoffman & Podgurski, 2012; Stiles & Petrila, 2011). These factors pose risks to maintaining study-subjects' protection and raise new kinds of privacy threats.

Respect for the privacy of research participants in clinical studies is internationally acknowledged in standards of ethics (Hodge & Gostin, 2008). Although privacy is generally recognized as a basic human right (United Nations, 1948), it is considered a highly complex concept by its very nature (Hodge & Gostin, 2008; Kaye, 2012; Nissenbaum, 2004; Solove, 2006). According to Beauchamp and Childress (2013), privacy refers to individuals' right to be free from intrusion or interference by others and is closely connected with the principles of autonomy and human dignity. Beyond that privacy has been described in terms of several perspectives and dimensions (Beauchamp & Childress, 2013; Kaye, 2012; Leino-Kilpi et al., 2001). For example, Beauchamp and Childress (2013) have characterized it as encompassing the following forms: informational, physical, decisional, proprietary, and associational privacy. In a clinical-research context, the focus is typically on informational privacy (Bélanger & Crossler, 2011; Hodge & Gostin, 2008)—that is, individuals' right to control their personally identifying health information, which can be highly personal and intimate (Hodge & Gostin, 2008). The issue extends further, however. Solove (2006) has presented a taxonomy of privacy to encourage a more coherent and comprehensive understanding of the concept. He identified four main groups of activities, each with various subgroups: (a) information collection, (b) information processing, (c) information dissemination, and (d) invasion.

[1]University of Eastern Finland, Kuopio, Finland
[2]Kuopio University Hospital, Finland
[3]Tampere University of Applied Sciences, Finland
[4]Social and Health Care Services, Kuopio, Finland

Corresponding Author:
Sanna-Maria Nurmi, Department of Nursing Science, University of Eastern Finland, Kuopio Campus, P.O. Box 1627, 70211 Kuopio, Finland.
Email: sannu@student.uef.fi

Confidentiality in a clinical-research context refers to the researchers' corresponding duty to protect study-subjects' right to privacy (Hodge & Gostin, 2008). This duty comprises those legal and ethical obligations that arise through specific relationships between researchers and human subjects (Stiles & Petrila, 2011). Confidentiality is directly related to the collection, use, and storage of personal data (Elliot, Mackey, O'Hara, & Tudor, 2016). According to Emanuel and Wendler (2008), protecting confidentiality in clinical research is an ongoing process that includes securing databases, locking filing cabinets, coding specimens and data forms, and interviewing participants in private spaces where they cannot be overheard. However, this process can never bring absolute confidence (Stiles & Petrila, 2011): Protecting confidentiality in research practice is a challenging task fraught with essential trade offs (Emanuel & Wendler, 2008; Hodge & Gostin, 2008; Stiles & Petrila, 2011). One key aspect of these compromises is that careless handling of identifying health data can cause direct and indirect harm to human subjects, create concerns that their health data will be misused, and discourage them from participating in future research (Hodge & Gostin, 2008).

The importance of privacy and confidentiality in clinical research is emphasized in national and international ethics codes and legal requirements (Holm, 2016; Windows Media Audio, 2013), especially in light of changes occurring in clinical research practice (Brall et al., 2017; Cohen, 2013; Stiles & Petrila, 2011). Although much attention has been paid to this topic in the literature, the research landscape is fragmented, and a synthesis addressing today's topical aspects of privacy and confidentiality in clinical research is lacking. Therefore, we carried out an integrative review to assess topical aspects of privacy and confidentiality in clinical research.

Method

The integrative review was guided by Whittemore and Knafl's (2005) methods. Our review process consisted of five stages: identifying the research problem, searching the literature, evaluating the data, analyzing the data, and presenting the synthesis of results.

Identification of the Research Problem

First, we conducted preliminary searches to understand prior literature and identify the research problem. We found that the existing knowledge was fragmented, that empirical studies accounted for a small proportion of the work and varied greatly in their methods, and that theoretical papers dealt with important aspects of the phenomenon. These findings supported our decision to conduct an integrative review, aimed at identifying and bringing together empirical and theoretical studies applying different methods (Soares et al., 2014; Whittemore & Knafl, 2005). This approach also enabled us to synthesize various perspectives into a systematic knowledge base and identify current topical elements of the subject studied (Whittemore & Knafl, 2005).

The Literature Search

The second stage was an electronic search for scientific peer-reviewed studies in English-language publications from January 2012 to February 2017. The search, using PubMed, Scopus, and CINAHL (see Figure 1), was designed by the research team and a university's specialist research librarian. The search terms chosen were based on previous literature and preliminary searches. The search terms were "clinical research," "clinical trials," "privacy," and "confidentiality."

We conducted the selection in stages, applying the inclusion criteria of the papers being peer-reviewed empirical and theoretical works with a title, abstract, or aim statement that indicated the article related to privacy and confidentiality in clinical research wherein the subjects are competent adults able to give informed consent. We excluded papers that focused on biobank research, because of the broad consent involved, wherein subjects may agree to a wide range of future research without any individual research project being specified (Hofmann, 2009). In addition, studies with subjects below 18 years of age were excluded from consideration.

Two of the authors (S.-M.N. and A.H.) screened 1,089 papers, narrowing the sample to 70 papers on the basis of the title, then 37 by reading the abstract, and finally 16 on the basis of the full text (see Figure 1). This selection was agreed upon by all of the authors.

Data Evaluation

The third stage was evaluation of the 16 papers selected. As is typical with integrative review methods, detail-level criteria were not appropriate for the appraisal of quality, because of the diversity of the primary sources represented (Hopia, Latvala, & Liimatainen, 2016; Whittemore & Knafl, 2005). Therefore, to examine the quality of these papers, we used the six descriptive criteria for methodological structure presented in an integrative review by Kangasniemi, Pakkanen, and Korhonen (2015). Each element was evaluated on a three-point scale with the items "yes," "poor," and "not reported" (see Table 1). No papers were excluded from consideration on the basis of this evaluation.

Characteristics of the Sample

Of the 16 papers, nine were of an empirical nature and seven were theoretical. Five of the empirical studies were quantitative (Grande et al., 2015; Martínez, Sánchez, & Valls, 2013;

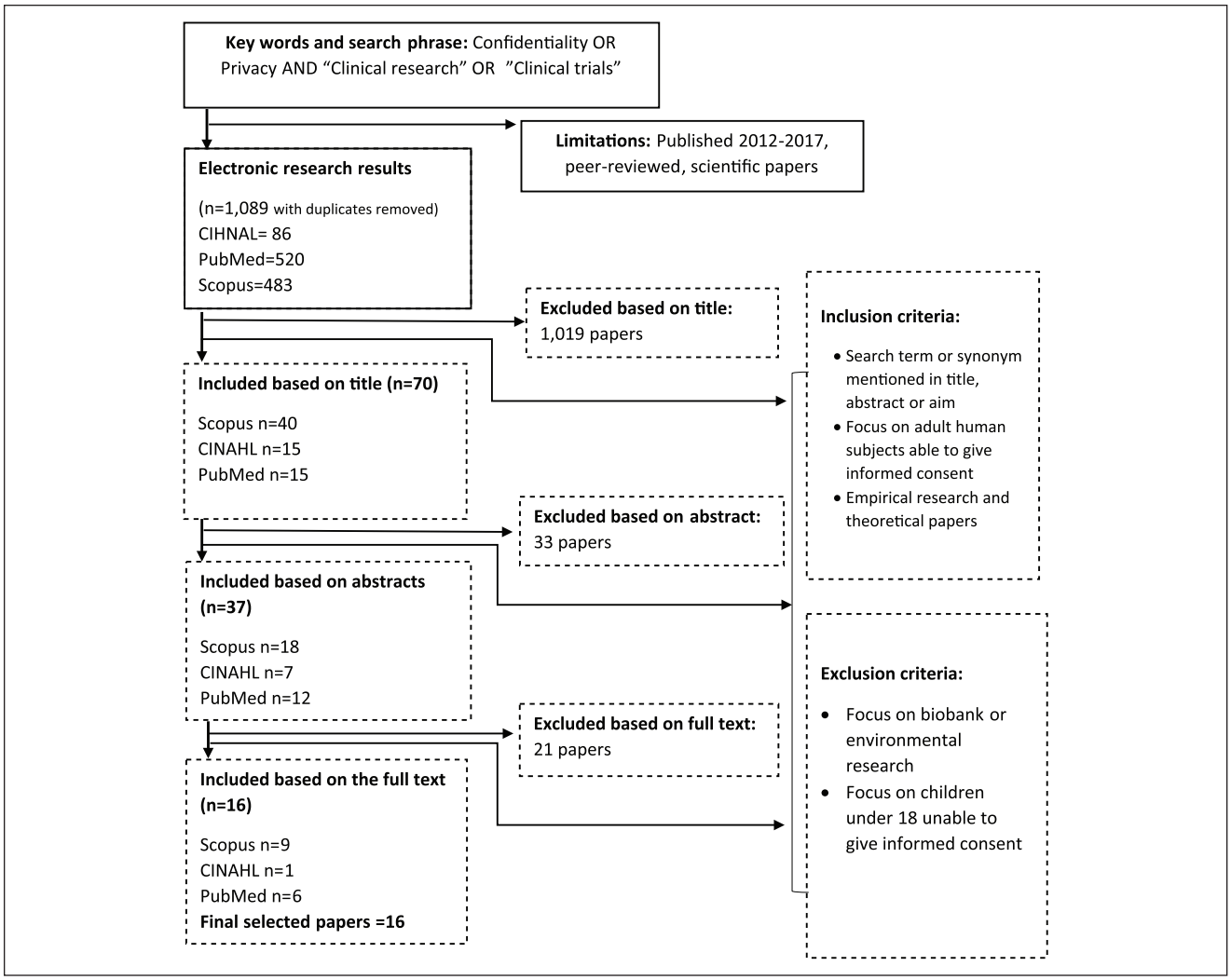

Figure 1. Flow diagram of literature searches and data selection.

Noh, Chun, & Jeong, 2014; Rho, Jang, Chung, & Choi, 2015; Yuan et al., 2017), and four were qualitative (Frizzo-Barker, Chow-White, Charters, & Ha, 2016; Halkoaho, Pietilä, Vesalainen, & Vähäkangas, 2012; Kuchinke et al., 2014; Nair & Ibrahim, 2015). Four studies examined were conducted in the United States, two each in the United Kingdom and South Korea, and one each in Australia, Belgium, Canada, Finland, Germany, India, Spain, and the United Arab Emirates.

Analysis of the Data

The fourth stage consisted of analysis of the data by means of the constant comparison method, which is suitable for varied data, arising from multiple methodologies (Whittemore & Knafl, 2005). The five-step analysis process comprised data reduction, data display, data comparison, drawing of conclusions, and verification (Whittemore & Knafl, 2005), with the first two steps being conducted simultaneously. All 16 papers were read several times for a holistic understanding of the data set, and then tabulation and extraction were performed to produce publication information (authorship, year and country of publication, and methods; see Table 1). Next, the content providing information on the purpose of the study was identified, coded with a term that described the data, and entered in an extensive matrix to allow a systematic view of the body of literature and methodical comparison of the primary sources (Whittemore & Knafl, 2005). The third step employed an iterative process of comparison in which similar codes were grouped together to form categories. As categories emerged, comparison was performed again, across categories, to reveal patterns throughout the data set. Comparison was conducted at both the level of each paper or piece of data and also between papers, to support development of our final understanding of the meaning of

Table 1. Description of the 16 Selected Papers.

Authorship, year, and country	Purpose	Design and tools	Sample size and characteristics	Quality appraisal criteria ("y" = yes, "p" = poor, "nr" = not reported)
Anuradha (2013), India	To propose a framework that uses the anonymization technique for protecting patient privacy.	Theoretical work (discussion)	Literature, 12 references, published in 2004-2012.	(nr) Aims and objectives clearly described (nr) Study design adequately described (nr) Research methods appropriate (nr) Theoretical framework explicit (nr) Limitations presented (p) Implications discussed
Coppieters and Levêque (2013), Belgium	To give an overview of the main privacy and data-protection issues that researchers need to take into account when working with health data.	Theoretical work (commentary)	Literature, 13 references, published in 1997-2011.	(y) Aims and objectives clearly described (y) Study design adequately described (nr) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed
Frizzo-Barker, Chow-White, Charters, and Ha (2016), Canada	To analyze emerging sociotechnical and privacy issues related to a collaborative clinical-research project in genomics.	Qualitative design	Clinical genome researchers and policy officials ($n = 43$).	(y) Aims and objectives clearly described (p) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (p) Limitations presented (y) Implications discussed
Grande et al. (2015), the United States	To determine whether individuals with a history of cancer are less willing to share their health information and whether they give the sensitivity of health information greater weight than people without a history of cancer do.	Quantitative design (survey)	Randomly assigned nationally representative participants ($n = 2,945$) with ($n = 187$) and without ($n = 2,789$) a history of cancer.	(y) Aims and objectives clearly described (y) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (y) Limitations presented (y) Implications discussed
Halkoaho, Pietilä, Vesalainen, and Vähäkangas (2012), Finland	To evaluate the ethics statements of principal investigators.	Qualitative design	A total of 56 applications/ studies.	(y) Aims and objectives clearly described (y) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (p) Limitations presented (y) Implications discussed
Kuchinke et al. (2014), Germany	To develop a model describing core concepts and principles of data flow, data privacy, and confidentiality in primary-care clinical research.	Model-based approach	Exploration of European Union (EU) legal requirements for data protection and privacy-data access policies, alongside existing privacy frameworks for research projects.	(y) Aims and objectives clearly described (y) Study design adequately described (y) Research methods appropriate (p) Theoretical framework explicit (y) Limitations presented (y) Implications discussed

(continued)

Table 1. (continued)

Authorship, year, and country	Purpose	Design and tools	Sample size and characteristics	Quality appraisal criteria ("y" = yes, "p" = poor, "nr" = not reported)
De Lusignan et al. (2015), the United Kingdom	To create a comprehensive framework for determining the ethics and privacy status of a project and for providing guidance on data access.	Theoretical work	Literature, 20 references, published in 1999-2015.	(y) Aims and objectives clearly described (nr) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed
Martínez, Sánchez, and Valls (2013), Spain	To propose a general framework that enables the anonymization of structured nonnumerical medical data from a semantic perspective.	Statistical methods	A real-world medical data set with structured clinical outcomes and SNOMED CT as the medical knowledge base.	(y) Aims and objectives clearly described (y) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed
McGraw et al. (2015), the United States	To explore both the ethical foundation and regulatory framework intended to protect privacy in pragmatic clinical trials and to review examples of novel approaches to respecting people involved in research that may have the added benefit of respecting patient privacy considerations.	Theoretical work	Literature, 55 references, published in 1970-2015.	(y) Aims and objectives clearly described (nr) Study design adequately described (nr) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed
Nair and Ibrahim (2015), United Arab Emirates	To explore factors influencing clinical trials' privacy and confidentiality in the United Arab Emirates.	Qualitative design, with review of clinical trials' study protocols	A review of data preservation for clinical trials in comparison with health-record preservation regulations in the United Arab Emirates and other countries.	(y) Aims and objectives clearly described (nr) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed
Noh, Chun, and Jeong (2014), Republic of Korea	To propose a controlled secure aggregation protocol guaranteeing both privacy and accuracy when researchers outsource their clinical research data for sharing.	Statistical methods	A controlled secure aggregation protocol for sharing of clinical-research data, to balance interests between hospitals and researchers.	(y) Aims and objectives clearly described (y) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed

(continued)

Table 1. (continued)

Authorship, year, and country	Purpose	Design and tools	Sample size and characteristics	Quality appraisal criteria ("y" = yes, "p" = poor, "nr" = not reported)
O'Keefe and Rubin (2015) Australia	To provide a review of several approaches to reducing disclosure risk when making data available for research and to present a taxonomy for such approaches.	Theoretical work	Literature, 127 references, published in 1967-2015.	(y) Aims and objectives clearly described (nr) Study design adequately described (nr) Research methods appropriate (y) Theoretical framework explicit (y) Limitations presented (y) Implications discussed
Rho, Jang, Chung, and Choi (2015), Republic of Korea	To ascertain the effect of the various types of knowledge, attitudes, and levels of trust related to personal health information on that information's use for clinical research.	Quantitative design (surveys)	267 participants: a clinical researcher group ($n = 113$), nonclinical researcher group ($n = 72$), and patient group ($n = 82$).	(y) Aims and objectives clearly described (y) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (y) Limitations presented (y) Implications discussed
Tucker et al. (2016), the United Kingdom	To briefly describe existing legislation, guidance, and common practices and then suggest best practices relevant for protecting patient privacy when sharing clinical-trial data.	Theoretical work	Literature, 34 references, published in 2008-2016.	(y) Aims and objectives clearly described (y) Study design adequately described (nr) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed
Wolf et al. (2015), the United States	To examine Certificates of Confidentiality and related statutory protections for greater understanding and suggest ways to strengthen the certificates' protections and to describe researchers' obligations to protect the confidentiality of data they collect.	Theoretical work	Literature, 140 references, published in 1966-2015.	(y) Aims and objectives clearly described (y) Study design adequately described (nr) Research methods appropriate (y) Theoretical framework explicit (nr) Limitations presented (y) Implications discussed
Yuan et al. (2017), the United States	To determine whether searchable encryption can instill privacy in clinical-research networks without sacrificing their usability.	Quantitative design	A technique implemented in working software to enable privacy-preserving cohort discovery (PPCD) services in large distributed clinical research- networks on the basis of elliptic curve cryptography (ECC).	(y) Aims and objectives clearly described (y) Study design adequately described (y) Research methods appropriate (y) Theoretical framework explicit (y) Limitations presented (y) Implications discussed

the data. In the conclusion and verification steps, we identified patterns and relationships that extended beyond the description to higher levels of abstraction (Whittemore & Knafl, 2005). The research group worked collaboratively to complete the data comparison, conclusion, and verification stages.

Results

The Evolving Nature of Health Data in Clinical Research

Privacy of human subjects and confidentiality of their personal data have become one of the central ethical aspects of clinical research, due to the increased use of electronic health data in clinical research (Anuradha, 2012; Coppieters & Levêque, 2013; Frizzo-Barker et al., 2016; Martínez et al., 2013; Meystre et al., 2014; Nicholls et al., 2016; O'Keefe & Rubin, 2015; Rho et al., 2015; Tucker et al., 2016). In greater and greater quantities, health data from sources such as medical and scientific documents, molecular research (De Lusignan, Liyanage, Di Iorio, Chan, & Liaw, 2016; Frizzo-Barker et al., 2016; O'Keefe & Rubin, 2015), and genomic studies (Anuradha, 2012; Meystre et al., 2014) are being collected and stored in electronic form in health care and clinical-research practice (Grande et al., 2015; Meystre et al., 2014; O'Keefe & Rubin, 2015). The widespread availability of electronic health data has created new opportunities for secondary use of said data for clinical-research purposes. In addition, there are mounting calls for allowing secondary use of clinical-research data in future studies (Tucker et al., 2016).

Another development is the changing nature of the data at issue, with increased availability of big data (Frizzo-Barker et al., 2016), freely available data from open-access databases (Frizzo-Barker et al., 2016; Noh et al., 2014; O'Keefe & Rubin, 2015), integration of material from various data sources (De Lusignan et al., 2016; O'Keefe & Rubin, 2015; Rho et al., 2015; Yuan et al., 2017), and the rapid movement of genome science from research labs and biobanks to clinical settings (Frizzo-Barker et al., 2016). These forms of health data serve as valuable resources for clinical research and create new opportunities, yet they also bring new challenges to protecting the privacy of each study subject and the confidentiality of his or her personal data (Anuradha, 2012; Martínez et al., 2013; Meystre et al., 2014; O'Keefe & Rubin, 2015; Rho et al., 2015; Yuan et al., 2017). The highly multifaceted body of health data generated today is complicated to understand, process, store, and analyze, rendering it more difficult than ever for researchers and clinic personnel to protect subjects' privacy and the confidentiality of their health data (Frizzo-Barker et al., 2016; Kuchinke et al., 2014; Rho et al., 2015; Wolf et al., 2015).

Sharing of Health Data

The papers in our data set acknowledge the value of shared data but also identify concerns about data sharing (Noh et al., 2014; Tucker et al., 2016). Health data in emerging clinical research have been shared at several levels: within organizations (O'Keefe & Rubin, 2015) and between them (O'Keefe & Rubin, 2015; Tucker et al., 2016; Yuan et al., 2017), often across national borders (De Lusignan et al., 2016; Nair & Ibrahim, 2015; Rho et al., 2015; Tucker et al., 2016; Yuan et al., 2017). Also, numerous stakeholders, differing in background, are involved in implementing research (De Lusignan et al., 2016; Frizzo-Barker et al., 2016), and this multiactor environment sets new challenges for protection of privacy (De Lusignan et al., 2016; Frizzo-Barker et al., 2016; Kuchinke et al., 2014; Tucker et al., 2016). Among these are challenges of sharing data across different technical infrastructures (O'Keefe & Rubin, 2015; Yuan et al., 2017) and in research projects wherein various stakeholders participate in condensing, processing, and integrating health data. Also, an issue is created when the handling and storage of data involve many systems, often very different, without any control by the entity that originally submitted the data (Kuchinke et al., 2014).

Data-holders are responsible for implementing all sharing of data with third parties in an ethical manner. They are required to specify a transparent process for handling and sharing the data, execute that process appropriately (Frizzo-Barker et al., 2016; Noh et al., 2014; Tucker et al., 2016; Wolf et al., 2015), and carry out control and surveillance procedures to make sure that the process complies with legal requirements (Coppieters & Levêque, 2013; De Lusignan et al., 2016; Nair & Ibrahim, 2015; Tucker et al., 2016). In addition, holders of data are responsible for assessing the risk of individuals being identified, for sharing data in a secure and controlled manner, and for employing strategies to minimize the risk of loss of individuals' privacy. Studies have shown that data-holders take on other duties too, providing support to help researchers navigate and understand the data used in their research (Tucker et al., 2016) and the legal requirements related to data use (Wolf et al., 2015).

The papers selected address health data being shared via open-access databases that are freely available to anyone (Frizzo-Barker et al., 2016; Noh et al., 2014). Making data available on the Internet, with no legal agreements in place, greatly increases the number of people with access to said data (Tucker et al., 2016). These databases bring new opportunities and cost savings for clinical research (Noh et al., 2014; O'Keefe & Rubin, 2015); however, they also raise ethics questions, such as where responsibility for the stewardship and governance of data ultimately lies, how privacy is to be protected with regard to these freely available databases, and what happens if corporations access the material

for purposes not originally envisioned (O'Keefe & Rubin, 2015).

Although the public has become more comfortable with sharing highly personal information online (e.g., via social media; Frizzo-Barker et al., 2016; Tucker et al., 2016), concerns have been expressed that individuals often lack even a basic understanding of what is considered personal health information (Rho et al., 2015) and of the risks and benefits of sharing their information on the Internet (Frizzo-Barker et al., 2016; Rho et al., 2015). The papers point to growing concerns that the proliferation of publicly available information online has made it possible to reconnect anonymized data to the subjects via data-linkage techniques (Kuchinke et al., 2014; O'Keefe & Rubin, 2015; Rho et al., 2015; Tucker et al., 2016)—that is, that one can trace items from "scrubbed data" back to the individual by combining material from separate data sets (O'Keefe & Rubin, 2015; Rho et al., 2015; Tucker et al., 2016). Unadvised sharing of information is problematic because of such growing possibilities of combining and linking data (data matching). For example, a combination of one's health interest, age, and name can be used to triangulate data, leading to a higher probability of compromises to the subject's identity and privacy (Anuradha, 2012; Frizzo-Barker et al., 2016; Rho et al., 2015).

The Challenges of Data Anonymization

Anonymization of data is a generally effective method of protecting study-subject privacy, and the literature recognizes its increased importance in the era of large data sets and increased data sharing (Martínez et al., 2013; Tucker et al., 2016; Wolf et al., 2015). The papers point out that the definitions of anonymization and de-identification are not always clear (Halkoaho et al., 2012; O'Keefe & Rubin, 2015; Tucker et al., 2016). Also, the difference between the two is not clear in some cases, and the literature often uses the corresponding terms interchangeably, in several contexts (Tucker et al., 2016).

Data anonymization is a technically complex task in which true anonymity demands personnel with an understanding of data management and basic statistics (Meystre et al., 2014; Tucker et al., 2016). Several approaches and methods of anonymization have been reported, among them being generalization-based and suppression-based methods (Anuradha, 2012). These are designed to maximize the amount of information provided while minimizing the risk of re-identification of individual subjects (Meystre et al., 2014). The techniques needed vary, because, for example, medical data can be presented as unstructured textual documents or in structured patient records that compile values for a set of normalized attributes (e.g., symptoms, diagnosis, and treatment; Martínez et al., 2013). Most of the anonymization methods in use were developed to deal with numerical data (Anuradha, 2012; Martínez et al., 2013), but electronic health data typically include many nonnumerical textual items also, such as diagnoses, symptoms, and free-form description/answers. Therefore, there is a need for tackling the anonymization of textual values too (Martínez et al., 2013). Addressing this, Martínez and colleagues (2013) have presented a semantic framework to protect the privacy of electronic health records with nonnumerical attributes.

Data anonymization can never be considered absolute or guaranteed, and some risk of re-identification of the individual always exists (Martínez et al., 2013; Tucker et al., 2016). Re-identification could occur through, for example, statistical matching of the attributes remaining (Martínez et al., 2013). That is, removing identifying attributes, such as a social security number, is not enough to keep the data anonymous, since some unique combinations of other elements, such as rare diagnoses and personalized treatments, could lead to an individual's identity being disclosed (Martínez et al., 2013; Noh et al., 2014; O'Keefe & Rubin, 2015; Tucker et al., 2016). The variability and high dimensionality of medical data render the appearance of identifying combinations of attribute values especially likely, and these could well enable disclosure when evaluated together (Martínez et al., 2013; Tucker et al., 2016).

At the other end of the spectrum, some concern has been expressed that anonymization methods may harm the data or reduce the material's utility for research (Meystre et al., 2014). Although Meystre et al. (2014) argued that the original data should be anonymized such that the risk of a subject's re-identification is reduced to a minimum, high levels of anonymization may result in data that cannot be used to answer scientific questions or that lead to misleading interpretations (Tucker et al., 2016). Accordingly, the papers reviewed advocate anonymized data sets that allow for as much high-quality information as possible, to be as useful as possible for research purposes and thereby enable reliable results and maximum benefit (Meystre et al., 2014; Tucker et al., 2016).

Interstakeholders' Collaboration as a Prerequisite

The authors of the selected papers found that the collaboration and involvement of many research stakeholders is a prerequisite for protection of study-subjects' privacy in clinical research (De Lusignan et al., 2016; Frizzo-Barker et al., 2016; Halkoaho et al., 2012; Kuchinke et al., 2014). Among the central stakeholders cited are researchers (De Lusignan et al., 2016; Wolf et al., 2015), patients, the clinic staff, ethics committees, privacy and ethics experts, data controllers and custodians, and the public. Although these stakeholders are presented as having various roles and responsibilities in protecting privacy and confidentiality in clinical-research processes (De Lusignan et al., 2016), researchers' ethical and legal obligation to protect study-subject privacy and

confidentiality of personal data is highlighted in particular (Wolf et al., 2015). The papers describe well-functioning collaboration among researchers, the data-holders who give approval for data access, and ethics committees as critical with regard to obtaining ethics approval and data-sharing agreements (De Lusignan et al., 2016; Tucker et al., 2016). These are presented as vital for guaranteeing that subjects' privacy and the confidentiality of the health data are considered during a study's planning and implementation phase and that the legal requirements are met (Coppieters & Levêque, 2013; De Lusignan et al., 2016; Halkoaho et al., 2012). Moreover, collaboration between researchers and the clinical staff actually implementing the research was seen as a prerequisite for privacy protection in research practice, though rendered challenging by differences in background, organizational culture, and goals among the various stakeholders (Frizzo-Barker et al., 2016).

The Complexity of Regulation

The papers described international- and national-level variation in privacy-protection regulations (Coppieters & Levêque, 2013; Kuchinke et al., 2014; McGraw et al., 2015; Nair & Ibrahim, 2015; O'Keefe & Rubin, 2015; Tucker et al., 2016). For instance, they referred to differences between regulation in the European Union (Coppieters & Levêque, 2013; O'Keefe & Rubin, 2015; Tucker et al., 2016) and the United States (McGraw et al., 2015; O'Keefe & Rubin, 2015; Tucker et al., 2016). Differences were found also within the European Union, related to national-level implementation of the Data Protection Directive depending on local interpretations of how to define anonymity, access to data, exemptions, and so on (Coppieters & Levêque, 2013; Kuchinke et al., 2014).

For this reason, access to research data has been hampered by a fragmented legal framework in Europe, nonuniform interpretation of regulations, changes in the guidance provided, and lack of clarity (among researchers, regulators, patients, and the general public alike; Coppieters & Levêque, 2013; Kuchinke et al., 2014). A similar situation has long existed in the United States, where interpretation of the Health Insurance Portability and Accountability Act of 1996 privacy rules has not been uniform, and integration with other federal regulations is difficult (Kuchinke et al., 2014; McGraw et al., 2015).

Current practice for protection of study-subject privacy in clinical research was seen as relying heavily on policies and guidelines, which were deemed incapable of covering all scenarios in emerging research practices (Yuan et al., 2017). Therefore, Coppieters and Levêque (2013) have emphasized the importance of ethical reflection during research projects. At the same time, current data-protection regulation has been regarded as complex, voluminous (O'Keefe & Rubin, 2015), incoherent (De Lusignan et al., 2016), country-specific (Coppieters & Levêque, 2013; Kuchinke et al., 2014; Tucker et al., 2016), and restricted to specific situations such as research contexts wherein the patient has provided informed consent (Coppieters & Levêque, 2013; Kuchinke et al., 2014). In addition, privacy legislation was presented as slow and too rigid to keep up with the expanding role of health data, rapidly changing technology, and the increasing availability of information (Kuchinke et al., 2014; Wolf et al., 2015). Furthermore, concerns were expressed that members of ethics committees (Wolf et al., 2015), health care professionals (Coppieters & Levêque, 2013), and researchers (Rho et al., 2015) have a limited understanding of the regulations and legal frameworks related to privacy and how they should be implemented. Confusion often surrounds privacy regulations addressing the use of personal health data in clinical research (Coppieters & Levêque, 2013), and misunderstanding in this respect was identified as present even among health care professionals and researchers (Coppieters & Levêque, 2013; Halkoaho et al., 2012; Rho et al., 2015).

These challenges point to the critical question of how researchers have been able to plan and implement clinical research in line with regulations (Coppieters & Levêque, 2013; De Lusignan et al., 2016; Halkoaho et al., 2012; Tucker et al., 2016) so as to account for research-specific aspects of privacy and sound implementation in research practice (Coppieters & Levêque, 2013; Halkoaho et al., 2012). This was seen as challenging, especially in international research projects involving stakeholders working across national borders (Tucker et al., 2016).

The Tension Between Social Benefits and Privacy

The central ethics question considered in the papers selected was the balance between social benefits and protection of individuals' privacy (Coppieters & Levêque, 2013; Frizzo-Barker et al., 2016; Grande et al., 2015; Kuchinke et al., 2014; McGraw et al., 2015; Meystre et al., 2014; Nair & Ibrahim, 2015; O'Keefe & Rubin, 2015; Tucker et al., 2016). In practice, it was considered important to allow the effective use of health data in clinical research for the benefit of society while still protecting the privacy and confidentiality of individuals (De Lusignan et al., 2016; Kuchinke et al., 2014; McGraw et al., 2015; Meystre et al., 2014; Nair & Ibrahim, 2015; O'Keefe & Rubin, 2015; Tucker et al., 2016). This balance was presented as a trade off, in that methods that reduce disclosure risk could also reduce the data's utility (O'Keefe & Rubin, 2015).

In the papers examined, achieving balance between beneficial uses of electronic health data and the protection of individuals' privacy was presented as requiring reexamination of ethics frameworks, more flexible privacy rules, and innovative approaches to protecting privacy and confidentiality (McGraw et al., 2015; Tucker et al., 2016). These papers

characterized current regulatory frameworks as supporting strong respect for individuals' autonomy and privacy (Coppieters & Levêque, 2013; Grande et al., 2015; McGraw et al., 2015). For instance, Grande et al. (2015) argued that the regulations are not designed to strike a balance between an individual's right to privacy and the collective social interest of research. They opined that policy initiatives ought to consider the possibility of discrepancies between a subject's altruistic desire to have his or her information used for the greater social benefit and the importance of his or her privacy being considered and protected. Finally, several papers identified public trust as necessary for promotion of clinical research, stating that this trust can be maintained via such means as making sure that health data are used in an ethically appropriate manner (De Lusignan et al., 2016; Martínez et al., 2013).

Discussion

Two topical aspects of privacy emerged as receiving emphasis in current clinical research. First, we will discuss how privacy can be safeguarded as clinical-research practices change. Second, we will focus on the balance between the benefits to society and individual-level privacy.

How Privacy Can Be Safeguarded in a World of Changing Clinical Research Practice

The privacy of human subjects has become one of the crucial considerations in clinical research because of changes in research practice. Some of the central changes are related to the evolving nature of health data in clinical research—in particular, increased collection and use of electronic health data, the phenomenon of big data, and greater focus on data sharing (Anuradha, 2012; Coppieters & Levêque, 2013; De Lusignan et al., 2016; Frizzo-Barker et al., 2016; Martínez et al., 2013; Meystre et al., 2014; O'Keefe & Rubin, 2015; Rho et al., 2015; Tucker et al., 2016). These changes create new opportunities for clinical research, but at the same time, they can threaten the privacy of individuals and the confidentiality of their personal health data. Privacy breaches resulting from re-identification could lead to harm ranging from individuals' social embarrassment and shame to stigmatization and even damage to their social and economic status, such as loss of employment and health insurance (McGraw et al., 2015; O'Keefe & Rubin, 2015). They also are likely to undermine public trust in the robustness of the data-protection measures adopted by research institutions (Vayena & Blasimme, 2017).

Our results identify an awareness that the increased use of information technology and its capacity to collect, analyze, and disseminate information on individuals has brought about new kinds of privacy threats (Anuradha, 2012; Coppieters & Levêque, 2013; De Lusignan et al., 2016; Frizzo-Barker et al., 2016). At the same time, the amount of medical information generated from each individual is increasing (Porsdam Mann, Savulescu, & Sahakian, 2016). In parallel with this development, individuals are increasingly sharing highly personal information online, which could lead to a greater risk of loss of control of their data (Frizzo-Barker et al., 2016; Tucker et al., 2016). This issue is compounded by the fact that some may not understand what personal information is (Rho et al., 2015), how their data from various sources can be considered in combination and linked, and the risks and benefits of sharing health data on the Internet (Frizzo-Barker et al., 2016). In addition, individuals are often poorly informed about the use of their health data for secondary purposes (Frizzo-Barker et al., 2016; Grande et al., 2015). Studies have shown low levels of personal awareness about the practices employed and about how the electronic health data are used (Aitken, de St Jorre, Pagliari, Jepson, & Cunningham-Burley, 2016; Riordan et al., 2015). Previous studies have also identified that individuals have been willing to share their data for research purposes without displaying concern about privacy breaches, especially if the research participant perceived there to be public benefits from the research (or the potential for these) and trusted the individual or organization conducting the research and supported their use of the data (Aitken et al., 2016). An interesting question in connection with these established patterns is whether the meaning of privacy is changing for individuals or if, instead, these attitudes stem from their not possessing enough information about privacy and the risk of privacy breaches. Mulligan, Koopman, and Doty (2016) have recognized changes to the nature of privacy in relation to evolving technological and social conditions. The authors concluded that the public must be adequately informed about how personal health data will be used and the potential risks related to that via education and information campaigns (Aitken et al., 2016; Riordan et al., 2015).

More than ever before, clinical research is likely to take place on multiple sites, in several countries, and to involve organizationally challenging activity that engages multiple research stakeholders, all with their own knowledge and backgrounds (De Lusignan et al., 2016; Frizzo-Barker et al., 2016; Kuchinke et al., 2014; Nair & Ibrahim, 2015; O'Keefe & Rubin, 2015; Rho et al., 2015; Tucker et al., 2016; Yuan et al., 2017). In a research environment of this sort, the data sharing extends across highly varied technical infrastructures (O'Keefe & Rubin, 2015; Yuan et al., 2017); various stakeholders sort, process, and integrate the health data; and data get handled and stored on diverse systems, without any control by the entity that submitted the data (Kuchinke et al., 2014). According to Cowie et al. (2017), guaranteeing privacy, overcoming the challenges associated with linking diverse systems, and maintaining the infrastructure needed for repeated use of high-quality data are

some of the biggest challenges associated with using electronic health records in clinical research. Our results point to data-holders in these research environments (e.g., hospitals) as having a role in protecting an individual's privacy, and the papers emphasize the need for confidentiality. For example, data-holders are responsible for establishing transparent data-handling and data-sharing processes, implementing those processes (Frizzo-Barker et al., 2016; Noh et al., 2014; Tucker et al., 2016; Wolf et al., 2015), controlling and monitoring the procedures involved, and making sure that all processes have been carried out in line with legal requirements (Coppieters & Levêque, 2013; De Lusignan et al., 2016; Nair & Ibrahim, 2015; Tucker et al., 2016).

The selected articles illustrate that the meaning of de-identification and anonymization are not always clear (Halkoaho et al., 2012; O'Keefe & Rubin, 2015; Tucker et al., 2016). Elliot et al. (2016) defined de-identification as "a process of removing or masking direct identifiers in personal data such as a person's name, address, or other unique number associated with them," where the process includes also pseudonymization, a technique wherein "direct identifiers are replaced with a fictitious name or code that is unique to an individual but does not of itself directly identify them." Furthermore, they define anonymization as

> a process of ensuring that the risk of somebody being identified in the data is negligible. This invariably involves doing more than simply de-identifying the data, and often requires that data be further altered or masked in some way to prevent statistical linkage. (Elliot et al., 2016)

The growing importance of well-conducted anonymization in the era of big data and data sharing has been clearly identified (Elliot et al., 2016). However, anonymization remains a technically complex task wherein qualified research stakeholders must make the data anonymized (Meystre et al., 2014). Furthermore, as Elliot et al. (2016) have argued, anonymization is always strongly context-dependent. Therefore, only by considering the data and the use environment can one come to a well-informed decision about whether and what anonymization is needed. In addition, the evolving nature of health data and the increased use of said data in clinical research are creating their own challenges to the anonymization of data (Vayena & Blasimme, 2017).

Ethics consideration have an important role in the process of anonymization and also once data are anonymized. The papers point to at least three reasons for this (Elliot et al., 2016). The first ethics question is related to study-subjects' opportunity to control the anonymized data (Vayena & Blasimme, 2017). For instance, some data subjects may not want their personal data to be reused in general, by specific third parties, or for particular purposes (Elliot et al., 2016; Vayena & Blasimme, 2017). Addressing this issue, Vayena and Blasimme (2017) stated that individuals' control over the purpose of use cannot be exercised if the data have been anonymized and that anonymization may actually hinder autonomy in some cases. The second issue is that, as noted above, anonymization is never absolute and there is always a risk of identification of the individual and of privacy breaches (Elliot et al., 2016; Martínez et al., 2013; Tucker et al., 2016; Vayena & Blasimme, 2017). That said, there is still a need to develop effective and reliable anonymization techniques (Wade et al., 2017) and also to assess the risk of re-identification and minimize the risk of loss of individual-level privacy (Elliot et al., 2016; Tucker et al., 2016). On a positive note, studies have shown that data-holders provide support to help researchers navigate and understand the data they use in their research (Tucker et al., 2016). Third, data anonymization can enable the efficient use of electronic health data in clinical research and, through this, create benefits for society.

Privacy protection for study subjects in clinical research is guided by both ethical and legal considerations. Our study highlighted privacy issues and information governance as among the most complex aspects of implementing the use of health data in clinical research, partly because the regulations related to data privacy are subject to significant changes and international variation (Cowie et al., 2017). One critical question that emerged is that of how researchers have been able to plan and implement clinical research in line with the regulations in force (Coppieters & Levêque, 2013; De Lusignan et al., 2016; Halkoaho et al., 2012; Tucker et al., 2016), such that they took the privacy and confidentiality legislation relevant for their research into account in the real-world implementation (Coppieters & Levêque, 2013). Another concern identified is that current regulations are too slow to react and too inflexible to keep pace with the expanding role of health data, rapid changes in technology, and increasing availability of health details (Kuchinke et al., 2014; Wolf et al., 2015).

There have been significant advances in European Union data-protection law that will have an impact on researchers and health care professionals. Superseding earlier rules, the General Data Protection Regulation (GDPR) entered into full effect in May 2018 (GDPR 2016/679; Rumbold & Pierscionek, 2017). The GDPR is intended to harmonize the rules across the European Union (EU), to reduce the legal fragmentation, complexities, and uncertainties that existed between member states under the Data Protection Directive (Chassang, 2017; Rumbold & Pierscionek, 2017). In addition, the GDPR is designed to reinforce the rights of data subjects to privacy in a digitalized and evolving environment such that they can retain control over their personal data. This should help to preserve equilibrium between the need to protect data subjects' rights in a digitalized and globalized world and the

interest of permitting the processing of personal data, including sensitive data, for scientific research (Chassang, 2017). The GDPR applies to the protection of data of all individuals located in the European Union, and every research project that collects data in the EU must comply with it, even if the user of the data does not have a presence in the EU (The European Parliament and European Council, 2016).Therefore, some concern has been expressed that the new regulations may complicate international collaboration in the clinical-research field. Either way, it is clear that the use of personal data in clinical research demands more detailed planning, clear documentation, and transparency than ever before. Also, collaboration among stakeholders from early in the process is even more essential, especially in a multinational research process. In itself, the GDPR brings significant changes and privacy-protection implications for organizations that collect, process, and store personal data. Because of these changes, organizations need to review and revise all strategies, policies, processes, and technical solutions related to their handling of personal data. In addition, implementation of the GDPR necessitates new kinds of expertise, along with clearly defined roles and responsibilities within the organization and solid education of the personnel (Tikkinen-Piri, Rohunen, & Markkula, 2018).

Collaboration among research stakeholders was found to be a prerequisite for protecting the privacy of subjects of clinical research. For example, smooth collaboration of researchers, the data-holder providing approval for data access, and the ethics committee was cited as critical to the early stages of research (De Lusignan et al., 2016; Tucker et al., 2016); this enables certainty that all the relevant aspects of privacy and confidentiality are taken into account during the planning phase (Coppieters & Levêque, 2013; De Lusignan et al., 2016; Halkoaho et al., 2012). That conclusion is consistent with earlier literature, which identified collaboration as vital for ethically conducted clinical research (Cowie et al., 2017; Emanuel & Wendler, 2008; Nurmi et al., 2017). Accordingly, it is important to foster interstakeholder collaboration to better safeguard the privacy of study subjects in clinical research.

The Balance Between Societal Benefits and Individuals' Privacy

The topical aspect of ethics that emerged in the results of this study was balance between the benefits to society and the protection of individuals' privacy in clinical research (Coppieters & Levêque, 2013; Frizzo-Barker et al., 2016; Grande et al., 2015; Kuchinke et al., 2014; McGraw et al., 2015; Nair & Ibrahim, 2015; O'Keefe & Rubin, 2015; Tucker et al., 2016). As previous work did, the literature we examined indicated that the social benefits to be gained via increased use of health data in clinical research are significant (Cowie et al., 2017), with considerable public interest in the effective use of health data in that research. However, at the same time, we found increased concern that the growing use of health data in clinical research increases the risk of harm to subjects' privacy and to the confidentiality of their personal data (De Lusignan et al., 2016; Kuchinke et al., 2014; McGraw et al., 2015; Meystre et al., 2014; Nair & Ibrahim, 2015; O'Keefe & Rubin, 2015; Tucker et al., 2016).

It is also worthy of attention that many funding bodies and government policies specify open access or data sharing as a condition for funding (Ross, 2016; Smith et al., 2017). Open science is indeed an important goal in clinical-research practice (Ross, 2016; Smith et al., 2017). Increasing openness in research can enhance the reliability, transparency, and social impact of the research (Ministry of Education and Culture, 2014). However, promoting open science requires extensive cooperation within the research community, along with further development of research environments, research services, and research infrastructure (Ministry of Education and Culture, 2014); collaboration in these contexts; and project-specific planning of appropriate design, conduct, and reporting (Ross, 2016; Smith et al., 2017). At the same time, open access and increased data sharing bring noteworthy privacy concerns for individuals. Hence, Riso et al. (2017) have argued that there is a clear need to determine which trade-offs between individuals' rights and the common good are acceptable and to address how the thresholds for such trade-offs are best determined.

Our results indicate that achieving balance between social benefits and individual-level protection calls for reexamination of ethics frameworks, more flexible privacy rules, and new approaches to protecting privacy and confidentiality (McGraw et al., 2015; Meystre et al., 2014; Tucker et al., 2016). More research is needed into how individuals wish their information to be used for altruistic, socially beneficial purposes and how policy initiatives can accommodate these wishes (Grande et al., 2015), while both promoting public trust in clinical research and protecting individuals' privacy and the confidentiality of their health data (De Lusignan et al., 2016; Frizzo-Barker et al., 2016; Grande et al., 2015; Rho et al., 2015). It is possible to increase public trust by, for example, enabling greater involvement in the planning of beneficial research (De Lusignan et al., 2016) and through public debate and education about privacy, its meaning, privacy risks, and regulatory frameworks (Frizzo-Barker et al., 2016; Rho et al., 2015). These measures can create public support for the future use of research data but also convey greater awareness of the privacy aspect of clinical research and provide opportunities for public engagement and deliberation (Aitken et al., 2016). Above all, it is crucially important that appropriate safeguards be put in place to guarantee that health data will always be used ethically and that

individuals' privacy and confidentiality will be protected in clinical research (De Lusignan et al., 2016; Martínez et al., 2013).

Limitations of the Study

The limitations of our integrative review were related to the search strategy and the quality of the works examined. We planned the search strategy with a specialist librarian, and the authors of this article jointly selected the studies, analyzed the data, and conducted the quality evaluation. This may have improved the methodological rigor. We limited our electronic searches to the years 2012-2017 because we wanted to identify the latest studies. Also, we included only papers published in English, and no manual searches were carried out. Although, as Whittemore and Knafl (2005) have stated, electronic searches may identify only about half of the relevant studies (because of inconsistent search terminology and indexing problems), our electronic literature searches did yield great variety in the results. Another limitation that we faced was related to the included studies' quality: it varied, and the quality appraisal demonstrated some weaknesses in the study designs, samples, and methods. Nonetheless, all 16 papers selected were included in the review because of the quite limited number of studies conducted in this field. There was a risk of bias, but the similarities in findings across the papers examined supports the representativeness of our results and decreased that bias.

Best Practices

Protecting the privacy and confidentiality of research participants is a central aspect of ethical clinical research. Although this seems obvious, changing research practice and the evolving nature of health data make privacy protection even more challenging. In response, new, more innovative methods are needed for safeguarding privacy and confidentiality in clinical research. Our findings suggest that collaboration among the various stakeholders in research is a prerequisite for protecting privacy in clinical research and, therefore, must be encouraged. It is crucial also that organizations hosting clinical-research studies develop a transparent process for handling and sharing health data, apply that process well, and also provide control and surveillance of the procedures involved—to be sure that their processes remain in line with legal guidelines. In addition, the tension between study-subject privacy protection and benefits to society must be eased. This requires, among other things, public trust and more flexible privacy rules. However, the most important factor for reducing tension between regulations and clinical research practice is a guarantee of adequate research-ethics training for researchers and clinical-study staff, with solid recognition of the moral rationale behind the regulations.

The Research Agenda

Our study focused on the topical aspects of privacy and confidentiality, providing a synthesis that draws the previously fragmented body of empirical and theoretical studies together into a systematic knowledge base. Most work on privacy in a clinical-research context has looked at research participants' informational privacy and their right to control their personally identifying health information. Further research should extend the focus to other facets of privacy, such as its physical, psychological, and social dimensions. This would encourage a holistic understanding of privacy in clinical research. Also, some privacy problems occurring today are fundamentally different from those of the past, so more research is needed into how privacy and confidentiality may be breached during the clinical-research process. Such awareness could aid in creating processes through which privacy and confidentiality are protected throughout the research. Furthermore, there is a need for more in-depth studies examining how individuals understand privacy in clinical research and the ways in which they want their health information used in clinical research for the benefit of society. The central ethics question that researchers should explore is how to protect privacy and confidentiality in constantly changing clinical-research practice.

Educational Implications

Hospitals should provide education for researchers and clinical staff in relation to privacy and confidentiality regulations. This is particularly important in countries affected by significant developments, such as changes in European Union data-protection law. In addition, clinical researchers need new kinds of knowledge, understanding, skills, and methods if they are to protect study-subjects' privacy, in response to the evolving nature of health data and the increased use thereof in clinical research. Data-holders should help researchers navigate and understand the data sets they use in research. Policy makers and regulators, in turn, should attempt to develop clear and flexible regulations that guide privacy protection in clinical research. In parallel work, the scientific community could usefully strive to clarify and harmonize terminology (such as that related to anonymization and de-anonymization). As the public has limited knowledge related to privacy and confidentiality in clinical research, better understanding and education of laymen in this field are necessary for establishing public trust and for ethically conducted clinical research. For instance, it is vital for hospitals active in clinical research to cultivate and maintain open interaction with the local community in relation to privacy and confidentiality in clinical research. We argue that making the public and the health-policy sector more aware of the ethics factors in clinical research should be a high priority.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research project received funding from the Finnish Cultural Foundation, North Savo Regional Fund.

References

Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R., & Cunningham-Burley, S. (2016). Public responses to the sharing and linkage of health data for research purposes: A systematic review and thematic synthesis of qualitative studies. *BMC Medical Ethics*, *17*(1), Article 73. doi:10.1186/s12910-016-0153-x

Anuradha, R. B. (2012). Securing privacy for confidential databases using anonymization. *Middle-East Journal of Scientific Research*, *12*, 1792-1795. doi:10.5829/idosi.mejsr

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics*. New York, NY: Oxford University Press.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, *35*, 1017-1041. doi:10.2307/41409971

Brall, C., Maeckelberghe, E., Porz, R., Makhoul, J., & Schröder-Bäck, P. (2017). Research ethics 2.0: New perspectives on norms, values, and integrity in genomic research in times of even scarcer resources. *Public Health Genomics*, *20*, 27-35. doi:10.1159/000462960

Chassang, G. (2017). The impact of the EU general data protection regulation on scientific research. *Ecancermedicalscience*, *11*, Article 709. doi:10.3332/ecancer.2017.709

Cohen, J. E. (2013). What privacy is for. *Harvard Law Review*, *126*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2175406

Coppieters, Y., & Levêque, A. (2013). Ethics, privacy and the legal framework governing medical data: Opportunities or threats for biomedical and public health research? *Archives of Public Health*, *71*, 1-4. doi:10.1186/0778-7367-71-15

Cowie, M. R., Blomster, J. I., Curtis, L. H., Duclaux, S., Ford, I., Fritz, F., & Zalewski, A. (2017). Electronic health records to facilitate clinical research. *Clinical Research in Cardiology: Official Journal of the German Cardiac Society*, *106*, 1-9. doi:10.1007/s00392-016-1025-6

De Lusignan, S., Liyanage, H., Di Iorio, C. T., Chan, T., & Liaw, S.-T. (2016). Using routinely collected health data for surveillance, quality improvement and research: Framework and key questions to assess ethics, privacy and data access. *Journal of Innovation in Health Informatics*, *22*, 426-432. doi:10.1891/1541-6577.23.3.203

Elliot, M., Mackey, E., O'Hara, K., & Tudor, C. (2016). *The anonymisation decision-making framework*. Manchester: UKAN. Retrieved from http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf

Emanuel, E. J., & Wendler, D. (2008). An ethical framework for biomedical research. In E. J. Emanuel, C. C. Grady, R. A. Crouch, R. K. Lie, F. G. Miller, & W. D. Wendler (Eds.), *The Oxford textbook of clinical research ethics* (pp. 123-135). New York, NY: Oxford University Press.

The European Parliament and European Council. (2016). *General data protection regulation*. Retrieved from http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

Frizzo-Barker, J., Chow-White, P. A., Charters, A., & Ha, D. (2016). Genomic big data and privacy: Challenges and opportunities for precision Medicine. *Computer Supported Cooperative Work*, *25*, 115-136. doi:10.1007/s10606-016-9248-7

Grande, D., Asch, D. A., Wan, F., Bradbury, A. R., Jagsi, R., & Mitra, N. (2015). Are patients with cancer less willing to share their health information? Privacy, sensitivity, and social purpose. *Journal of Oncology Practice*, *11*, 378-383. doi:10.1200/JOP.2015.004820

Halkoaho, A., Pietilä, A.-M., Vesalainen, M., & Vähäkangas, K. (2012). Ethical aspects in tissue research: Thematic analysis of ethical statements to the research ethics committee. *BMC Medical Ethics*, *13*(1). doi:10.1186/1472-6939-13-20

Hodge, J., & Gostin, L. (2008). Confidentiality. In E. J. Emanuel, C. C. Grady, R. A. Crouch, R. K. Lie, F. G. Miller, & D. D. Wendler (Eds.), *The Oxford textbook of clinical research ethics* (pp. 673-281). New York, NY: Oxford university press.

Hoffman, S., & Podgurski, A. (2012). Balancing privacy, autonomy, and scientific needs in electronic health records research. *Electronic Health Records Research*, *85*, 87-143. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1923187

Hofmann, B. (2009). Broadening consent—And diluting ethics? *Journal of Medical Ethics*, *35*, 125-129. doi:10.1136/jme.2008.024851

Holm, S. (2016). Data protection legislation. In *Encyclopedia of life sciences* (pp. 1-3). Chichester, UK: John Wiley. http://www.els.net

Hopia, H., Latvala, E., & Liimatainen, L. (2016). Reviewing the methodology of an integrative review. *Scandinavian Journal of Caring Sciences*, *30*, 662-669. doi:10.1111/scs.12327

Kangasniemi, M., Pakkanen, P., & Korhonen, A. (2015). Professional ethics in nursing: An integrative review. *Journal of Advanced Nursing*, *71*, 1744-1757. doi:10.1111/jan.12619

Kaye, J. (2012). The tension between data sharing and the protection of privacy in genomics research. *Annual Review of Genomics and Human Genetics*, *13*, 415-431. doi:10.1146/annurev-genom-082410-101454

Kuchinke, W., Ohmann, C., Verheij, R. A., van Veen, E.-B., Arvanitis, T. N., Taweel, A., & Delaney, B. C. (2014). A standardised graphic method for describing data privacy frameworks in primary care research using a flexible zone model. *International Journal of Medical Informatics*, *83*, 941-957. doi:10.1016/j.ijmedinf.2014.08.009

Leino-Kilpi, H., Välimäki, M., Dassen, T., Gasull, M., Lemonidou, C., Scott, A., & Arndt, M. (2001). Privacy: A review of the literature. *International Journal of Nursing Studies*, *38*, 663-671.

Martínez, S., Sánchez, D., & Valls, A. (2013). A semantic framework to protect the privacy of electronic health records with non-numerical attributes. *Journal of Biomedical Informatics*, *46*, 294-303. doi:10.1016/J.JBI.2012.11.005

McGraw, D., Greene, S. M., Miner, C. S., Staman, K. L., Welch, M. J., & Rubel, A. (2015). Privacy and confidentiality in pragmatic clinical trials. *Clinical Trials*, *12*, 520-529. doi:10.1177/1740774515597677

Meystre, S. M., Ferrández, Ó., Friedlin, F. J., South, B. R., Shen, S., & Samore, M. H. (2014). Text de-identification for privacy protection: A study of its impact on clinical text information content. *Journal of Biomedical Informatics*, *50*, 142-150. doi:10.1016/j.jbi.2014.01.011

Ministry of Education and Culture. (2014). *Open science and research leads to surprising discoveries and creative insights*. Retrieved from http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/75210/okm21.pdf?sequence=1&isAllowed=y

Mulligan, D. K., Koopman, C., & Doty, N. (2016). Privacy is an essentially contested concept: A multi-dimensional analytic for mapping privacy. *Philosophical Transactions. Series A: Mathematical, Physical, and Engineering Sciences*, *374*. doi:10.1098/rsta.2016.0118

Nair, S. C., & Ibrahim, H. (2015). Assessing subject privacy and data confidentiality in an emerging region for clinical trials: United Arab Emirates. *Accountability in Research*, *22*, 205-221. doi:10.1080/08989621.2014.942452

Nicholls, S. G., Langan, S. M., Benchimol, E. I., Moher, D., Glasziou, P., Altman, D., & Hoey, J. (2016). Reporting transparency: Making the ethical mandate explicit. *BMC Medicine*, *14*, Article 44. doi:10.1186/s12916-016-0587-5

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*, 101-139.

Noh, G., Chun, J. Y., & Jeong, I. R. (2014). Sharing privacy protected and statistically sound clinical research data using outsourced data storage. *Journal of Applied Mathematics*, *2014*, 1-12. doi:10.1155/2014/381361

Nurmi, S.-M., Halkoaho, A., Kangasniemi, M., & Pietilä, A.-M. (2017). Collaborative partnership and the social value of clinical research: A qualitative secondary analysis. *BMC Medical Ethics*, *18*, Article 57. doi:10.1186/s12910-017-0217-6

O'Keefe, C. M., & Rubin, D. B. (2015). Individual privacy versus public good: Protecting confidentiality in health research. *Statistics in Medicine*, *34*, 3081-3103. doi:10.1002/sim.6543

Porsdam Mann, S., Savulescu, J., & Sahakian, B. J. (2016). Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*, 1-17. doi:10.1098/rsta.2016.0130

Rho, M. J., Jang, K. S., Chung, K.-Y., & Choi, I. Y. (2015). Comparison of knowledge, attitudes, and trust for the use of personal health information in clinical research. *Multimedia Tools and Applications*, *74*, 2391-2404. doi:10.1007/s11042-013-1772-6

Riordan, F., Papoutsi, C., Reed, J. E., Marston, C., Bell, D., & Majeed, A. (2015). Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. *International Journal of Medical Informatics*, *84*, 237-247. doi:10.1016/j.ijmedinf.2015.01.008

Riso, B., Tupasela, A., Vears, D. F., Felzmann, H., Cockbain, J., Loi, M., . . . Rakic, V. (2017). Ethical sharing of health data in online platforms—Which values should be considered? *Life Sciences, Society and Policy*, *13*, Article 12. doi:10.1186/s40504-017-0060-z

Ross, J. S. (2016). Clinical research data sharing: What an open science world means for researchers involved in evidence synthesis. *Systematic Reviews*, *5*, Article 159. doi:10.1186/s13643-016-0334-1

Rumbold, J. M. M., & Pierscionek, B. K. (2017). A critique of the regulation of data science in healthcare research in the European Union. *BMC Medical Ethics*, *18*, Article 27. doi:10.1186/s12910-017-0184-y

Scott, F. I., McConnell, R. A., Lewis, M. E., & Lewis, J. D. (2012). Increasing complexity of clinical research in gastroenterology: Implications for the training of clinician-scientists. *The American Journal of Gastroenterology*, *107*, 496-500. doi:10.1038/ajg.2011.450

Smith, C. T., Nevitt, S., Appelbe, D., Appleton, R., Dixon, P., Harrison, J., . . . Tremain, E. (2017). Resource implications of preparing individual participant data from a clinical trial to share with external researchers. *Trials*, *18*, Article 319. doi:10.1186/s13063-017-2067-4

Soares, B. C., Hoga, L. A. K., Peduzzi, M., Sangaleti, C., Yonekura, T., & Silva, D. R. A. D. (2014). Integrative review: Concepts and methods used in nursing descriptors Integrative review: Concepts and methods used in nursing CritiCal review. *Revista da Escola de Enfermagem da USP*, *48*, 329-339. doi:10.1590/S0080-623420140000200020

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, *154*, 477-588. doi:10.2307/40041279

Stiles, P. G., & Petrila, J. (2011). Research and confidentiality: Legal issues and risk management strategies. *Psychology, Public Policy, and Law*, *17*, 333-356. doi:10.1037/a0022507

Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, *34*, 134-153. doi:10.1016/J.CLSR.2017.05.015

Tucker, K., Branson, J., Dilleen, M., Hollis, S., Loughlin, P., Nixon, M. J., & Williams, Z. (2016). Protecting patient privacy when sharing patient-level data from clinical trials. *BMC Medical Research Methodology*, *16*, 1-10. doi:10.1186/s12874-016-0169-4

United Nations. (1948). *Universal declaration of human rights*. Author. Retrieved from http://www.un.org/en/universal-declaration-human-rights/

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

Wade, J., Elliott, D., Avery, K. N. L., Gaunt, D., Young, G. J., Barnes, R., . . . Donovan, J. L. (2017). Informed consent in randomised controlled trials: Development and preliminary evaluation of a measure of Participatory and Informed Consent (PIC). *Trials*, *18*, 1-13. doi:10.1186/s13063-017-2048-7

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, *52*, 546-553. doi:10.1111/j.1365-2648.2005.03621.x

Windows Media Audio. (2013). *Declaration of Helsinki*. Retrieved from https://www.wma.net/policies-post/wma-declaration-

of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Wolf, L. E., Patel, M. J., Williams Tarver, B. A., Austin, J. L., Dame, L. A., & Beskow, L. M. (2015). Certificates of confidentiality: Protecting human subject research data in law and practice. *The Journal of Law, Medicine & Ethics*, *43*, 594-609. doi:10.1111/jlme.12302

Yuan, J., Malin, B., Modave, F., Guo, Y., Hogan, W. R., Shenkman, E., & Bian, J. (2017). Towards a privacy preserving cohort discovery framework for clinical research networks. *Journal of Biomedical Informatics*, *66*, 42-51. doi:10.1016/j.jbi.2016.12.008

Author Biographies

Sanna-Maria Nurmi is a PhD student in health science at the University of Eastern Finland. Her main research interest has been focused on the ethical aspects of clinical research. In this article, she has contributed on the data collection, data analysis, and interpretation of findings. She also drafted and revised the manuscript.

Mari Kangasniemi is working as a university lecturer at the University of Eastern Finland. Her main research interest has been focused on the ethics and change of work in health care and nursing. In this article, she has contributed on the analysis and interpretation of the findings and critical revision of the manuscript.

Arja Halkoaho is a PhD and docent in health science. Currently she is working as a principal lecturer in Tampere University of Applied Sciences in Finland. Her main research interests include research ethics and health promotion. She designed the study together with research group. She contributed the data collection, interpretation of findings, and revision of the manuscript.

Anna-Maija Pietilä is a professor of preventive nursing science at the University of Eastern Finland. Her research has been focused on ethics and health promotion in the different stages of individuals' life span. In this article, she has contributed on the analysis and interpretation of the data and critical revision of the manuscript.