Stakeholder’s Perceptions of Value and Risks in Data Governance for the Secondary Use of Health Data

Hannu Nieminen\textsuperscript{a}, Rima Sermontyte-Baniule\textsuperscript{b}\textsuperscript{c} and Nina Helander\textsuperscript{d}

\textsuperscript{a}Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
\textsuperscript{b}School of Economics and Business, Kaunas University of Technology, Kaunas, Lithuania
\textsuperscript{c}Information and Knowledge Management, Tampere University, Tampere, Finland

Keywords: Health Data Governance, Secondary Health Data, Stakeholder Perceptions, Value Creation, Review.

Abstract: The study is a literature study assessing the value expectations and risks perceived by the different stakeholders related to the governance of secondary use of health data. A key value expectation for all stakeholders was found to be that data provides public benefits and “common good”, especially through academic research. Especially for the researchers improvement of health equity in the society was also an important value expectation. For patients and also for decisionmakers security and privacy related risks were often mentioned. For all stakeholders the risk of stigma for different groups in the society and for the patient herself was seen to be important. Constant and clear communication towards all stakeholders about what data is collected, how it is used, what the expected benefits are and how the risks are managed needs to be a key element of health data governance solutions. All stakeholders see the importance of involving also the patient representatives to the governance of health data. Data governance should be developed towards a continuous and transparent collaborative process, where all stakeholders voice is heard, and they can affect the decisions.

1 INTRODUCTION

Data is often referred as a key resource for creating value at the level of individuals, organizations and societies. Value creation is a multifaceted and complex concept (Sidorchuk 2015, Climent and Haftr 2021) as multiple elements affect value perception, such as functionality, aesthetics, symbolic, financial, social, and emotional aspects (Karababa and Kjeldgaard, 2014). Value creation can be defined as the trade-off between benefits captured and sacrifices made/risks realized (see e.g. Helander and Kukko,2009).

To ensure effective value creation, good governance of data is essential. Data governance can be defined as “the practice of managing data assets throughout their lifecycle to ensure that they meet organizational quality and integrity standards” (Abraham et al. 2019). Data governance is a highly cross-functional effort to increase the value of data, to minimize data-related costs and risks and to be able to utilize data as a strategic asset (Abraham et al. 2019, Carretero et al. 2017, Zhang et al. 2022).

Health data collected by public and private health care organizations and governments is an extremely valuable resource that could, in addition to its primary use in health care, be used in the research aiming to improve health outcomes, for improving the quality, safety and cost-effectiveness for health care systems and, also, for supporting the development of new products and services. These types of uses are typically referred to as secondary use of health data.

From the governance perspective secondary use of health data is in many ways a special case: health information consists of highly protected personal health data, maintaining the privacy and security of many types of health data can be difficult, there are exceptional needs for data security, data sources are heterogenous, interoperability between different countries health care systems is a challenge and the regulatory framework still remains fragmented (Pinto et al. 2021, OECD 2015).
Different stakeholders (patients, medical professionals, researchers, health care managers and decisionmakers) have their own perspectives towards the value creation. Value expectations of patients / citizens for the secondary use of data have been summarized in recent systematic reviews (Aitken et al. 2016, Skovgaard et al. 2019, Perrin and Mathieu 2021, Kalkman et al. 2022). Patients have been found to be in general supportive towards the secondary use of health data, provided that the data is used for “common good” purposes, and that they trust the organizations (Aitken et al. 2016, Skovgaard et al. 2019, Perrin and Mathieu 2021, Kalkman et al. 2022). Key concerns were related to confidentiality, individual’s control over their data (Aitken et al. 2016, Perrin and Mathieu 2021, Kalkman et al. 2022) and misuse of data in a way that puts some groups of people in bad light (Skovgaard et al. 2019). Also, citizens often lack knowledge on secondary use of health data and on how the privacy and security of the data is protected (Aitken et al. 2016, Perrin and Mathieu 2021, Kalkman et al. 2022).

While there are recent literature reviews about the benefits and risks expectation from the patient perspective, reviews combining the perspectives of stakeholders are abundant. The ways to create value for the various stakeholders has recently gained growing attention in the academic debate on value creation in general (e.g. Busch et al. 2016), but not yet much in the context of secondary use of health data.

This research study aims to assess, based on a literature study, the value expectations and risks perceived by the different stakeholders related to the governance of health data for secondary use. Understanding the value expectations and perceived risks provides input for building governance solutions, which meet the expectations of all stakeholders as well as possible.

In the next section, the research method is presented, and after this the findings of the literature review are presented. Finally, the research results and limitations of the study are discussed and the key conclusions are summarized.

2 RESEARCH METHOD

To find the right articles related to the topic, the following search terms were combined to search articles published during the past 10 years from the Web of Science database: health data, governance, (secondary use OR re-use OR data sharing), (value OR benefit), (risk OR problem), (stakeholder OR patient OR citizen OR professional OR manager OR industry OR decisionmaker OR physician).

Articles fulfilling all the following inclusion criteria were included in the further analysis: 1) Article utilizes or refers to health data governance for the secondary use of data, 2) Perceived value, benefits, risks or problems of secondary use are studied from the perspective of at least one of the identified stakeholder groups, 3) Article presents original results from empirical research.

In addition, searches were done by scanning the references of included papers. After the analysis, a total of 23 papers were selected for further study. A thematic synthesis approach (Thomas and Harden 2008) was adopted. First the key findings describing the values, benefits, risks and problems perceived by the different stakeholders were extracted from the publications. After this, the findings were classified under descriptive themes emerging from them, and in the final phase the themes were interpreted in relation to the research question and analytical themes were developed.

3 FINDINGS

19 of the selected papers were studying the patients/citizens, 10 medical researchers and 5 decisionmakers / managers (8 papers studied the perceptions of several stakeholders).

Different types of mostly qualitative methods were used in the studies. Most common methods were structured interviews (4), focus groups (5), surveys (8) and multi-method studies combining interviews and focus groups (6). Total aggregated N for studies using surveys was 15794, varying between 280 and 8004. For other methods the aggregated N was 655, varying between 18 and 73 in the publications.

3.1 Value / Benefit Expectations

For patients, a total of five themes describing their expectations for value/benefit from the secondary use of data were identified: The most common value expectation (9 papers) was that the data contributes to public good through academic research (e.g. Audrey et al. 2016, Grande et al. 2013, Karampela et al. 2019). Related to this, the second most important expectation was that data provides public benefit improving the health of the population and health equity (7) (e.g. Evans et al. 2020, Tully et al. 2018, Spencer et al. 2016). Patients strongly except, that public good for the society is achieved from the data they share, and that the governance solutions help...
them to trust that this expectation is fulfilled. Improvement of the quality of care or outcomes for patients was also mentioned in 5 papers (e.g. Adanijo et al. 2021, Velarde et al. 2021). Other value expectations from patients included data advancing innovation (4) (e.g. Colombo et al. 2010, Johansson et al. 2021) and data improving the quality of research (3) (e.g. Manhas et al. 2018). Patients often were suspicious about the use of data for advancing commercial innovation, but with more information about the benefits and safeguards they became more accepting towards it (Tully et al. 2018).

For researchers the public benefit of improving the health of the population and health equity was the most often mentioned value expectation (4/10 papers). Health data was seen as an essential public resource to protect and produce population health (Evans et al. 2020,Mbuthia et al. 2019) and a key tool to help reach better health equity e.g. for low-to-middle income settings (e.g. Jao et al. 2015). Other benefits included data improving care and saving resources (3 papers, e.g. Neves et al. 2019), improvement of societal decision-making and regulation (Hate et al. 2015) and improving the quality of research (Adanijo et al. 2021). For decisionmakers the most often mentioned value expectations were improvement of care and saving of resources (3/5, e.g. Mazor et al. 2017), data improving the quality of research (2) and public good benefits (2).

Table 1: Most common value/benefit expectations.

<table>
<thead>
<tr>
<th>Patients (19 papers)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Public good through academic research</td>
<td>9</td>
</tr>
<tr>
<td>Public benefit improving the health of the population and health equity</td>
<td>7</td>
</tr>
<tr>
<td>Improvement of the quality of care or outcomes for patients</td>
<td>5</td>
</tr>
<tr>
<td>Researchers (10 papers)</td>
<td></td>
</tr>
<tr>
<td>Public benefit of improving the health of the population and health equity</td>
<td>4</td>
</tr>
<tr>
<td>Data improving care and saving resources</td>
<td>3</td>
</tr>
<tr>
<td>Decisionmakers (5 papers)</td>
<td></td>
</tr>
<tr>
<td>Improvement of care and saving of resources</td>
<td>3</td>
</tr>
<tr>
<td>Data improving the quality of research</td>
<td>2</td>
</tr>
<tr>
<td>Public good benefits</td>
<td>2</td>
</tr>
</tbody>
</table>

Some of the studied papers did not specifically address the value expectations, but were more focusing on risks, foreseen problems and recommendations for data governance solutions.

### 3.2 Foreseen Risks and Problems

Several themes were identified for the foreseen risks and problems in the governance of health data. Patients most often mentioned privacy and confidentiality related risks (10 papers). These included especially risks of being able to identify a person even from anonymized data (e.g. Cheah et al. 2018) and data and privacy breaches and infringements (e.g. Seltzer et al. 2019). Mistrust in the organizations governing data so that patients cannot affect how and by whom the data is used was the second most commonly mentioned risk (8 papers, e.g. McCracken and Hopkins 2021, Shah et al.2019). Due to lack of transparency and awareness of the ways how the data will be used, data could be misused by companies, government or other actors for “bad intentions” (Evans et al. 2020). Third most common theme (6 papers) was the risk of data leading to stigmatizing or discriminating some groups in the society, for example gays, ethnic groups or people living in a certain area (e.g. Audrey et al. 2016, Cheah et al. 2018). Not achieving public good due to the data used for profit-making purposes was seen as a risk in 5 papers, e.g. Mazor et al. 2017. Also, negative effect on care for example due to the long lifespan of data “permanently marking” the individual (Evans et al. 2020), patients not understanding what they are consenting to and later inconvenience for the patient were seen as risks.

Table 2: Most common foreseen risks and problems.

<table>
<thead>
<tr>
<th>Patients (19 papers)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy and confidentiality</td>
<td>10</td>
</tr>
<tr>
<td>Mistrust in organizations: patients cannot affect how and by whom the data is used</td>
<td>8</td>
</tr>
<tr>
<td>Data stigmatizing or discriminating some groups in the society</td>
<td>6</td>
</tr>
<tr>
<td>Not achieving public good due to the data used for profit-making purposes</td>
<td>5</td>
</tr>
<tr>
<td>Researchers (10 papers)</td>
<td></td>
</tr>
<tr>
<td>Data stigmatizing or discriminating some groups in the society</td>
<td>5</td>
</tr>
<tr>
<td>Decisionmakers (5 papers)</td>
<td></td>
</tr>
<tr>
<td>Security and privacy problems</td>
<td>4</td>
</tr>
<tr>
<td>Lack of transparency and awareness in the way how data is used</td>
<td>4</td>
</tr>
<tr>
<td>Problems in data quality</td>
<td>3</td>
</tr>
<tr>
<td>Data stigmatizing or discriminating some groups</td>
<td>3</td>
</tr>
</tbody>
</table>

For medical researchers the variation in the foreseen risks was high. 5 papers mentioned the risk of data leading to stigmatizing groups in society as a risk (e.g. Jao et al. 2015). Other risks included...
excessive costs for the governance (3, e.g. Ballantyne et al. 2020), additional burden for the patient (4), patients not understanding what they are consenting to (3), privacy and confidentiality (3), mistrust in the organization governing the data (3), lack of transparency in the ways how data is used (2), data affecting care in a negative way (3), prioritizing profit (3) and problems in research quality (2).

**Decisionmakers** identified as risks and problems security and privacy problems (4), lack of transparency and awareness in the way how data is used (4), problems in data quality affecting research results (3) for example inherent biases in collecting data misleading decisions (Evans et al. 2020) and data leading to stigmatizing groups in the society (3).

### 3.3 Recommendations for Data Governance

Based on the study findings, many papers provided recommendations and requirements for the data governance. From **patients** the most common recommendation was the need to provide clear and understandable information about how the data is shared and what the key benefits are (6 papers, e.g. Audrey et al. 2016, Kim et al. 2015). Existing safeguarding procedures need to be highlighted (Adanijo et al. 2021) and a review process to oversee the use of data needs to be at place (Johansson et al. 2021). **All stakeholders**, including **patient representatives**, should be involved in the governance of data (Adanijo et al. 2021, Hate et al. 2015) and patients should have choice on what data is shared and have a possibility to opt out (3 papers). **Ensuring trust in the organizations governing the data** (4) and **safety and security** (4) is essential. Sanctions should be at place in case of data misuse (Colombo et al. 2019). In order to maintain trust good communication is important (Hate et al. 2015).

**Researchers** agree with the patients that **all stakeholders should be involved in the governance of data** (4, e.g. Manhas et al. 2018). Community engagement in the data governance committees is seen as an essential element of ethical practice (e.g. Jao et al. 2015). **Ensuring patient consent** is a key requirement (3, e.g. Hate et al. 2015, Stevenson 2015). Effective processes are needed for the governance of requests for data re-use in later projects, and a broad form of consent would make this easier (Jao et al. 2015). Other recommendations include **need to provide clear and understandable information how data is shared and what the benefits are** and **ensuring security and safety** through strict safeguards.

**Decisionmakers** provide very similar recommendations as researchers, emphasizing the involvement of all stakeholders in the governance, ensuring clear and understandable information for all stakeholders, ensuring safety and security, ensuring patient consent and building trust in all entities participating in the data governance process.

As a general observation, the reviewed papers did not provide detailed level descriptions on the health data governance models. Thus, there still remains in the literature a gap of the discussion on successful health data governance model in practice.

| Patients (19 papers) | | | |
|----------------------|---|---|
| Clear and understandable information about how the data is shared and what the key benefits are | 6 |
| Ensuring trust in the organizations governing the data | 4 |
| Ensuring safety and security | 5 |
| Patients should involved in the governance, and should have a choice on what data is shared | 3 |

| Researchers (10 papers) | | | |
|------------------------|---|---|
| All stakeholders should be involved in the governance of data | 4 |
| Ensuring patient consent | 3 |

| Decisionmakers (5 papers) | | | |
|---------------------------|---|---|
| All stakeholders should be involved in the governance of data | 2 |
| Ensuring clear and understandable information for all stakeholders | 2 |

### 4 DISCUSSION

Understanding of the stakeholder’s expectations for the values and risks for the secondary use of health data is important to ensure the development of high-quality solutions for health data governance, which can be accepted both by the citizens/patients and by the medical professionals.

Recent literature reviews have indicated that for the patients a key value expectation is that the secondary use of health data should ensure public benefits and “common good”, and that trust in the organizations utilizing the data is essential for them (Aitken et al. 2016, Skovgaard et al. 2019, Perrin and Mathieu 2021, Kalkman et al. 2022). Our review results indicate, that this key value expectation is also shared by the other stakeholders. For the researchers improvement of health equity was also seen as an important value expectation. Big data is seen as a valuable resource to improve and develop the society towards being more equal. These types of altruistic
value goals are mentioned more often than values related to care improvement or innovation, and especially patients seem to be somewhat suspicious towards industry’s use of data for innovation purposes. Most probably being more transparent in explaining the benefits of innovation, ways how data would be used and the safeguards at place would make the stakeholders more positive towards these types of uses.

Summarising the results of value expectations of patients, researchers and decisionmakers we have identified that despite common value expectations there exist unique value expectations, specifically for patients and researchers. Analyzed papers report such value expectation by patients as ability of data to advance innovations. Interestingly, it is not so much reported by researchers or decisionmakers which perhaps are more involved into research and innovation processes than patients. Other interesting insight is the fact, that researchers report such value / benefit expectation as improvement of societal decision-making and regulation which is not reported by decisionmakers themselves as value / benefit. Contrary, decision makers see the benefit of secondary data more as helping to improve the quality of research. In other words, researchers and decision makers do not assign certain value as applicable to them even though it is closely related.

Regarding the perceived risks, for patients and also for decisionmakers security and privacy related risks were most often mentioned. This corresponds well with the earlier reviews (Aitken et al. 2016, Perrin and Mathieu 2021, Kalkman et al. 2022). Researchers, however, did not see these risks as so important. Possible reason for this is that professionals know in more detail the ways how privacy is protected. Interestingly, for all stakeholder groups the risk of stigma for different groups in the society was seen to be important. This would come from the misuse of data or from the inherent biases and limitations in the collected data.

Full transparency towards all stakeholders on the ways what data is collected and used and on the expected benefits would be essentially important for building the trust. Specifically, literature analysis highlights, that a worry of not achieving public good due to the data used for profit-making commercial purposes is expressed by patients, but not reported by researchers or decisionmakers. Constant and clear communication towards all stakeholders needs to be a key element for the future data governance models.

When data is collected, it is required that patient signs an informed consent. This consent typically is only for needs of one study, and it describes in detail what data is collected, how it is protected and what it is used for. Broad forms of consent are also in use, where patient gives the right for using data also for secondary studies. While consent forms provide information for the patients about the future use of data, this “one governance contact with the patient” principle is not enough. Advanced systems, where the patient can monitor the requests for data use and opt out or adapt her preferences are also being developed (e.g. Williams et al. 2015).

Based on our review, stakeholders see the importance of involving also the patient representatives to the governance of health data. Data sharing could be overseen by a committee involving e.g. (Hate et al 2015) decisionmakers, internal researchers, patient representatives, representatives from the communities where data is collected and ethicists. As Joa et al. (2015) states: “governance processes need to include openness, solidarity, fairness, and truth-telling”. Building this kind of stakeholder involvement and continuous and transparent communication process is a key challenge for the future health data governance solutions and thus, important avenue for further research.

The results of this literature study will be utilized and further verified in our ongoing empirical study on stakeholders perceptions on value, benefits and risks of secondary use of health data. Study is focusing on long-term home care of chronic diseases, and data is being collected in 5 countries (Finland, Sweden, France, Lithuania and Spain) as part of the DiHECO (Digital Healthcare Ecosystem research and innovation capability building) project.

**Limitations of the Study**

Even though this literature based research at hand was able to give some guidelines for further research, the research also faced limitations. For example, the number of articles describing the perceptions of researchers / medical professionals (10) and decisionmakers (5) was smaller than the number of articles analyzing the patients perspective (19), leading to potential emphasis on patients’ views.

**5 CONCLUSIONS**

Based on our literature study, a key value expectation for the secondary use of health for all stakeholders is that data provides public benefits and “common good”, especially through academic research. Especially for the researchers improvement of health equity in the society is also seen as an important value expectation. For patients and also for decisionmakers...
different security and privacy related risks were the most often mentioned risks, while researchers did not mention these risks so often. For all stakeholder groups the risk of stigma for different groups in the society and for the patient herself was seen to be important.

Constant and clear communication towards all stakeholders about what data is collected, how it is used, what the expected benefits are and how the risks are managed need to be a key element of health data governance solutions. Communications is the essential enabler for building the needed trust between the stakeholders. All stakeholders see the importance of involving also the patient representatives to the governance of health data. Data governance should be developed towards a continuous and transparent collaborative process, where all stakeholders voice is heard, and they can affect the decisions. Building this kind of stakeholder involvement and continuous and transparent communication process is a key challenge for the future health data governance solutions.

ACKNOWLEDGEMENTS

This work is a part of the DiHECO project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 952012.

REFERENCES


Adanijo, A., McWilliams, C., Wykes, T., Julka, S. (2021) Investigating mental health service user opinions on clinical data sharing: Qualitative focus group study. JMIR Mental Health, vol. 8, iss. 9


