Technology in Society 66 (2021) 101625

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

Governance mechanisms for sharing of health data: An approach towards selecting attributes for complex discrete choice experiment studies

Jennifer Viberg Johansson [a,*], Nisha Shah [b], Eik Haraldsdóttir [c], Heidi Beate Bentzen [d], Sarah Coy [b], Jane Kaye [b,e], Deborah Mascalzoni [a,f], Jorien Veldwijk [a,g,h]

[a] Centre for Research Ethics & Bioethics, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden
[b] Centre for Health, Law and Emerging Technologies (HeLEX), Faculty of Law, University of Oxford, Oxford, UK
[c] Social Science Research Institute, University of Iceland, Reykjavik, Iceland
[d] Norwegian Research Center for Computers and Law, Faculty of Law, University of Oslo, Oslo, Norway
[e] Centre for Health, Law and Emerging Technologies (HeLEX), Melbourne Law School, University of Melbourne, Australia
[f] Institute of Biomedicine, Eurac Research, Bolzano, Italy
[g] Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, the Netherlands
[h] Erasmus Choice Modelling Centre, Erasmus University Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Discrete choice experiment
Attribute development
Governance of health data
Preferences

ABSTRACT

Background: Discrete Choice Experiment (DCE) is a well-established technique to elicit individual preferences, but it has rarely been used to elicit governance preferences for health data sharing.
Objectives: The aim of this article was to describe the process of identifying attributes for a DCE study aiming to elicit preferences of citizens in Sweden, Iceland and the UK for governance mechanisms for digitally sharing different kinds of health data in different contexts.
Methods: A three-step approach was utilised to inform the attribute and level selection: 1) Attribute identification, 2) Attribute development and 3) Attribute refinement. First, we developed an initial set of potential attributes from a literature review and a workshop with experts. To further develop attributes, focus group discussions with citizens (n = 13), ranking exercises among focus group participants (n = 48) and expert interviews (n = 18) were performed. Thereafter, attributes were refined using group discussion (n = 3) with experts as well as cognitive interviews with citizens (n = 11).
Results: The results led to the selection of seven attributes for further development: 1) level of identification, 2) the purpose of data use, 3) type of information, 4) consent, 5) new data user, 6) collector and 7) the oversight of data sharing. Differences were found between countries regarding the order of top three attributes. The process outlined participants' conceptualisation of the chosen attributes, and what we learned for our attribute development phase.
Conclusions: This study demonstrates a process for selection of attributes for a (multi-country) DCE involving three stages: Attribute identification, Attribute development and Attribute refinement. This study can contribute to improve the ethical aspects and good practice of this phase in DCE studies. Specifically, it can contribute to the development of governance mechanisms in the digital world, where people's health data are shared for multiple purposes.

1. Introduction

With rapid advances in digital health and computing technologies, there has been considerable debate about privacy issues relating to the wide sharing of health data in Cyberspace [1]. It has been argued that people's attitudes and preferences for sharing data digitally should be considered in the process of policymaking [2]. While there has been considerable empirical research on people's attitudes towards the reuse of personal data for research [3–6], there are no studies inquiring about peoples' preferences for the governance of sharing health data by a variety of users such as technology companies or through national data access hubs. In the context of health, to our knowledge, no studies have

* Corresponding author. Centre for Research Ethics & Bioethics, Uppsala University, Box 564, SE-751 22, Uppsala, Sweden.,
E-mail addresses: jennifer.viberg@crb.uu.se, deborah.mascalzoni@crb.uu.se (J. Viberg Johansson).

https://doi.org/10.1016/j.techsoc.2021.101625
Received 3 March 2021; Received in revised form 24 May 2021; Accepted 29 May 2021
Available online 16 June 2021

investigated how people make trade-offs between privacy and data sharing when the same data is reused in various settings. This may in part be due to the methodological limitations of previously employed methods in the privacy field. Hence, this has left the privacy field with a knowledge gap that can only be answered adequately with a more innovative methodological approach.

There are number of aspects or characteristics to consider for people when deciding to share their health information in various contexts. Hence, to identify how individuals balance the characteristics of various aspects of their health information, we hypothesized that a discrete choice experiment (DCE) could be a suitable method to use. DCE is a well-established stated preference method to elicit preferences [7], widely applied in transportation, marketing and health (care) [8–10]. The method is derived from the random utility theory [11,12]. Respondents in a survey are confronted with a set of alternatives from which to choose the alternative they prefer. Each alternative consists of different combinations of levels of a set of multiple attributes [11,13]. To date, DCEs have not been used to understand peoples' preferences for sharing health data digitally. DCEs can manage complex choice scenarios, where each decision carries both advantages and disadvantages for the individual. Rather than 'just' asking people what they think is important, this method gives insight on how important different factors are and what trade-offs people make between these factors [14]. We therefore chose to use a DCE to investigate people's preferences for sharing health data digitally, and we found the method to be well suited for such innovative privacy research [15]. When setting out to use this well-established methodology in a new field, we identified another knowledge gap, namely the lack of a thorough explanation of the development process of DCE attributes and levels. We seek to fill that knowledge gap with this paper for the benefit of others developing a DCE or seeking to understand if it would be an appropriate methodology for yet other fields where it has not previously been used.

The DCE method is limited regarding how many factors (i.e. attributes and levels) can be included in the choices presented; therefore, a careful selection process is needed. Best-practice guidelines describing the process of developing a DCE instrument involve reviewing relevant literature and conducting qualitative research with the target population to identify important attributes [16]. There is a tendency that qualitative preparatory work required prior to dissemination of a DCE is often presented briefly and limited in DCE studies. Moreover, it is often poorly described in relation to the analytical process involved in reaching the final DCE survey [7,16,17]. A structured and transparent process for attribute selection and considerations is crucial for evaluating the validity and trustworthiness of the DCE study and to be able to compare studies on similar topics. By explaining each step of the preparatory work necessary in order to select attributes, from the qualitative to the quantitative work suggestion of attribute and levels, the content validity of the coming DCE can be evaluated.

The aim of our study was to complement an attribute development approach that outlines, not only how to select attributes for a DCE study, but also present the researchers' practical experience between attribute identification and selection. Therefore, this study will make an important first step by contributing to ongoing discussions within preference research regarding the need for a transparent approach towards attribute selection. While we recognise that for some well-studied topics like vaccination, transportation or particular clinical settings, the attribute identification and selection process is considered straightforward, describing this process is still required as part of evidence-based good-practice guidelines. However, structured attribute identification and selection approaches are particularly necessary in unexplored or more complex areas like cybersecurity research. For that reason, this study focuses on developing an attribute selection approach by identifying attributes and levels that are most relevant to citizens in the context of sharing health data digitally.

2. Method

This study forms part of a larger research project, aimed at identifying the governance mechanisms that inspire trust and resilience in the digital health sphere. The countries included in this study were Iceland, Sweden and the UK. Researchers from each country developed and performed the data collection. The data analysis and confirmation of the results was conducted collaboratively through regular meetings. The study team comprised complementing expertise including law, philosophy, ethics and social science. The conduct and reporting of the study followed the guidelines for conducting a DCE study [16,17].

A three-step approach (see Fig. 1) was utilised to inform attribute and level selection: Step 1) Attribute identification, Step 2) Attribute development and Step 3) Attribute refinement.

2.1. Attribute identification - Step 1

2.1.1. Literature review and discussion with experts

A rapid scoping review was conducted to identify potential attributes related to governance mechanisms in the digital health sphere [18]. The search was conducted in MedLine, CINAHL, EMBASE, Web of Science and PsychINFO. The search was limited to English-language full text papers, human adult studies, and articles between January 1, 2016 and December 7, 2018. The search was started in 2016 because the EU General Data Protection Regulation 2016/679 (GDPR) entered into force across Europe in the spring of 2016. The inclusion criteria referred to empirical articles that presented results of quantitative and qualitative research about stakeholder perspectives on health data sharing. Participants in the empirical articles could be members of the public, patients, end-users, consumers, experts (i.e., policy makers/commissioners/healthcare professionals/organisations' opinions). We included articles that examined attitudes, opinions or beliefs about health data sharing in different health contexts, for different purposes. See the key search terms in Appendix 1 in the Supplementary files. The reviewers (NS, JVJ, JK and EH) initially screened the titles and abstracts of 473 research papers, see Fig. 2.

In the first step, 50 (random 10%) articles were assessed for inclusion/exclusion by each reviewer for the purpose of cross-referencing. Through group discussion, consensus was reached about which articles to include: articles that demonstrated the important factors associated with sharing health data digitally. We then divided the remaining articles among NS and JVJ, who screened the rest of the titles and abstracts for inclusion individually. Thereafter, we selected 98 articles for full text review after which 35 were excluded for not being relevant to the topic of governance mechanisms for sharing data, as they were opinion articles, or were not accessible. There were three categories of articles that we identified after the initial review of abstracts and titles: 1) identification and discussion of governance mechanisms for data sharing and secondary uses of data, 2) assessment of people's/expert's attitudes and behaviours towards data sharing/secondary data use, include articles about consent preferences, 3) technical articles about development of software that helps to address governance issues for sharing of data.

Finally, data were extracted from 63 articles. To explore the complexity of how people's preferences change in different settings, we employed Nissenbaum's theory of contextual integrity. Nissenbaum's theory highlights the role that context plays in our privacy choices. Different contexts (the medical doctor office or the internet) are regulated differently and people may have very different expectations and preferences with regard to privacy [19]. The data extraction was informed by Nissenbaum's contextual integrity theory as each included article was read. The information about the type of data, the actors involved, the context in which the data originated and in which it was shared, and the governance mechanism discussed were extracted into a matrix. The outcomes of the research articles were analysed to derive the overall attitudes towards, behaviours related to, and consequences of

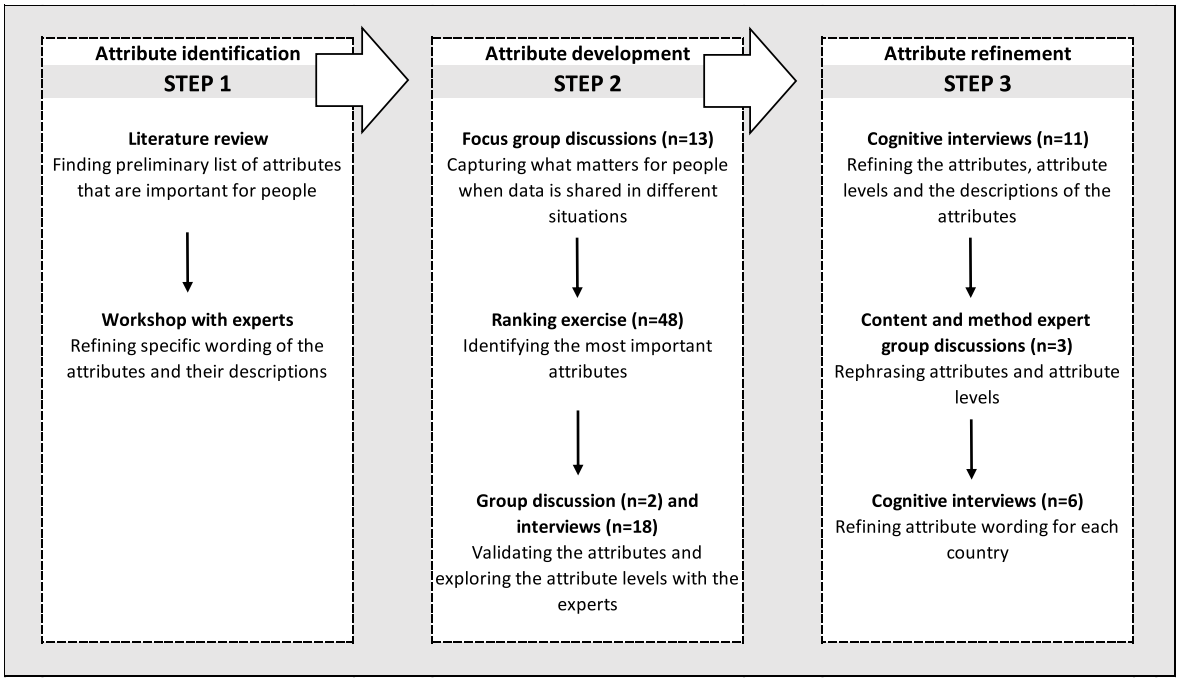


Fig. 1. A process map of the three steps in the development of attributes for a discrete choice experiment.

sharing health-related data in different contexts. The initial list of attributes was drafted and discussed with members of our research team and during an online meeting with two experts on privacy issues in data sharing (see Table 1).

2.2. Attribute development - Step 2

2.2.1. Focus groups and ranking exercise with the target population

In the second step, the focus was on further capturing what mattered to citizens. Focus group discussions were conducted in Sweden, Iceland and the UK between June and September 2019. All procedures were in accordance with the 1964 Declaration of Helsinki, as revised in 2013, and ethical approval was obtained in all participating countries. Informed consent for research participation and consent for data processing was obtained from all participants prior to being included. Materials were presented and discussions were held in the official language of each study country.

In total, we conducted 13 focus group discussions (n = 6 in the UK, n = 4 in Sweden, n = 3 in Iceland). Overall, there were 71 participants in the study (n = 32 in the UK, n = 17 in Sweden, n = 22 in Iceland).

Focus groups are useful for gathering information from lay people about complex topics, and observing the dynamics amongst participants when they express their attitudes and options [20,21]. Open-ended questions were posed to the group from a semi-structured interview guide. The interview questions were formulated with the goal of capturing the individual's perspective on sharing health data digitally. We used a vignette strategy approach, where the group followed the journey of a fictional person in three scenarios involving the sharing of health data. The vignette strategy was selected because we wanted to capture the attitudes and potential attitudinal changes in the different data scenarios. In addition, it was meant to aid participants' understanding of the ways in which health data may flow between data processors and third parties. The entire research team was involved with considering question topics that would be helpful to prompt discussion and the formulation and order of the questions.

To capture differences in experiences with sharing health data among participants, we applied a purposeful sampling strategy with participants in different age groups (minimum age 18 years old), different educational levels, and people from urban and rural areas. We used snowball recruitment (Sweden), random selection from register (Sweden and Iceland) and a multimodal approach in England using snowballing through community links and workshops, and advertising in different locations to get maximum variation among the respondents, with each session lasting approximately 90–120 min. The focus group discussions were recorded and transcribed verbatim. Identification of possible attributes and levels was carried out by JVJ, NS, and EH using ATLAS.ti and Excel software. To avoid the introduction of errors that translation might bring, transcripts of the English focus groups were discussed amongst JVJ, NS, and EH until consensus was found in a coding strategy. Non-English transcripts were reviewed by native speakers in the respective study countries. The grouped material of the possible attributes were thematically analysed by JVJ: initial reading followed by line-by-line coding and sub-grouped into attribute and levels [22]. Researchers identified attributes by considering: 'What in this material describes important factors for sharing health data, and how are they described?' The descriptions of the attributes were then discussed by all authors. The descriptions presented by participants were used to inform us of the amount of levels and wording of the levels for the DCE.

To identify the most important attributes, Nominal Group Technique [23] was applied during the end of the focus groups. Participants were asked to perform a ranking exercise in the last 15 min of each focus group session. They were asked to rank order attributes from a pre-defined list (from Step 1) from most to least important (see Appendix 2 in Supplementary material). Participants were given the opportunity to add to the list of pre-defined attributes if they thought any attribute was missing (see Table 1). After the ranking, a discussion was held around the reasoning for each participant's top three rankings. Participants who did not want to take part in the ranking exercise or who misunderstood the task were excluded (n = 9 in the UK, n = 2 in Sweden, n = 2 on Iceland). After collecting rankings from all groups, we calculated both a total mean score of all attributes and the frequency in percentage of the top three most important attributes for each country.

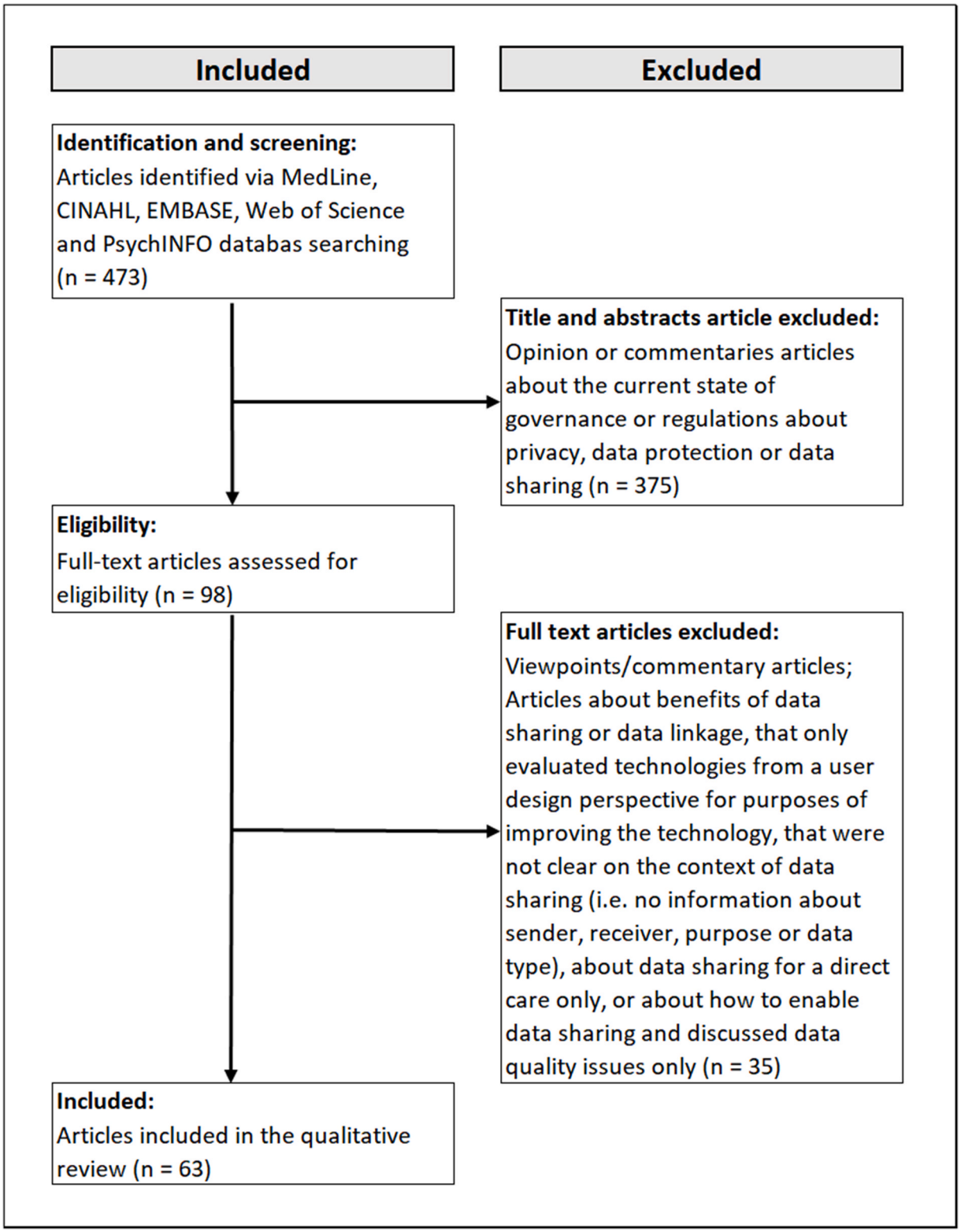


Fig. 2. Literature search flowchart of attribute identification – Step 1.

2.3. Attribute refinement - Step 3

2.3.1. Expert discussion and cognitive interviews

By confirming the attributes identified in step 2 and to explore attribute levels, roundtable discussions (n = 1 in England, n = 1 in Iceland) and individual interviews (n = 6 in the UK, n = 12 in Sweden) were conducted with experts. Experts were selected by differentiating the areas of expertise, including data management and project coordination, research, philosophy, law, policy making and journalism (see Appendix 3 for the full list in the Supplementary materials). Interviews followed a semi-structured interview guide with open-ended questions informed by the literature review and the expert discussion in step 1. The interview questions were formulated with the goal of capturing the experts' specific perspective and experience of sharing health data

Table 1
List of potential attributes and levels of the initial iteration of the DCE.

Attribute no.	Potential attributes	Potential levels
Attribute1	Type of information	name, DOB, education level, where I live, my ethnicity, political views, religion, sexual orientation, my interests, things I buy, people I see/am friends with, my health status, my mental health status, genetic information, medication, clinical data
Attribute2	Level of identification	identifiable data, de-identifiable data
Attribute3	Sender or collector that shares my data	companies, research, healthcare provider
Attribute4	Recipients of my data	companies, research, healthcare provider
Attribute5	Purpose of using the data	commercial advertising or marketing, medical research, policy makers, development such as quality improvement or service planning
Attribute6	Profit-making	companies profit, companies and data subjects profit, data subjects profit
Attribute7	Who is benefitting when I share the data	individual benefit, benefits for user of data, societal benefits
Attribute8	Oversight	ethics committee, data access committee, national authorities, governmental ministries or departments or groups
Attribute9	Consent	dynamic, broad, T&C consent, re-informed consent
Attribute10	Soft-law	non-binding agreements between parties - usually international declarations, recommendations, and resolutions
Attribute11	Hard-law	high monetary penalties, other strong penalties of accountability, legally enforced in front of a court
Attribute12	Public Engagement	public outreach, social marketing, involvement in design of policies or representation on boards, surveys about preferences of data sharing, other
Attributes that UK participants added during one focus group		
Attribute 13	Transparency	
Attribute 14	Data Management	

digitally.

As a next step, an academic seminar with experts on ethics, a webinar with experts in cybersecurity and privacy research, and a two-day workshop with stated preference method experts were organised to further discuss the outcomes, and to make decisions on the attributes relevant for the DCE as well as how they would be described.

Finally, we rephrased and refined the attributes and levels by using cognitive interviews (n = 2 in the UK, n = 5 in Sweden, n = 4 in Iceland) with citizens. Cognitive interviews offer the opportunity to identify areas or descriptions that respondents do not understand or find confusing. The interview guide included both think-aloud and probing techniques [24,25]. The participants were asked to say out loud what they were thinking as they completed the questionnaire. The interviewers observed and took notes when the participant became surprised, hesitant, skipped a question or if something was difficult to understand. Probing questions were asked: What are you thinking now? Why did you make this choice? Would you choose that if I weren't here? Are you considering all of the information presented? Finally, we asked an evaluation question at the end of the interview. We modified the instrument and refined the attribute wording for each country based on the cognitive interviews. Researchers took notes and shared summaries of participants' feedback, adjusting the survey accordingly to aid understanding. Changes were made at this stage related to the wording used, spelling, length of the scenario, and attribute description and the order of the descriptions.

3. Results

3.1. Attribute identification - Step 1

Previous empirical research indicates that peoples' willingness to share data for secondary use is dependent on what type of data is being linked, the level of identification, and the new purpose for the data being shared [3,26,27]. A study that investigated public preferences regarding data linkage for health research showed that the type of information shared is the most important factor to be considered [4]. Other studies showed instead that the new purpose for which data is to be used is an important factor. People are less willing to make data available to commercial companies and insurance companies whose purposes for using the data may be unclear or not aligned with public expectations [27–29].

Besides the long list of potential attributes, this step helped us differentiate between different governance mechanisms, e.g. soft and hard law, consent and review of transfer of data. In a manner that was relevant to participants, 12 potential attributes were further examined in Step 2.

3.2. Attribute development - Step 2

The seven attributes that were most important to citizens, on average, across the countries were: 1) 'level of identification', 2) 'the purpose of data use', 3) 'type of information', 4) 'consent', 5) 'new data user', 6) 'health information collector' and 7) 'overview of data sharing'. Across the sample, participants ranked 'purpose', 'level of identification' and 'consent' as the three most important attributes, see Fig. 3. Findings were generally consistent across countries. However, participants from Sweden ranked the 'purpose of data use' as the most important attribute, while 'level of identification' was the most important for participants from Iceland and the UK. The attribute 'consent' was considered more important for participants from Iceland than participants from the UK and Sweden (Fig. 4).

Substantive emphasis was placed on trust; if the collector and the new user can be trusted, participants did not see issues with sharing. Being informed or knowledgeable about the sharing was also discussed before the ranking exercise was introduced.

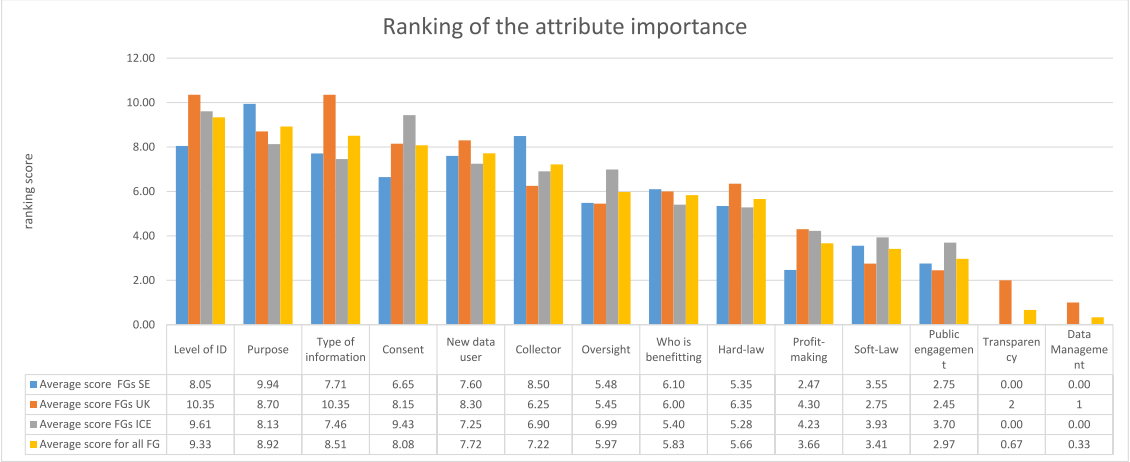
When provided a list of pre-defined attributes, only one group in the UK added two attributes ('transparency' and 'data management'), although these were later ranked low in that group.

It was difficult for citizens to discern the significance of attributes related to different regulations (hard and soft law). The participants thought that appropriate laws on data sharing were a prerequisite for data sharing in society at large.

Below, we outline participants' conceptualisation of the chosen attributes, and what we learned for our attribute development phase.

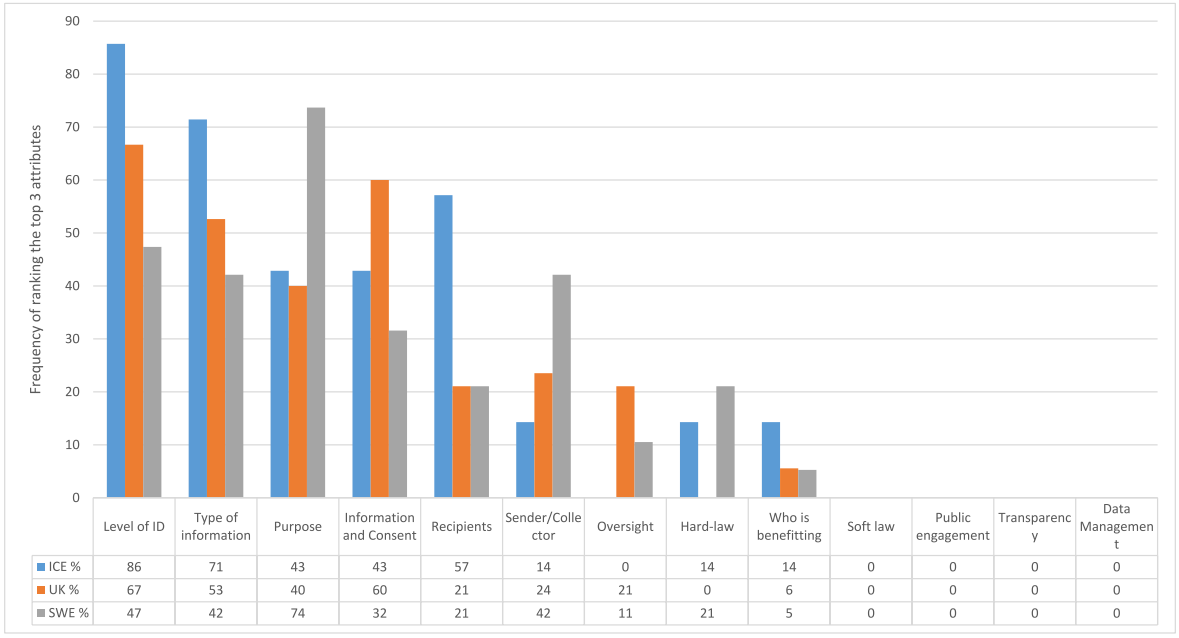
3.2.1. The purpose of data sharing

The 'purpose of data sharing' was an aspect frequently discussed in the focus groups. When considering the important attributes for decision-making about whether health data should be shared or not, participants consistently returned to the attribute of purpose. Participants would allow for data sharing as long as data are shared for good and reasonable purposes – good being defined as for the betterment of individual or public health, or to improve and resolve problems in society. However, participant discussions revealed that their



	Level of ID	Purpose	Type of information	Consent	New data user	Collector	Oversight	Who is benefitting	Hard-law	Profit-making	Soft-Law	Public engagement	Transparency	Data Management
Average score FGs SE	8.05	9.94	7.71	6.65	7.60	8.50	5.48	6.10	5.35	2.47	3.55	2.75	0.00	0.00
Average score FGs UK	10.35	8.70	10.35	8.15	8.30	6.25	5.45	6.00	6.35	4.30	2.75	2.45	2	1
Average score FGs ICE	9.61	8.13	7.46	9.43	7.25	6.90	6.99	5.40	5.28	4.23	3.93	3.70	0.00	0.00
Average score for all FG	9.33	8.92	8.51	8.08	7.72	7.22	5.97	5.83	5.66	3.66	3.41	2.97	0.67	0.33

Fig. 3. Citizens' average score on the ranking exercise. Higher points indicate that the factor is more important.



	Level of ID	Type of information	Purpose	Information and Consent	Recipients	Sender/Collector	Oversight	Hard-law	Who is benefitting	Soft law	Public engagement	Transparency	Data Management
ICE %	86	71	43	43	57	14	0	14	14	0	0	0	0
UK %	67	53	40	60	21	24	21	0	6	0	0	0	0
SWE %	47	42	74	32	21	42	11	21	5	0	0	0	0

Fig. 4. Frequency of the attribute that was ranked top three in the ranking exercise stratified by country.

considerations about acceptability of the purpose were more nuanced when this related to the greater good. Vague purposes having no clear justifications for data sharing were most likely to be unacceptable. Some participants expressed their disapproval when their data are used in a way that defines them, or to target advertisements.

We learned that having clear reasons for data use are important for the participants, and therefore would need to be part of the DCE further inquiry. Levels for further consideration of this attribute were: develop a new product or service, advertisement and improvement of healthcare.

3.2.2. Level of identification

Across all participants, the level of identification was also a key attribute in their decision regarding whether to share data. Participants expressed a greater willingness to share information about themselves if they were anonymous in the data set. Some participants had a high expectation to remain anonymous; one participant expressed that it is the crucial point in the decision regarding whether to share information. Other participants expressed that they would be very uncomfortable if they could be identified personally in the data. As a result, participants were not willing to have their personal information such as name, social security number, and home address be linked to their health data, as such information would increase their identifiability. Words that participants used as synonymous with anonymisation were: being unable to be identified, to be de-identified, to be anonymous, to be de-coded, not being traceable.

We learned that for the participants, the level of identification was crucial for sharing health data at all. For that reason, we believed this attribute might dominate people's decisions. Therefore, we decided to set this attribute at a constant fixed level for all choice tasks (choice context variable). Moreover, anonymity as to lifestyle, health and genetic data is difficult to achieve; the choice is between personally identified and personally identifiable data. The DCE was designed so that before respondents answered the choice tasks, they were informed that all health information would be coded; meaning that personal information, such as name and address, would be replaced with a code. The participants were also informed that this coded data is also called 'pseudonymised data' and that those who collected the information has a code key that is kept separately and safely, that can be used to identify people if need be. But those receiving and using the data do not know the identities of the participants.

3.2.3. Type of health information

The participants were asked to consider different information types in situations where their information had been shared in the past. They reported the following types of personal information: social security number, bank account number, GPS location data, income, DNA, medical record, fingerprints, e-mail, blood sample, lifestyle data (step counts, exercise and eating behaviours, sleeping pattern, running details, etc.), diagnoses, measurements about their physical and mental health. Step-counting and eating habits were considered not sensitive to some participants. One participant expressed that no particular type of data is more or less sensitive if one cannot be identified. Another did not think that genetic data was more sensitive than other information. However, many expressed that for them, medical data is more sensitive than lifestyle data.

They perceived that there was a risk associated with, for instance, being identified as having a medical condition as this may be used to discriminate someone in the future, while one's lifestyle can change. Moreover, mental health data was viewed as more sensitive than other types of data due to the risk of being stigmatised. Some participants expressed uncertainty about what constitutes sensitive information. They mentioned that it is difficult to know all unforeseen consequences of the use of different data types; therefore, it was challenging to ascertain the levels of sensitivity. The participants expressed that all information types are potentially sensitive in nature.

These results led us to incorporate four types of information in our DCE: lifestyle, mental health, physical health and genetic risk.

3.2.4. Health information collector

Some participants had experiences of their data being collected and shared by different organisations, such as local government collecting municipal data to monitor health, welfare, crime rates, etc. However, some participants were not aware that data might be shared from the original collector with a new recipient. Examples of data collectors mentioned by participants included healthcare and non-healthcare institutions, commercial companies, public welfare, scientific institutions and research projects. Influence on data sharing preference seems to be based on trust towards the organisation and previous experience. The participants expressed greater trust in organisations that are large and known. They expressed that it is difficult to foresee all the future consequences of accepting to share information. Therefore, they expressed the need to trust the collector and the new user to some extent.

Three different health information collectors that people can primarily come across and would therefore suit the DCE were: technological company, an academic research project, and health care providers.

3.2.5. New data user

Participants reported having little to moderate levels of knowledge about who had access to their health data. They formed different opinions about the acceptability of the data users, depending on the type of user, for example, government departments versus a pharmaceutical company. Some participants assumed that the data remained in the collector's repository and was not shared with other users. Whether or not the new user was accepted was linked to their reason for processing the data. When discussing trust in data users, it was deemed better to share data with a technological company with a public benefit purpose, such as finding a way to diagnose an illness than with a research project that had no direct benefit for individuals or society. Examples of different new users that might receive participants' data included government health ministries, commercial companies, insurance companies, pharmaceutical companies, universities, research projects and healthcare providers.

These actors were conceptualised into four relevant typologies, reflecting different levels to characterise this attribute in the DCE: a technological company, a pharmaceutical company, an academic research project and a national authority.

3.2.6. Consent

Being informed about what happens with one's data was of paramount interest to participants. One viewpoint was that if a data collector or secondary user 'uses my data, they use me'. Participants indicated they preferred to actively consent to data sharing. If that was not an option, they indicated they wanted to know what happens to their data and be able to at least opt out if they object to their data being shared or reused. Regardless of whether the purpose was perceived to be good, being informed about what the data is used for was always preferred. Participants also expressed the importance of having clear information before deciding to consent. Therefore, the following levels were further considered for this attribute: not being informed, being informed, being informed and the possibility to opt out, and being informed and asked to consent.

3.2.7. Oversight of data sharing

There were two diverse expectations of who will and can control whether health data is used and shared appropriately.

One perception was that it is only 'I' that should be in full control regarding the use of data about me, by accepting or not accepting to share. The perception was that each individual should evaluate the specific situation. Some participants expressed that the government should not control our society completely in all issues. Examples of taking control as an individual include: adjusting phone settings, turning off microphone or cookies, leaving the phone at home when taking a walk, and asking more questions before consenting.

The other perception was that individuals tasked with the responsibility to check and control data sharing and use should be part of an independent organisation that could observe the type of data being collected, transfer processes and oversee/ensure appropriate re-use of data. The expectation was that it must be an entity that can be held accountable in case something unexpected, such as a legal breach, happens. For some participants, it was very important to ensure oversight by a recognised body. A recognised body means an institutional body such as the parliament, the state, the government, an external reviewer, or someone officially responsible with the power to stop the data use and sharing. The discussions often concluded that the responsibility ought to be on several levels: individual, collector of data, the rules and laws of the government, and someone reviewing the transfer and use of the data.

Based on this discussion and the fact that this attribute is part of our research question (i.e. to investigate the relative importance of different governance mechanisms), this attribute was included in the DCE even though it was ranked seventh in the ranking exercise. Levels that were examined further in the next step included the identity of the oversight agent: a relevant committee including legal experts, members of the community, physicians/scientists/data scientists, organisations collecting data/the sender, or the government/national authority.

3.3. Attribute refinement - Step 3

Two attributes were particularly highlighted as being difficult for citizens to understand. The attribute 'purpose' was interpreted in two ways when testing our first attributes together in a DCE. Some individuals interpreted it as the purpose of the company as opposed to the purpose of sharing and using people's data. Therefore, we changed the wording of that attribute to 'the reason for data use'.

A further challenge was the attribute 'oversight', which proved difficult for people to understand without more explanation. During the attribute development process, we considered to include who will conduct the oversight or who approves the transfer of the data. Following extensive discussions among experts in the field, we rephrased this to express the presence of a review process for new uses and/or a review for the transfer of data. We concluded that what matters for citizens is that there will be a review, and of less importance is who exactly performs the review. Therefore, the suggested levels were: no

Table 2
The final attributes, attribute levels and their descriptions.

Attributes	Levels
1. **Health information collector** Different collectors can collect health information. The different collectors are:	**1. A technological company** with which you have used a service, programme or application for your phone or computer. You may have used a service through the company's website, where you have entered information about yourself. Alternatively, you have downloaded an application to your phone and it has collected information about your health. **2. An academic research project** where you have participated and they have collected health information about you. **3.**Your **healthcare provider** (hospital or GP) who has collected health information about you regarding your care.
2. **Data user** Your health information will be shared to a new data user. This new recipient may be:	**1. A technological company** that develops health applications, which can be used to predict diagnoses. **2. A pharmaceutical company** that develops and manufactures new medicines. **3. An academic research project** that produces new knowledge by testing hypotheses and theories about human health. **4. A national authority**, e.g. the public health authority, information and commissioner's office, etc., which is responsible for the health of the population. They can track peoples' health through population registers to prevent disease.
3. **The reason of data use** This aspect describes the reason why the data user wants to have access to your health information. The different reasons may be:	**1.** The reason can be to **develop a new product or service.** It can be a medical device, a drug, or application for your phone, or a new health service or programme. **2.** The reason can be to **promote, advertise or market** their product or service to personalise communication. The reason may be to direct advertising to a specific target group for a new service or product. **3.** The reason can be to **investigate a policy initiative.** Your health information can provide a basis for a new policy initiative at a national level. It may be to improve services for a specific part of the population or to identify new preventive measures to improve public health. **4. Evaluate the quality** of the data user's product or service, and for planning how resources should be distributed in the future.
4. **Information and consent** This aspect is about whether you will be informed if your health information is being shared.	**1.** You will **not be informed** that health information about you is being shared and used in a new context. **2.** You will **be informed** that health information about you is being shared and used in a new context. **3.** You will **be informed** that health information about you is being shared and used in a new context as well as be told that you **can opt-out** **4.** You will **be informed and asked to consent** that health information about you is being shared and used in a new context.
5. **Review of data sharing** Before your data is shared, there might be a review of the reason and how the data user will store and use your health information. The data user needs to apply for access to the health information. The reviewer(s) make a decision based on national law.	**1.** There will be **no review** of the data sharing. **2.** A committee **will review the transfer** of your health information to the new context. **3.** A committee **will review the transfer** and **the use** of your health information in the new context.

review of the data sharing, a committee will review the transfer of the health information to the new context, and a committee will review the transfer and the use of the health information in the new context.

Another challenge was whether 'type of information' should be an attribute on its own or a choice context variable. Attributes in DCE studies should fulfil several criteria, including 'being tradable' [30]. This means that people use the levels of an attribute to trade against levels of other attributes across alternatives. For instance, people might be willing to choose a treatment that is less effective because the risk of side effects is smaller (i.e. respondents trade effectiveness of the treatment against the risk of side effects). The 'type of information' was not considered as fulfilling the criteria of tradability as respondents reported that this attribute was a prerequisite for their decision to share their data (i.e. some people would never share genetic data, but always share physical data). At the same time, this is an extremely important characteristic of the situation in which people might be willing to share information in the first place. For that reason, 'type of information' was included as a context variable that will vary per given set of choices and not within each choice task. Indirectly, this approach provides the possibility to investigate if 'type of information' affects preferences for data sharing and governance mechanisms. See Table 2 for the final attributes and levels.

4. Discussion

In this paper, we show how we developed attributes and levels that are most important to citizens in the context of sharing health data digitally. This paper contributes to the cybersecurity debate by showing how we incorporated in the qualitative information into the DCE quantitative measure, through reflecting upon what truly matters to people when they need to make a decision including both privacy components and the utility of sharing health information.

Aitken et al. [4] investigated the public's preferences for sharing health information in the context of research, and identified appropriate attributes. Our results lead to the development of a similar attribute selection regarding who the new user of the data is, what type of information is shared, the purpose of data sharing, and the oversight of the process. In contrast, Aitken et al. included the attribute 'profit-making'. We considered including that attribute in our DCE, however, after discussions with the research team following the cognitive interviews in Step 3, we decided to exclude it. This attribute would lead to several design restrictions (i.e. non-profit new users would never make profit, while for-profit organisations would always make profit) and seemed to interact with 'the new data user' and 'the reason for data use'. For instance, if the new user is a technological company and the new use is to develop a new product or service, it is understood that the company will profit from the shared health information. Similarly, if health information is shared with healthcare providers for quality control, it is implicit that healthcare providers do not profit, but there is a benefit for the society and individuals. However, whether the actual transfer will involve a monetary exchange could still be a relevant attribute for new studies, even if that was not included in our DCE.

In line with our finding that participants expressed greater willingness to share information about themselves if data were anonymised,

previous studies indicated that individuals support data sharing for medical research, as long as the data are pseudonymised [31,32]. Many respondents also had quite high expectations for remaining anonymous. If the health information collectors and the new users can guarantee people's anonymity, and successfully manage to communicate this, it will facilitate data sharing. However, this is not a realistic expectation in many sharing situations in the big data era [33].

This study furthermore contributes to the methodology literature. This study reported on the three-step process for attribute and level selection for a DCE study, inspired by earlier studies that present the preparatory work for a DCE [30,34–36]. Currently, an established method for identification and development of attributes for DCE studies is lacking, even though a framework is recommended [17,37]. In the community of preference research, it is argued that the preparatory work for a more complex DCE study needs to be reported for the purpose of content validity and transparency of the upcoming DCE [38]. The attribute development process as well as its reporting need more attention as final attribute selection and wording emerges from small decisions in the process of developing a DCE. We agree with Helter and Boehler [37] that the overall process of attribute development, i.e., the step involved to move from raw data to final set of attribute (and levels), need to be presented. In this study, we performed an attribute selection approach in three steps: Step 1) Attribute identification, Step 2) Attribute development and Step 3) Attribute refinement. These three steps serve as a tool to distinguish the different phases that are required for attribute development. Step 1 serves the purpose of identifying all possible attributes. However, one must be careful to not only be guided by what has been done before. Therefore, the qualitative work in Step 2 is of utmost importance. This qualitative work in the target population ensures that the development of the attributes and levels is a bottom-up approach.

We suggest to always include a ranking exercise and calculate both the average ranking and the rank order frequencies per attribute. Such insights help to understand the total value of the attributes, and potential heterogeneity in preferences.

The last step is where the refinement of the attribute is implemented. This step is important to ensure the attributes are rightfully and clearly explained in the experiment, so that the target population can understand [30].

Contrary to other articles on preparatory work for DCEs [34,35,37], we want to emphasise the importance of closely monitoring and reporting on all decisions being made for attribute and level selection. This way we learn more about the phenomenon under study as well as the reflections of the researchers. A systematic approach for attribute development is desirable since such a process ensures researchers are transparent about all decisions behind selecting attributes and levels. In doing so, attributes and levels will be better contextualized, instead of taken for granted based on referencing previous published literature.

4.1. Limitations

A limitation of the current study is that whilst participants were recruited to achieve maximum variation, we anticipated that the sample was self-selected, had achieved moderate to high levels of education, and were fluent in the languages spoken in the study countries. Further research would be warranted to be inclusive of least studied groups and those from diverse socio-demographic backgrounds to understand better their unique experiences of the digital world and data sharing as defined in this paper.

DCE is a method that reflects reality more than using Likert-scale statements. However, as DCEs measure preferences in hypothetical scenario's, these stated preferences might differ from revealed or true preferences. Measuring revealed preferences therefore usually is considered preferable. However whenever this is not possible, DCEs might offer a valid alternative [39–41].

4.2. Conclusion

This study contributes to better understand decision-making of the general public about sharing their health data for reuse in different contexts. Further, the study identified a set of potential attributes for the development of a future discrete choice experiment. The three attributes that were most important, on average, in all three study countries were: 1) the purpose of data use, 2) level of identification and 3) the provision of information and consent.

Even more importantly, this study demonstrates a process to select attributes for a multi-country DCE involving three stages: Attribute identification, Attribute development and Attribute refinement. We presented arguments for publishing the process of selecting attributes and levels for a DCE. In doing so, future DCE studies would improve in terms of reliability, content validity and trustworthiness. Researchers developing a DCE or aiming to understand if a DCE would be an appropriate methodology to use in a field in which it has not previously been used, can lean on the process we describe for their own work.

Sample credit author statement

Jennifer Viberg Johansson: Conceptualisation, Methodology, Software, Data curation, Writing – original draft preparation, Visualization, Investigation. **Nisha Shah**: Conceptualisation, Investigation, Writing- Reviewing and Editing. **Eik Haraldsdóttir**: Conceptualisation, Investigation, Writing- Reviewing and Editing. **Heidi Beate Bentzen**: Conceptualisation, Validation, Writing- Reviewing and Editing. **Sarah Coy**: Conceptualisation, Investigation, Writing- Reviewing and Editing. Supervision: **Jane Kaye**: Conceptualisation, Validation, Writing-Reviewing and Editing. **Deborah Mascalzoni**: Conceptualisation, Validation, Writing- Reviewing and Editing. **Jorien Veldwijk**: Conceptualisation, Methodology, Software, Validation, Writing- Reviewing and Editing.

Acknowledgments

This work was supported by NordForsk [grant number 81105] and the Economic and Social Research Council (part of the UK Research and Innovation, United Kingdom). The funders had no influence on the design or writing of the article. Our greatest gratitude goes to all participants of the focus group discussions and the expert interviewees. Thank you for your time and for sharing your experiences and ideas generously. We would also like to acknowledge Nia Roberts, Outreach Librarian, Bodleian Health Care Libraries, for her help with the searching strategy of the literature review.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.techsoc.2021.101625.

References

[1] C. Petersen, V. Subbian, Special section on ethics in health informatics, J. Yearbk. Med. Inform. 29 (2020) 77.
[2] C.J. Haug, Whose data are they anyway? Can a patient perspective advance the data-sharing debate? J. New Engl. J. Med. 376 (2017) 2203–2205.
[3] M. Aitken, J. De St Jorre, C. Pagliari, R. Jepson, S. Cunningham-Burley, Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies, BMC Med. Ethics 17 (2016) 73.
[4] M. Aitken, G. McAteer, S. Davidson, C. Frostick, S. Cunningham-Burley, Public preferences regarding data linkage for health research: a discrete choice experiment, J Int. J. Popul. Data Sci. 3 (2018).
[5] N. Howe, E. Giles, D. Newbury-Birch, E. McColl, Systematic review of participants' attitudes towards data sharing: a thematic synthesis, J. Health Serv. Res. Pol. 23 (2018) 123–133.
[6] N. Shah, V. Coathup, H. Teare, I. Forgie, G.N. Giordano, T.H. Hansen, et al., Sharing data for future research-engaging participants' views about data

governance beyond the original project: a DIRECT Study, Genet. Med. 21 (2019) 1131–1138.
[7] V. Soekhai, E.W. de Bekker-Grob, A.R. Ellis, C.M. Vass, Discrete choice experiments in health economics: past, present and future, Pharmacoeconomics 37 (2019) 201–226.
[8] M.D. Clark, D. Determann, S. Petrou, D. Moro, E. de Bekker-Grob, Discrete choice experiments in health economics: a review of the literature, J. Pharmacoecon. 32 (2014) 883–902.
[9] E. de Bekker-Grob, M. Ryan, K. Gerard, Discrete choice experiments in health economics: a review of the literature, J. Health Econ. 21 (2012) 145–172.
[10] J. Viberg Johansson, S. Langenskiold, P. Segerdahl, M.G. Hansson, U.U. Hosterey, A. Gummesson, et al., Research participants' preferences for receiving genetic risk information: a discrete choice experiment, Genet. Med. 21 (2019) 2381–2389.
[11] D. Hensher, J. Rose, W. Greene, Applied Choice Analysis, second ed., Cambridge University Press, Cambridge, 2015.
[12] D. McFadden, Conditional Logit Analysis of Qualitative Choice Behavior, 1973.
[13] J.J. Louviere, G.J. Woodworth, Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data, J. Market. Res. 20 (1983) 350–367.
[14] M. Ryan, K. Gerard, M. Amaya-Amaya, Using Discrete Choice Experiments to Calue Health and Health Care, Springer, Dirdrecht, The Netherlands, 2010.
[15] Viberg Johansson, J., Bentzen, H.B., Shah, N., Haraldsdóttir, E., Jónsdóttir, G.A., Kaye, J., et al. Publics' Preferences for Sharing Health Data: a Discrete Choice Experiment. JMIR Med Inform. (2021) (In press).
[16] I.L. Hollin, B.M. Craig, J. Coast, K. Beusterien, C. Vass, R. DiSantostefano, et al., Reporting formative qualitative research to support the development of quantitative preference study protocols and corresponding survey instruments: guidelines for authors and reviewers, J. Patient-Patient-Center Outcomes Res. 13 (2020) 121–136.
[17] J. Bridges, A. Hauber, D. Marshall, A. Lloyd, L. Prosser, D. Regier, et al., Conjoint analysis applications in health-a checklist: a report of the ispor good research practices for conjoint analysis task force, Value Health 14 (2011) 403–413.
[18] A.C. Tricco, J. Antony, W. Zarin, L. Strifler, M. Ghassemi, J. Ivory, et al., A scoping review of rapid review methods, J. BMC Med. 13 (2015) 224.
[19] H. Nissenbaum, Privacy as contextual integrity, J Wash. L. Rev. 79 (2004) 119.
[20] J. Kitzinger, Qualitative research: introducing focus groups, J Bmj 311 (1995) 299–302.
[21] R.A. Krueger, M.A. Casey, Focus Groups: A Practical Guide for Applied Research, Sage publications, 2014.
[22] P. Burnard, P. Gill, K. Stewart, E. Treasure, B. Chadwick, Analysing and presenting qualitative data, Br. Dent. J. 204 (2008) 429–432.
[23] M. Hiligsmann, C. van Durme, P. Geusens, B.G.C. Dellaert, C.D. Dirksen, T. van der Weijden, et al., Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis, Patient Prefer. Adherence 7 (2013) 133–139.
[24] S. Cheraghi-Sohi, P. Bower, N. Mead, R. McDonald, D. Whalley, M. Roland, Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique, Fam. Pract. 24 (2007) 276–282.
[25] M. Ryan, V. Watson, V. Entwistle, Rationalising the 'irrational': a think aloud study of discrete choice experiment responses, Health Econ. 18 (2009) 321–336.
[26] E. Clayton, C. Halverson, N. Sathe, B. Malin, A systematic literature review of individuals' perspectives on privacy and genetic information in the United States, PloS One 13 (2018), e0204417.
[27] M. Shabani, L. Bezuidenhout, P. Borry, Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review, Expert Rev. Mol. Diagn. 14 (2014) 1053–1065.
[28] D. Goodman, C.O. Johnson, D. Bowen, M. Smith, L. Wenzel, K. Edwards, De-identified genomic data sharing: the research participant perspective, J. Commun. Genet. 8 (2017) 173–181.
[29] R.D. Jones, A.N. Sabolch, E. Aakhus, R.A. Spence, A.R. Bradbury, R. Jagsi, Patient perspectives on the ethical implementation of a rapid learning system for oncology care, J. Oncol. Pract. 13 (2017) E163–E175.
[30] J. Coast, H. Al-Janabi, E. Sutton, S. Horrocks, A. Vosper, D. Swancutt, et al., Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations, Health Econ. 21 (2012) 730–741.
[31] D. Kaufman, J. Murphy, J. Scott, K. Hudson, Subjects matter: a survey of public opinions about a large genetic cohort study, Genet. Med. 10 (2008) 831–839.
[32] K. Spencer, C. Sanders, E.A. Whitley, D. Lund, J. Kaye, W.G. Dixon, Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study, J. Med. Internet Res. 18 (2016) e66.
[33] M. Enserink, G. Chin, The end of privacy, J. Sci. 347 (2015) 490–491.
[34] C.S. Bennette, S.B. Trinidad, S.M. Fullerton, D. Patrick, L. Amendola, W. Burke, et al., Return of incidental findings in genomic medicine: measuring what patients value-development of an instrument to measure preferences for information from next-generation testing (IMPRINT), Genet. Med. 15 (2013) 873–881.
[35] E. Louis, J.M. Ramos-Goñi, J. Cuervo, U. Kopylov, M. Barreiro-de Acosta, S. McCartney, et al., A qualitative research for defining meaningful attributes for the treatment of inflammatory bowel disease from the patient perspective, J. Patient-Patient-Center Outcomes Res. (2020) 1–9.
[36] A. Rydén, S. Chen, E. Flood, B. Romero, S. Grandy, Discrete choice experiment attribute selection using a multinational interview study: treatment features important to patients with type 2 diabetes mellitus, J. Patient-Patient-Center Outcomes Res. 10 (2017) 475–487.
[37] T.M. Helter, C.E.H. Boehler, Developing attributes for discrete choice experiments in health: a systematic literature review and case study of alcohol misuse interventions, J. Subst. Use 21 (2016) 662–668.
[38] E.M. Janssen, J.F.P. Bridges, Art and science of instrument development for stated-preference methods, Patient-Patient Center Outcomes Res. 10 (2017) 377–379.
[39] E.W. de Bekker-Grob, J.D. Swait, H.T. Kassahun, M.C. Bliemer, M.F. Jonker, J. Veldwijk, et al., Are healthcare choices predictable? the impact of discrete choice experiment designs and models 22 (2019) 1050–1062.
[40] M.S. Lambooij, I.A. Harmsen, J. Veldwijk, H. de Melker, L. Mollema, Y.W. van Weert, et al., Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared 15 (2015) 1–8.
[41] B.H. Salampessy, J. Veldwijk, A.J. Schuit, K. Van Den Brekel-dijkstra, R.E. Neslo, G. A. De Wit, et al., The predictive value of discrete choice experiments in public health: an exploratory application 8 (2015) 521–529.