

# Who tweets what about personalised medicine? Promises and concerns from Twitter discussions in Denmark

Digital Health  
Volume 9: 1–12  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076231169832  
journals.sagepub.com/home/dhj



Lea Skovgaard  and Anders Grundtvig

## Abstract

Digital health data are seen as valuable resources for the development of better and more efficient treatments, for instance through personalised medicine. However, health data are information about individuals who hold opinions and can challenge how data about them are used. Therefore it is important to understand public discussions around reuse of digital health data. Social media have been heralded as enabling new forms of public engagement and as a place to study social issues. In this paper, we study a public debate on Twitter about personalised medicine. We explore who participates in discussions about personalised medicine on Twitter and what they tweet about. Based on user-generated biographies we categorise users as having a ‘Professional interest in personalised medicine’ or as ‘Private’ users. We describe how users within the field tweet about the promises of personalised medicine, while users unaffiliated with the field tweet about the concrete realisation of these ambitions in the form of a new infrastructure and express concerns about the conditions for the implementation. Our study serves to remind people interested in public opinion that Twitter is a platform used for multiple purposes by different actors and not simply a bottom-up democratic forum. This study contributes with insights relevant to policymakers wishing to expand infrastructures for reuse of health data. First, by providing insights into what is discussed about health data reuse. Second, by exploring how Twitter can be used as a platform to study public discussions about reuse of health data.

## Keywords

Personalised medicine < personalised medicine, twitter < media, public debate, reuse of health data, Denmark

Submission date: 26 September 2022; Acceptance date: 29 March 2023

## Introduction

Digital health data in the form of electronic medical records, registry data, genetic information and patients’ self-collected data are seen as valuable resources for the development of better and more efficient treatments, for instance through personalised medicine.<sup>1</sup> With the aim of individualising prevention, diagnostics and treatment based on multiple forms of digital data,<sup>2,3</sup> personalised medicine is heralded as having the potential for delivering better treatment at lower cost.<sup>1</sup> The introduction of personalised medicine rests on the availability of digital health data from a large part of the population, to be used for research and for comparison to the treatment of other patients. However, the collection and reuse of health data for secondary purposes raise questions about data security, risk of misuse of health data by private

companies or employers, as well as risk of violations of individuals’ autonomy if health data are used without their knowledge or for purposes they find inappropriate.<sup>4,5</sup> In recent years, there has also been an increased focus on how the collection of health data can have adverse effects through digital profiling<sup>4</sup> or by increasing inequality in healthcare.<sup>6,7</sup> In some cases, publics, consisting of the people the data concern, ‘bite back’ if data are used in

Department of Public Health, University of Copenhagen, Copenhagen K, Denmark

### Corresponding author:

Lea Skovgaard, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, Copenhagen K 1014, Denmark.  
Email: lsk@sund.ku.dk



ways they find inappropriate.<sup>8</sup> Public issues around the use of health data can ultimately jeopardise the opportunity to reuse health data for secondary purposes.<sup>9</sup> This shows the importance of exploring issues about reuse of health data to better enable the development of legitimate governance. In this paper, we explore issues around personalised medicine in Danish on Twitter. We explore *who* tweets *what* about personalised medicine. Accordingly, this study contributes with empirical insights into what is discussed about personalised medicine among Danish Twitter publics.

There has been an increasing interest in including publics' perspectives in decisions relating to development of science and technologies,<sup>10</sup> including in relation to personalised medicine.<sup>11,12</sup> The aim has been to make science more democratic<sup>13,14</sup> and to produce more robust science.<sup>15,16</sup> In this article, we explore how personalised medicine is discussed among people who themselves choose to express their perspectives.<sup>17</sup> We talk about publics instead of 'the public' because we see neither publics nor issues as constant entities in society. As we see it, multiple publics exist and these change over time. These publics consist of changing groups of people who assemble in response to issues and through different modes of participation.<sup>14,18,19</sup> In this article, we explore both what is discussed on Twitter in relation to personalised medicine and which publics gather on Twitter as a reaction to issues related to personalised medicine.

Previous research gives insight into how reuse of health data is discussed in news media and point to a tendency towards framing new developments positively. In Iceland, studies have described how the privately owned database deCode came under public scrutiny in public debates and through news media.<sup>20,21</sup> Hjörleifsen and colleagues describe that despite the public debate, media portraits of the deCode database mainly focused on the potential benefits with regard to medical and technical advancements as well potential economic benefits, while ethical and epistemological questions were less present.<sup>20</sup> An excessive focus on advancement in media coverage has also been described in relation to personalised medicine. In a study of how national daily news media in the UK covers news about personalised medicine and pharmacogenetics, Almomani and colleagues argue that newspaper articles overstate the benefits while writing less about potential risks.<sup>22</sup> In a Danish context, scholars have explored a public debate about personalised medicine in written Danish news media with a specific focus on the rationalities of the actors in the debate.<sup>23</sup>

In the past two decades, the use of digital methods to map issues as they appear on social media has been consolidated.<sup>24–26</sup> There has also been an increased interest in using social media to predict and study health related behaviour through 'inveillance'.<sup>27</sup> Digital methods are applauded for their ability to combine features of qualitative and quantitative research,<sup>28,29</sup> thereby allowing for a richer

analysis. Twitter is by now a commonly used platform to study issues and the platform has been heralded as democratic and as a new way of facilitating public engagement<sup>30,31</sup> since it allows publics to gather *ad hoc* in reaction to issues they find important.<sup>32</sup> Twitter has been used to study a broad range of issues spanning COVID-19,<sup>33,34</sup> breast cancer,<sup>35</sup> diabetes,<sup>36</sup> climate change<sup>37</sup> and sexual victimisation,<sup>38</sup> to mention only a few. Twitter has also been used to study a dispute around the care data scheme in the UK, a database set up to enable reuse of health data from primary care.<sup>39</sup> We still lack knowledge about how personalised medicine is discussed on Twitter. With this article, we add to the literature about how reuse of health data is discussed publicly, by exploring tweets in Danish about personalised medicine. We do this by studying aggregated frequency of specific hashtags and nouns. Based on user-generated biographies we categorise Twitter users as having a 'Professional interest in personalised medicine' or as 'Private' users. While the users working within the field, tweet about the promises of personalised medicine, users unaffiliated with personalised medicine tweet their concerns about the actual conditions for the implementation of personalised medicine.

Denmark serves as an interesting case for studying what is discussed in relation to reuse of health data because of the already high level of digitisation in the public sector<sup>40</sup> and Denmark's long history of using health data for purposes other than treatment, e.g., for research and planning.<sup>41</sup> A part of the introduction of personalised medicine in Denmark was the establishment of a new infrastructure, the National Genome Center, which was set up to centralise storage of genetic information as well as facilitate reuse of this information in research and for clinical purposes.<sup>42</sup> The National Genome Center accordingly was the pivotal point in the strategy to realise personalised medicine. The establishment of this infrastructure required changes in the Danish legal framework for the storage of genetic information. The suggested legal changes were debated diligently in the Danish news media,<sup>23</sup> and the legislation which was finally passed included an addition concerning patients' ability to opt out of the storage of genetic information,<sup>43</sup> in contrast to the legislation which was originally proposed.<sup>44</sup> Personalised medicine in Denmark is assuredly promoted as being offered by the public sector<sup>45,46</sup> yet the implementation rests on funding from both public and private sectors. In 2018 a private foundation invested 1 billion Danish kroner in the establishment and development of the National Genome Center<sup>47</sup> to be used in the period 2020–24.<sup>48</sup> No private funding for the further operations of the centre has been announced. There will continue to be public funding for future operations, however only to a limited degree.<sup>48</sup> Consequently, the future of personalised medicine in Denmark is currently characterised by uncertainties with regard to the future costs of running the National Genome Center, and how this should be divided between public and

private funds. Accordingly, there have been two periods of political indeterminacy, firstly in 2018, and secondly in 2022, since when the governance of personalised medicine in Denmark is again open for negotiation.

Understanding what is discussed, and by whom in a country where reuse of health data to a large part goes uncontested, can be helpful for other countries wishing to expand data infrastructures facilitating reuse of health data in similar ways. With this article, we contribute to the literature about how reuse of health data is discussed publicly. We do this by exploring which topics are discussed in relation to personalised medicine on Twitter and by exploring who becomes engaged through Twitter. We start by describing our methodology. Afterwards, we present our empirical material showing *what* is discussed on Twitter by *whom*. We end by arguing that Twitter should not be seen as a bottom-up democratic forum, despite the fact that the platform facilitates debate from different publics.

## Methodology

### Data collection

To explore what is discussed on Twitter we collected tweets about personalised medicine in Danish. Twitter is a platform which allows users to make posts (tweets) with a maximum of 240 characters. Users can follow other users and share their posts (retweets).<sup>49</sup> Twitter allows users to apply hashtags to their posts as brief summaries of the content of the tweet.<sup>50</sup> We collected data with Twitter's Academic Access API which allows researchers to access all tweets in Twitter's database.<sup>51</sup> The search query was developed through an iterative process of adding and removing potential search terms. The initial search query was inspired by the search terms used by Skovgaard and Hoeyer (2022)<sup>52</sup> in their study of public debate in news media about personalised medicine. We collected tweets using their search terms and studied the most frequent hashtag used in the dataset. Based hereon we added hashtags related to reuse of health data and personalised medicine to the search query. We studied the amount of tweets each search term would add to the dataset. Then we manually went through a random sample of the tweets collected through the different search words to study whether the tweets were interesting for discussions about personalised medicine or whether the tweets were mainly discussing other topics. This process led us to the search query: 'personlig medicin OR personligmedicin [personalised medicine] OR genomcentre OR genom centre [genome centre] OR præcisionsmedicin OR præcisions medicin [precision medicine]'.

Tweets were collected from the period 3rd November 2009, where the first tweet using one of the words from our search query was posted, until 30th March 2022. We

included all tweets using words from the search query to be able to explore developments in discussions about personalised medicine over time including changes in the frequency of tweets using the search terms.

The data collection had two steps:

1. First, we collected all tweets corresponding to our search query (sub-dataset A). This includes tweets, retweets, replies and the corresponding metadata, that contained one of the search terms in the textual content of the tweet. We used Twitter's language identification algorithm to limit the dataset to tweets written in Danish.
2. With the 'Conversation ID's' from sub-dataset A we collected all tweets posted as a response to each of the tweets in sub-dataset A and each of the responses to the replies and so on.

Sub-dataset A and sub-dataset B constitute what we refer to as 'the dataset'.

### Categorisation of Twitter users

To explore who discusses personalised medicine on Twitter we categorised user profiles as either having a professional interest in personalised medicine or as private users. Twitter users are prompted by Twitter to write a short description of themselves. These descriptions provide insights into the people behind the tweets. Inspired by the ongoing work of Vandeweerdt, Hjort and Dinesen (2022)<sup>53</sup> we used the self-generated descriptions to categorise users. We used a combination of manual and computational methods to do this. First, we created categories manually, based on our knowledge about actors in the field of personalised medicine. Then we used computational assisted categorising techniques to create categories based on user biographies. We did this by using the Natural Language Processing function, part of Speech Tagging, from the library 'Spacy'<sup>54</sup> to extract nouns and proper nouns from the user biographies. Then we used the topic modelling algorithm Latent Dirichlet Allocation from the library Gensim<sup>55</sup> to put these nouns and proper nouns into topics. The manual and computational assisted categories were compared and used to make three overall user categories: 'Professional interest in personalised medicine', 'Private' and 'Other'. Accordingly, the decisions to make these three categories were made manually by the authors with inspiration from the computationally made categories.

The category 'Professional interest in personalised medicine' contains both users identifying as individuals and as institutions. It consists of the subcategories 'Research and clinic', 'Governance' and 'Interest groups'. 'Research and clinic' includes users describing their affiliation with universities, or an affiliation with the healthcare system. The

category of ‘Governance’ covers institutions or individuals describing themselves as working at institutions, which are administering or governing the Danish welfare state: Politicians, Regions, Municipalities, and institutions facilitating use of health data such as the States Serum Institute and the National Genome Center. ‘Interest groups’ are patient organisations, industry working with biomedicine (including pharmaceutical companies and companies developing health technologies), and professional organisations related to the biomedical field. The group ‘Private’ consists of users specifying their professional status in areas not related to the ones described in the above categories, or users using words about leisure activities, and users who did not have any information in their biography. Users without a description in their biographies can have a professional interest in personalised medicine. However, if users do not themselves reveal their professional background – thereby allowing others to learn about their profession – we assume that they want their posts to be read as coming from a private profile, and we chose to categorise them accordingly. Finally, ‘Others’ consists of users whose biographies describe them as journalists, and news media, as well as bots, and profiles set up to promote open positions.

After the users had been automatically categorised one author manually browsed all the biographies, to correct miscategorisation. We are aware of one user, a doctor who has been very active and critical in the debate in news media about personalised medicine, who did not write anything in their biography and therefore ended up being categorised as ‘Private’. We know this user through our previous interest in the topic even though we could not verify that it is actually their account. According to our criteria, placing this user in the ‘Private’ category is a miscategorisation. Yet, not all users on Twitter know them, and it is likely that their comments in the debate are read without knowledge about their profession. Therefore, we did not recategorise the user. We used this knowledge to study how their tweets affected the overall map and found that they were present in more critical parts of the map. We know of this one user but it is likely that other similar miscategorisations exist in the material of which we are not aware and have therefore not been able to explore in further depth.

### Issue mapping

In order to analyse what people are tweeting about personalised medicine we created two different network graph visualisations: one of the co-occurrence of hashtags and one of the co-occurrence of nouns in the tweets in the dataset. We included tweets, retweets and replies in the maps. We used the open-source tool Cortext Manager<sup>56</sup> for both term extraction and network graph mapping.

First, we extracted hashtags by limiting the maximal word length to one, allowing monograms, not applying linguistic preprocessing and only including words that start

with a hash ‘#’. Then we generated a co-occurrence network of the top 350 most used hashtags (including the search terms). We found 350 to be the most optimal balance between readability and complexity. In the network graphs, each hashtag is a node (triangle) and there is an edge (line) between the nodes if they co-occur (are used together) in a tweet. The placement of each node is determined by the force-directed spatialisation algorithm, Fruchterman Reingold, which forces nodes to cluster into groups if they are frequently used together in the tweets. This creates thematic clusters that are defined by being more related to each other than to the rest of the network. The size of the nodes represents how often they occur in the dataset. The colour of the nodes represents the grouping or cluster of the nodes detected by the Louvain community detection algorithm.

We first created a network graph depicting co-occurrence of hashtags in the tweets. Because only 48.7% of tweets use hashtags, we chose also to study the commonly used nouns from the tweets. We repeated the mapping exercise by extracting the 350 most used Danish nouns from the tweets (including the search terms). We call these maps of co-occurrence of hashtags or nouns ‘Base maps’. From the base maps, we identified clusters and analysed what we viewed as the prevailing theme(s). This was done both by looking at the wording of the hashtags and nouns and by studying tweets manually. Accordingly, we used the qualitative reading of the tweets to qualify our analysis of the network graphs, thereby using both the quantitative and the qualitative features of the data.<sup>29</sup> Next, we studied whether users from the different categories were more predominantly using hashtags or nouns from specific clusters. We did this by creating heat maps where we overlaid the base map with a heat map algorithm that colours the map. The heat map algorithm calculates a score for a third variable in the dataset, and colours the map according to how the nodes on the map correspond with the third variable. For the first map (Figure 1) we used the user categories as the third variable for colouring the heat map. The scale (right side of the figures) shows the third variable score. On this map, the score indicates how much more likely the hashtags or nouns are to be produced by users with professional interest or private, compared to the expected average.<sup>56</sup> For example, if the score is 2 in Figure 1(b) then there is twice the chance of the hashtag being produced by a private user than the expected average. This allows us to visually detect not just *what* people are discussing but also *who* is discussing what. The heat maps were also created by using the build-in ‘Network Mapping’ function in CorText where the process described above was repeated with an additional chi2 heat map algorithm to the map.

Discussions on Twitter are multi-faceted and include content from various user-types including social bots<sup>57</sup> and in our case, news media sites and job facilitating profiles. Content from these profiles is inherently a part of how issues are discussed on Twitter<sup>58</sup> but it is important to

explore how such profiles affect the debate. We looked for bots both qualitatively and quantitatively. Qualitatively, we read all user biographies and studied the network graphs including the tweets they compose of manually. Quantitatively, we checked for frequency and tweeting patterns in the dataset. Neither in the qualitative nor in the quantitative reading of the data did we observe bot-behaviour. We cannot dismiss that bot-generated content has been part of discussions about personalised medicine in Danish on Twitter but we consider it unlikely. In the following analysis, our main focus is on the categories 'Private' and 'Professional interest in personalised medicine' and we will only show heat maps from these two categories. We choose also to include tweets from users in the 'Other' category in the base maps because we acknowledge that job facilitating profiles are part of how personalised medicine is discussed on Twitter, although we do not analyse this category closely.

Even though we are filtering for Danish tweets there are still some Swedish tweets in the dataset. One tweet in Swedish, which had many replies and many retweets, makes up its own cluster in one map (see Figure 2). The tweet is about the publishing of a book by Anna Ardin, a Swedish woman who accused Julian Assange of sexual assault.<sup>59</sup> The cluster is unrelated to the rest of the map and contains one tweet in Swedish, and several retweets of this. This shows that the language identification algorithm is imperfect and the dangers of assuming that issues are delimited through search words.

### Who discusses what about personalised medicine?

The dataset consists of 9841 tweets from 2003 distinct users in the period 3rd November 2009 to 30th March 2022. Out of the 2003 users discussing personalised medicine on Twitter, 880 were categorised as having a professional interest in personalised medicine, 961 as being private users, and 162 as 'Other'. The debate on Twitter facilitates participation of a public where more than half of the participants do not describe themselves with words indicating that they have a professional interest in personalised medicine. However, people with a professional interest in personalised medicine tweet more about the topic: users in the 'Private' group made 3589 tweets in the period and users in the group 'Professional interest in personalised medicine' made 5568 tweets. There is a remarkable difference in when users tweet about personalised medicine. In 2018, 3242 tweets were posted about the topic. Accordingly, a third of all the tweets in the dataset is from 2018. That is more than four times as many tweets than from 2021 where 748 tweets were posted. In the first three months of 2022, from which we have data, 101 tweets were posted. This could indicate that the establishment of the National Genome Center and the legislative changes it required in

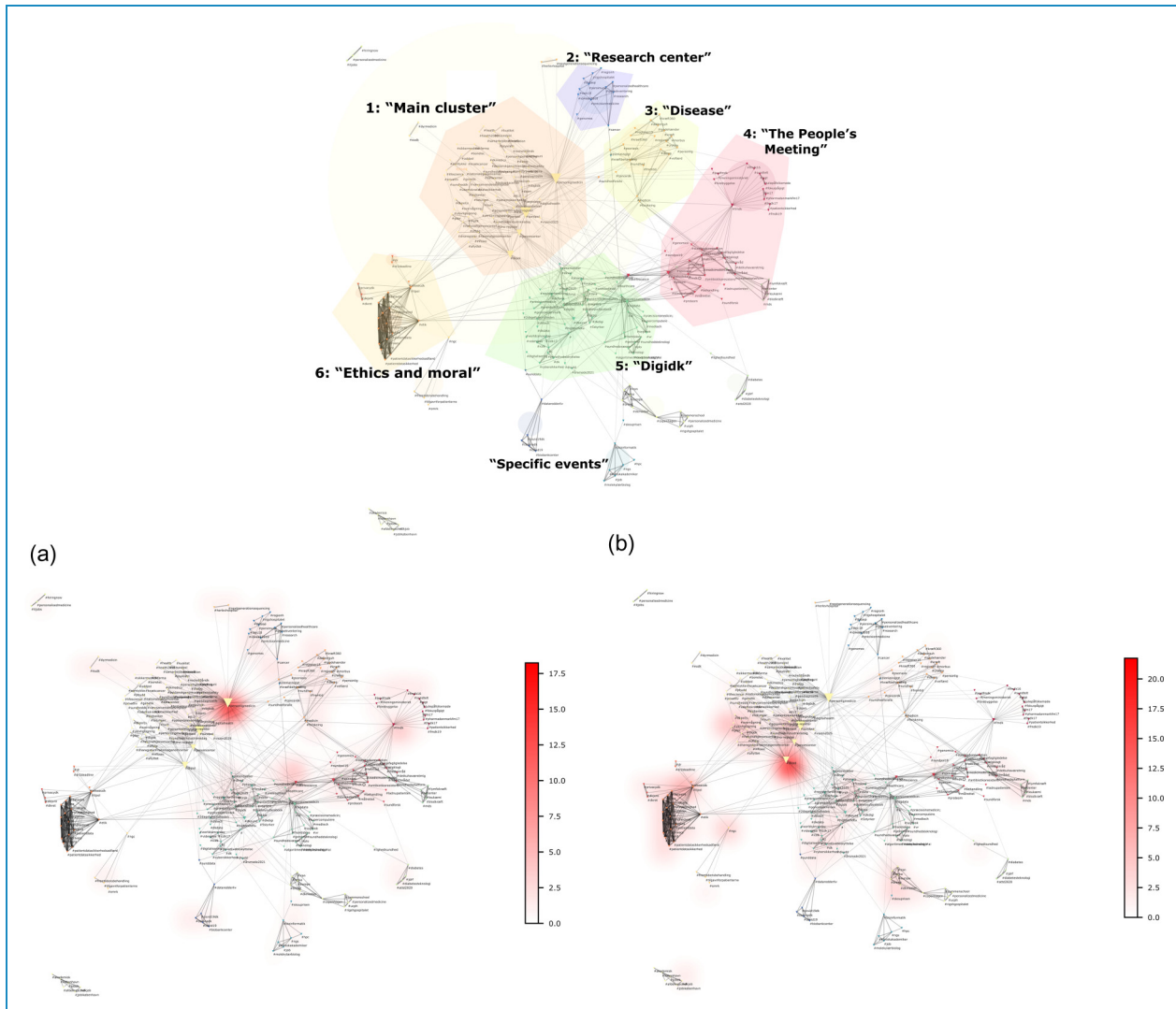
2018 mobilised tweets about personalised medicine while the current uncertainties regarding funding have not led to an increase in tweets about personalised medicine.

In the following, we will present the results from two network graphs: One based on co-occurrence of hashtags and one based on co-occurrence of nouns. We describe the content of the graphs focusing on the grouping of hashtags or nouns in clusters. We also present heat maps showing which hashtags or nouns are predominantly used by users from the different categories.

### Network graphs based on hashtags

Figure 1 shows a hashtag network of co-occurrence. Figure 1(a) and (b) show heat maps on the base map for hashtags. From the hashtag-clusters we learn how people we categorised as having a professional interest in personalised medicine tweet about the topic in relation to existing infrastructures and organisations. We also learn that it is predominantly users from the private category who apply hashtags which indicate concerns regarding privacy, surveillance, and consent in relation to personalised medicine.

From Figure 1 we have identified six clusters. Cluster one is 'The main cluster' where there is the most activity. In this cluster, we find our search word #personlig medicin (personalised medicine) and #genom center (genome centre) (in the various Danish spellings). In the cluster we also find the hashtags #sundpol used for topics relevant to health policy in Denmark and #dkpol used for political topics in Denmark, indicating that Twitter users see personalised medicine as a political topic or wish to make it a political topic. Interestingly, users we categorised as working with personalised medicine are more likely to be using the hashtag #personlig medicin (Figure 1(a)) whereas users categorised as private users use the hashtags #genom center or #nationalt genom center (Figure 1(b)). It seems that while users working within the field tweet about a concept, users from the private category tweet about the concrete physical realisation of the concept in the form of a new infrastructure. On the top right of the cluster, there are hashtags referring to organisations or infrastructures relevant to the implementation of personalised medicine. Examples of organisations are #dsam (professional organisation for general practitioners) and #danske patienter (patient organisation). Users apply these hashtags to either be in conversations with these actors or to clarify the stance of the organisation. Accordingly, in the top right of the cluster, the hashtags can be used to identify institutions and organisations which appear central in tweets about personalised medicine. From Figure 1(a) we see that it is predominantly users from the category 'Professional interest in personalised medicine' who use hashtags from this part of the cluster. On the left side of the main cluster, the hashtags #samtykke



**Figure 1.** Base map of co-occurrence of hashtags in tweets concerning personalised medicine in Danish. (a) Hashtag base map (Figure 1) with heat map layer (coloured areas) indicating which hashtag-clusters are predominantly used by users in the category 'Professional interest in personalised medicine'. (b) Hashtag base map (Figure 1) with heat map layer (coloured areas) indicating which hashtag-clusters are predominantly used by users in the category 'Private'.

(consent), #privatliv (privacy), #datasikkerhed (data security), #overvågning (surveillance), #udemokratisk (undemocratic) and #gdpr appear, which cover tweets expressing different kinds of concerns related to personalised medicine. Tweets from this side of the cluster tell us that people are concerned that the National Genome Center, and the collection of genetic information will be used for surveillance and breach privacy. It also shows us that there have been discussions around whether there should be consent to storage of genetic data in the National Genome Center, and how this should be formed. From Figure 1(b) we learn that it is mainly users we categorised as private who apply hashtags indicating these concerns.

Cluster two we call 'Research centre' because the hashtags circle around a specific research centre #persimune that does research in personalised medicine. The hashtags show that the centre is placed at #regionh (Capital Region of Denmark) at #rigshospitalet (hospital in the capital city of Denmark). Accordingly, this cluster focus on a specific institution which is a part of the introduction of personalised medicine in Denmark. Cluster three we call 'Disease' because it contains hashtags about how personalised medicine can be beneficial for specific diseases or disease areas such as #morbus and #crohn, #psorais and #kræft (cancer). Additionally, the 'Disease'-cluster contains hashtags about specific institutions such as the #regionerne (the Danish Regions), which are responsible for hospital

management in Denmark and #aalborguh (University Hospital of Aalborg) working with or doing research into these diseases. As apparent from Figure 1(a), it is predominantly users we categorised as having a professional interest in personalised medicine who use the hashtag from the ‘Research centre’-cluster and the ‘Disease’-cluster.

In cluster four, ‘The People’s Meeting’ (Folkemødet) we found hashtags relating to a recurring public event in Denmark, which aims at facilitating ‘democratic conversations’ on a variety of political topics.<sup>60</sup> The cluster has two parts: the top right part with general hashtags indicating participation in the event, sometimes with a specific year, e.g., #fm19. The bottom left part of the cluster is about specific topics or cases discussed, together with personalised medicine at specific events at ‘The People’s Meeting’. The tweets from these clusters are advertisements for events at ‘The People’s Meeting’ or users indicating that they have attended events about personalised medicine at the public meeting. Even though the aim of the event is to ‘decrease the distance between citizens and decision-makers’<sup>60</sup> we found that it was mainly people working with personalised medicine, which include decision-makers, who applied hashtags referring to the event.

The fifth cluster we call ‘Digidk’. This cluster circles around generic business and technology hashtags like #dkbiz, #dkdigi, #forskpol (research politics), #dktech, #dkdata. These hashtags are used for tweets about research politics, about research news, from users who have attended events, or from institutions promoting themselves in different ways. While the hashtags at the top of this cluster are mainly used by users in the category ‘Professional interest in personalised medicine’, the bottom of the cluster seems to be dominated by users we categorised as private (Figure 1(a) and (b)).

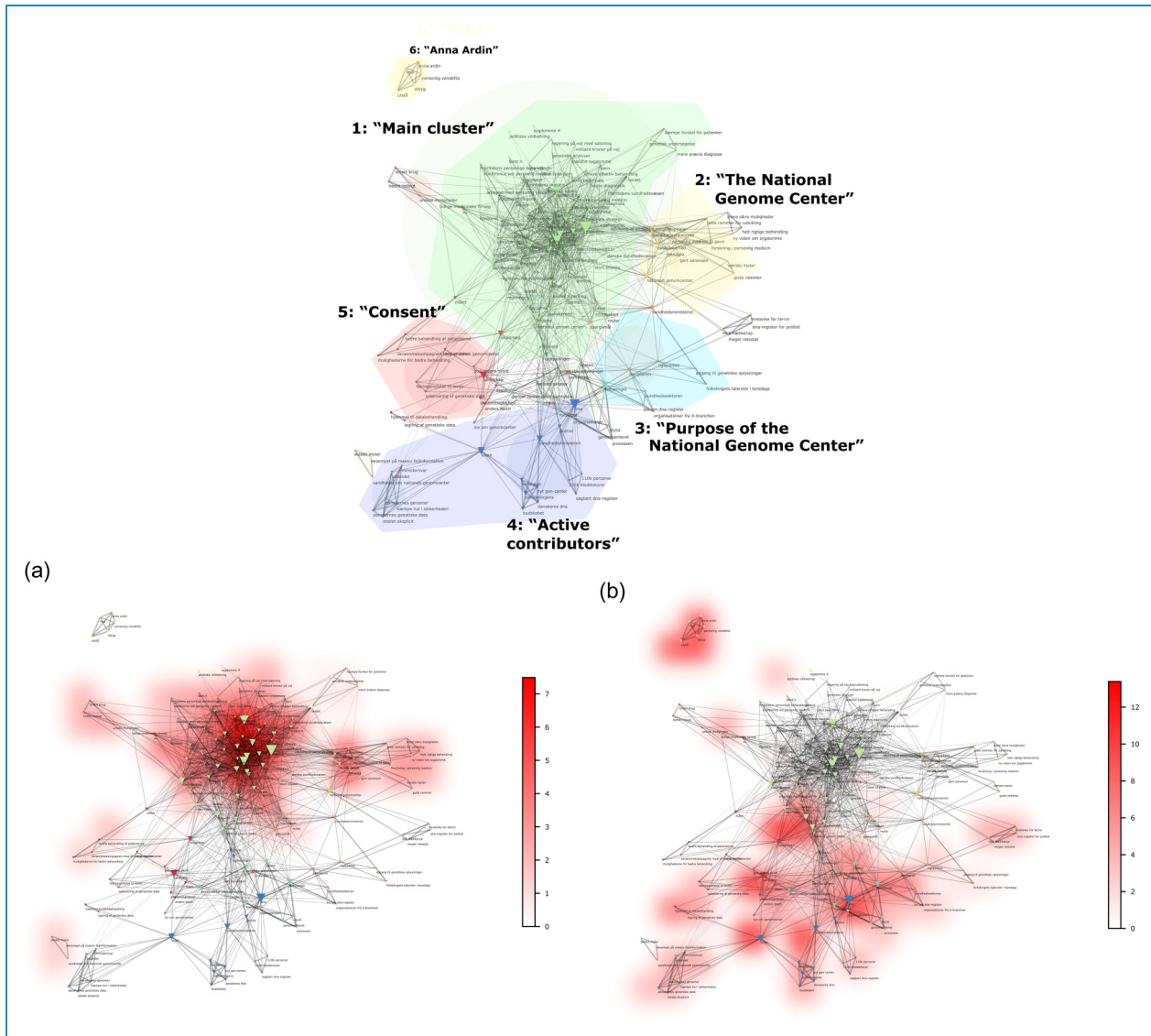
Cluster six we call ‘Ethics and moral’ and contains the hashtags #etik (ethics), #moral, #folkegavn (benefit for the people), #patientdatasikkerhed (patient information security), #privacy and #jura (law). The main part of this cluster comes from one user in the private category who shared the same tweet with many hashtags three times. In the tweet, the user accuses the National Genome Center of social blackmailing. The user is referring to discussions about whether patients have to consent to the storage of genetic data in the National Genome Center, according to the legislation about the centre passed in May 2018,<sup>44</sup> a point also discussed persistently in the debate in news media.<sup>23</sup> Like the left part of the main cluster related to concerns around the National Genome Center, the hashtags from the ‘Ethics and moral’-cluster are predominantly used by users in the private category (Figure 1(b)). Finally, there are some smaller clusters at the bottom of the map relating to specific events or topics: there are several about jobs within the field of personalised medicine, there is one related to COVID-19, one with tweets referring to a summer school at the University of Copenhagen

relating to personalised medicine and one with tweets about a conference on diabetes.

### Network graphs based on nouns

Figure 2 shows a network graph of co-occurrence of nouns. Figure 2(a) and (b) show heat maps on the base map for nouns. From the noun clusters, we learn about the striking divergence in how people from the two categories frame personalised medicine on Twitter. While people affiliated with personalised medicine tweet about benefits and highlight the positive potential, people we categorised as private users tweet about concerns and the conditions of implementing personalised medicine.

From Figure 2 we also identified six clusters. ‘The main cluster’ is again the one with the most activity. A big part of the words in this cluster frames personalised medicine positively using words like ‘fremtidens sundhedsvæsen’ (the healthcare system of the future), ‘færre bivirkninger’ (fewer side effects), ‘bedre behandling’ (better treatment), and ‘stort potentiale’ (big potential). In the middle of the cluster there are two relatively big nodes with the words ‘regionerne’ (The Regions) and ‘universiteterne’ (The Universities). These cover tweets describing the Regions and the Universities as institutions playing a part in the implementation of personalised medicine. The word ‘regionerne’ also covers tweets answering @regionerne, which is the Twitter page of the joint interest organisation of The Danish Regions. Several tweets also write about the publishing of the ‘nationale strategi’ (the national strategy) for personalised medicine. The second cluster we call ‘The National Genome Center’ and it contains tweets about the establishment of the infrastructure with the same name and the expected benefits, using words like ‘helt rigtige behandling’ (just the right treatment) and ‘gavn for patienterne’ (benefits for the patients). The top of the cluster accordingly frames the National Genome Center positively. As apparent from Figure 2(a), it is mainly users from the category of professionally interested in personalised medicine who use the nouns from the top part of the map, that is, the nouns which frame personalised medicine positively. ‘The National Genome Center’ cluster is connected to a sub-cluster towards the bottom of the map in which concerns about the potential of using the National Genome Center as ‘dna-register for politiet’ (dna-register for the police) in cases where there is ‘mistanke for terror’ (suspicion of terror) are discussed. The third cluster we call ‘Purpose of the National Genome Center’ and it covers tweets about how genetic information stored in the centre can be used in the future. This cluster is centred around the words ‘DNA’ and ‘Folketinget’ (the Parliament) and contains mainly critical tweets from people in the private category. Both the part of ‘The National Genome Center’-cluster where users express concerns and the ‘Purpose of the National Genome



**Figure 2.** Base map of co-occurrence of nouns in tweets concerning personalised medicine in Danish. (a) Noun base map (Figure 2) with heat map layer (coloured areas) indicating which noun-clusters are predominantly used by users in the category ‘Professional interest in personalised medicine’. (b) Noun base map (Figure 2) with heat map (coloured areas) indicating which noun-clusters are predominantly used by users in the category ‘Private’.

Center’-cluster accordingly discuss the scope and limitations of use of information from the National Genome Center. Nouns from these two clusters are mainly from tweets by users in the private category (Figure 2(b)).

The fourth cluster we call ‘Active contributors’ because it primarily contains tweets from two users, one very critical from the private category and one very positive, which we categorised as having a professional interest in personalised medicine. The active user from the private category is the doctor without description in their biography, mentioned in the methods section, who has been very active in the debate in news media. We call the fifth cluster ‘Consent’, because it contains tweets discussing consent to storage

of genetic information in the National Genome Center, using words like ‘god ordens skyld’ (to set the record straight), ‘samtykke’ (consent), ‘genom center uden samtykke’ (genome centre without consent). Nouns from this cluster are predominantly used by users in the private category (Figure 2(b)).

Summing up, there seems to have been most debate on Twitter around the time of the establishment of the National Genome Center. Besides this, Twitter facilitates different discussions about personalised medicine depending on whether users were categorised as working within the field or not. Users working within the field tweet about the topic in relation to existing infrastructures and



organisations. Additionally, they tweet about the promises of personalised medicine. On the other hand, users from the private category are discussing the concrete physical realisation of these ambitions in the form of a new infrastructure and express concerns about the actual conditions for the implementation of personalised medicine.

## Discussion

Social media are sometimes heralded as a new public sphere enabling democratic participation and the construction of counter-publics.<sup>30,31</sup> We find that the publics who assemble in reaction to personalised medicine mediated through Twitter include both users with a professional interest in personalised medicine and users we categorised as private. This is notable, considering that previous research argues that written Danish news media mainly facilitate participation of people working with personalised medicine.<sup>23</sup> Accordingly, our study indeed supports the claim that social media *do* facilitate participation for a broader public than news media. Twitter is easier to access than news media where editors decide what is relevant enough to go to print. This naturally facilitates a broader participation. News media are likely to demand other kinds of identification of contributors than Twitter does, where writing something in the biography is optional. Still, the proportion of people categorised as 'Private' suggests that personalised medicine is an issue which assembles more than only people working within the field, although it is still a very small proportion of the Danish population who become engaged through Twitter. The debates mediated through news media and Twitter are closely related however. Many tweets contain links to articles or opinion pieces from news media. Therefore, debate mediated through the two media should not be seen as separate.

Bresling and colleagues have argued that Twitter users in Denmark are an 'elite community of politicians, journalists, interest groups and citizens who are more interested in politics than the average person in Denmark'.<sup>33</sup> Other researchers have also pointed out that Twitter users across countries are not demographically representative and that Twitter users pay more attention to politics.<sup>61,62</sup> Although the users we have categorised as 'Private' might be people 'who are more interested in politics', Twitter does facilitate debate about personalised medicine among a people who are at least not easy to identify as either politicians, lobbyists or interest groups. Nevertheless, our study points to how Twitter is also a place for strategic communication from people already in powerful positions.

While personalised medicine did spark a somewhat large debate on Twitter in 2018, the issue today seems somewhat stabilised despite the uncertainties around the future funding. Most tweets from this dataset were posted in 2018 around the time the legislation enabling the establishment of the National Genome Center was passed.

Apparently, a potential for influencing the development of the future policies of personalised medicine does not in itself spark a debate online. Our data indicate that if personalised medicine is again being discussed, it is being discussed using different words than in 2018, by using other forums or offline where people not affiliated with personalised medicine are likely to have a poorer chance of expressing their concerns.

Personalised medicine is framed differently on Twitter depending on whether users work within the field or not. Firstly, users working with personalised medicine tweet considerably more about the topic than the private users. Users working within the field tweet about personalised medicine in relation to existing infrastructures, institutions and organisations. Additionally, many users working within the field use Twitter to promote personalised medicine, both by sharing news about developments and policies in the field and by highlighting the potential benefits for the patients and for the healthcare sector. It has previously been argued that personalised medicine is a buzzword containing promises of a better future, sometimes without substantial scientific evidence, and that the notion is used to mobilise investments and political good will towards specific developments in the healthcare sector.<sup>63,64</sup> It seems as if Twitter is a platform where such futures are promoted by people working in the field of personalised medicine. However, personalised medicine is not mainly discussed positively on Twitter, as Almomani et al. have argued is the case in the UK news media.<sup>22</sup> Concerns about the infrastructure set up to facilitate personalised medicine, potential use of genetic information for purposes other than research or treatment, concerns related to data security and the risk of privacy breaches, and not least tweets about design of a consent or opt-out, are also prevalent on Twitter. These are topics known to be discussed in relation to sharing of health data<sup>5,52,65</sup> and more specifically in relation to the implementation of personalised medicine.<sup>23</sup> Our study suggests that Twitter can be used as a platform for policymakers to gain insights into how these topics are perceived by people not directly affiliated with the implementation of personalised medicine.

The different ways the user groups tweet about personalised medicine are also made apparent by different use of our search terms: Users with a professional interest in personalised medicine use the hashtag #personalised medicine while private users use the hashtag #genome center. Our assumption was that the two words were describing the same issue. However, it seems the words are highlighting different issues: one where professionals are promoting personalised medicine as a promissory concept and as an ambition for the future healthcare sector, and one where concerns about the actual conditions for the implementation of personalised medicine, including the formation of a new infrastructure, are shared by private users. That the different user groups use different

words to tweet about the matter also introduces the risk of filter bubbles,<sup>66</sup> where users are not exposed to views opposing their own.

The results of this study are limited by the subjective choices made by the authors both in terms of how we define discussions related to personalised medicine<sup>26</sup> and how we categorise the Twitter publics. Our search terms delimit discussions about personalised medicine to users who tweet about it using these terms. Since we found that our search words were used (predominantly) by different user groups the inclusion of other search terms would likely have facilitated participation of other users. How people discuss issues also change over time. We cannot dismiss that the words used to discuss personalised medicine on Twitter have changed over the course time. The categorising of users into three categories is also a way of constructing specific publics and enable specific kinds of analysis. In the study, we focus on the aggregated frequency of specific hashtags and nouns. We encourage future studies to explore other aspects of discussions about personalised medicine on Twitter, for instance, the reachability of tweets or interactions with tweets in the form of likes.

## Concluding remarks

In this paper, we have used Twitter to explore discussions about personalised medicine in Denmark. The paper provides insights into what is discussed in relation to reuse of health data and into how Twitter can be used to study public opinion. We have found that different publics use the platform to tweet about different aspects of personalised medicine and that users working with personalised medicine are notably more active on this issue than private users. From our study we learn, that discussions about the physical realisation of ambitions for personalised medicine in the form of a new infrastructure and concerns related to privacy, data security and the scope of the use of health data are mainly led by private users. We also learn that Twitter is a platform used to advance certain political initiatives by people already in powerful positions, when users working in the field of personalised medicine use Twitter to promote it, and to highlight the potential future benefits. These findings support that Twitter can be used by policy-makers to get insights into discussions about reuse of health data from people unaffiliated with the field. Yet, it is important to note that the 961 users we categorised as ‘Private’ are likely to represent people with very specific characteristics,<sup>30</sup> and that they are by no means representative of the Danish population or the ‘general public’. Additionally, it seems that only certain topics are discussed on Twitter. The establishment of the National Genome Center and the following legal change in 2018 was one such thing. However, political uncertainty alone does not spark an issue into being. Today’s uncertainty about the future funding of personalised medicine, including

uncertainty about the role of private actors, does not seem to have resulted in a rise in tweets about the topic. If these discussions exist then they are outside of this platform. Therefore, our study serves to remind people interested in public opinion that Twitter is a platform used for multiple purposes by different actors and not simply a bottom-up democratic forum.

**Acknowledgements:** The authors would like to thank Klaus Hoeyer, Mette Nordahl Svendsen and the IPW-writing group at UVienna for valuable feedback on an earlier version of this article.

**Contributorship:** LS formulated the research interest, researched the literature and prepared the first draft of the manuscript. AG collected the data. Both authors analysed the data. Both authors reviewed and edited the manuscript and approved the final version.

**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.

**Ethical approval and consent to participate:** Not applicable. The article only reports aggregated data from publically available tweets and adhere to Twitter’s developer agreement and policy.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors’ research has received funding from the Carlsberg Foundation (Semper Ardens Grant CF17-0016).

**Guarantor:** LS.

**ORCID iD:** Lea Skovgaard  <https://orcid.org/0000-0001-5491-6928>

## References

1. European Commission. *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS: on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society*. 25 April 2018. Brussels.
2. Prainsack B. *Personalized medicine: empowered patients in the 21st century?* 1st ed. New York: New York University Press, 2017.
3. Parry B and Greenhough B. *Bioinformatics*. Resources series. Cambridge; Medford, MA: Polity, 2018.
4. Mittelstadt BD and Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics* 2016; 22: 303–341.
5. Piasecki J, Walkiewicz-Żarek E, Figas-Skrzypulec J, et al. Ethical issues in biomedical research using electronic health

- records: a systematic review. *Med Health Care Philos* 2021; 24: 633–658.
6. Lee SS-J. Obligations of the “gift”: reciprocity and responsibility in precision medicine. *Am J Bioeth* 2021; 21: 57–66.
  7. Petersen AR. *Digital health and technological promise: a sociological inquiry*. Milton Park, Abingdon, Oxon; New York, NY: Routledge, 2019.
  8. Tupasela A. *Populations as brands: marketing national resources for global data markets*. London: Palgrave Macmillan Cham, 2021.
  9. Langhoff TO, Amstrup MH, Mørck P, et al. Infrastructures for healthcare: from synergy to reverse synergy. *Health Informatics J* 2018; 24: 43–53.
  10. Jasanoff S. Technologies of humility: citizen participation in governing science. *Minerva* 2003; 41: 223–244.
  11. Budin-Ljøsne I and Harris JR. Ask not what personalized medicine can do for you – ask what you can do for personalized medicine. *Public Health Genom* 2015; 18: 131–138.
  12. Lehoux P. Beyond our mutual ignorance. Or, how would engaging the public benefit the personalized medicine community. *Curr Pharmacogenom Pers Med Moving* 2011; 9: 76–79.
  13. Irwin A. Constructing the scientific citizen: science and democracy in the biosciences. *Public Underst Sci* 2001; 10: 1–18.
  14. Marres N. The issues deserve more credit: pragmatist contributions to the study of public involvement in controversy. *Soc Stud Sci* 2007; 37: 759–780.
  15. Felt U and Fochler M. The bottom-up meanings of the concept of public participation in science and technology. *Sci Public Policy* 2008; 35: 489–499.
  16. Schicktanz S, Schweda M and Wynne B. The ethics of ‘public understanding of ethics’—why and how bioethics expertise should include public and patients’ voices. *Med Health Care Philos* 2012; 15: 129–139.
  17. Wynne B. Public participation in science and technology: performing and obscuring a political–conceptual category mistake. *East Asian Sci Technol Soc: Int J* 2007; 1: 99.
  18. Marres N. Issues spark a public into being: A key but often forgotten point of the Lippmann-Dewey debate. In: Latour B and Weibel P (eds) *Making things public: atmospheres of democracy*. Cambridge, Mass.: [Karlsruhe, Germany]: MIT Press; ZKM/Center for Art and Media in Karlsruhe, 2005: pp. 208–217
  19. Birkbak A, Petersen MK and Elgaard Jensen T. Critical proximity as a methodological move in techno-anthropology. *Techné: Res Philos Technol* 2015; 19: 266–290.
  20. Hjörleifsson S, Árnason V and Schei E. Decoding the genetics debate: hype and hope in Icelandic news media in 2000 and 2004. *New Genet Soc* 2008; 27: 377–394.
  21. Pálsson G and Rabinow P. The Icelandic genome debate. *Trends Biotechnol* 2001; 19: 166–171.
  22. Almomani B, Hawwa AF, Goodfellow NA, et al. Pharmacogenetics and the print media: what is the public told? *BMC Med Genet* 2015; 16: 32.
  23. Author deleted for anonymity.
  24. Marres N. *Digital sociology: the reinvention of social research*. Oxford, UK: Polity Press, 2017. <http://ebookcentral.proquest.com/lib/kbdk/detail.action?docID=4860618> (2017, accessed 22 March 2018).
  25. Rogers R. *Digital methods*. Cambridge, Massachusetts: The MIT Press, 2015.
  26. Venturini T and Munk AK. *Controversy mapping: a field guide*. Cambridge, UK; Medford, MA, USA: Polity, 2022.
  27. Barros JM, Duggan J and Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (infoveillance): systematic review. *J Med Internet Res* 2020; 22: e13680.
  28. Venturini T, Jacomy M, Meunier A, et al. An unexpected journey: a few lessons from sciences Po médialab’s experience. *Big Data Soc* 2017; 4: 2053951717720949.
  29. Venturini T and Latour B. The social fabric: digital traces and quali-quantitative methods. In: Arai K, Kapoor S and Bhatia R (eds) *Proceedings of the future En Seine*. Editions Future en Sein. Raleigh, NC, USA: Lulu.com 2009: pp.87–101.
  30. Bakardjieva M. Subactivism: lifeworld and politics in the age of the internet. *Inf Soc* 2009; 25: 91–104.
  31. Shirky C. The political power of social Media: technology, the public sphere, and political change. *Foreign Aff* 2011; 90: 28–41.
  32. Bruns A and Burgess J. The use of Twitter hashtags in the formation of ad hoc publics | QUT ePrints. In: Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference. Reykjavik: University of Iceland, <https://eprints.qut.edu.au/46515/> (2011, accessed 6 April 2018).
  33. Bresling SD, Enggard TR, Blok A, et al. How We Tweet About Coronavirus, and Why: A Computational Anthropological Mapping of Political Attention on Danish Twitter during the COVID-19 Pandemic. In: *Somatosphere*. <http://somatosphere.net/forumpost/covid19-danish-twitter-computational-map/> (2020, accessed 19 November 2021).
  34. Schweinberger M, Haugh M and Hames S. Analysing discourse around COVID-19 in the Australian twittersphere: a real-time corpus-based analysis. *Big Data Soc* 2021; 8: 20539517211021436.
  35. Modave F, Zhao Y, Krieger J, et al. Understanding perceptions and attitudes in breast cancer discussions on twitter. *Stud Health Technol Inform* 2019; 264: 1293–1297.
  36. Beguerisse-Díaz M, McLennan AK, Garduño-Hernández G, et al. The ‘who’ and ‘what’ of #diabetes on Twitter. *Digital Health* 2017; 3: 2055207616688841.
  37. Walter S, Lörcher I and Brüggemann M. Scientific networks on Twitter: analyzing scientists’ interactions in the climate change debate. *Public Underst Sci* 2019; 28: 696–712.
  38. Bogen KW, Bleiweiss KK and Leach NR. #Metoo: disclosure and response to sexual victimization on twitter. *J Interpers Violence* 2021; 36: 8257–8288.
  39. Hays R and Daker-White G. The care.data consensus? A qualitative analysis of opinions expressed on Twitter. *BMC Public Health* 2015; 15: 38.
  40. United Nations. *E-government survey 2020. Digital government in the decade of action for sustainable development*. New York: United Nations, 2020.
  41. Bauer S. From administrative infrastructure to biomedical resource: danish population registries, the “scandinavian laboratory,” and the “epidemiologist’s dream”. *Sci Context* 2014; 27: 187–213.
  42. Sundheds- og Ældreministeriet. National strategi for Personlig Medicin 2017-2020. 2016.

43. Sundheds- og Ældreministeriet. L 146 Lov om ændring af sundhedsloven (Organiseringen i Sundheds- og Ældreministeriet, oprettelse af Nationalt Genomcenter m.v.). 2018.
  44. Sundheds- og Ældreministeriet. L 146 Forslag til lov om ændring af sundhedsloven. 2017.
  45. Jensen LG and Svendsen MN. Personalised medicine in the Danish welfare state: political visions for the public good. *Crit Public Health* 2021; 1–12. DOI: 10.1080/09581596.2021.1937524.
  46. Terkildsen MD, Burau V, Væggemose U, et al. The welfare state driving “me” and “we” medicine – a critical discourse analysis. *New Genet Soc* 2020; 39: 404–423. 1–20.
  47. Nationalt Genom Center. 1 mia. kr. på vej til skræddersyet behandling til danske patienter, <https://ngc.dk/nyheder/2018/december/1-mia-kr-paa-vej-til-skraeddersyet-behandling-til-danske-patienter> (2018, accessed 5 April 2022).
  48. Sundheds- og Ældreministeriet. National Strategi for Personlig Medicin 2021-2022. Personlig medicin til gavn for patienterne: Klar diagnose, målrettet behandling, styrket forskning (2021, accessed 3 April 2022).
  49. Twitter. Using Twitter, <https://help.twitter.com/en/using-twitter> (2022, accessed 19 May 2022).
  50. Twitter. How to use hashtags, <https://help.twitter.com/en/using-twitter/how-to-use-hashtags> (2022, accessed 19 May 2022).
  51. Twitter. Twitter API for Academic Research | Products, <https://developer.twitter.com/en/products/twitter-api/academic-research> (2022, accessed 29 March 2022).
  52. Skovgaard LL and Hoeyer K. Data authority: public debate about personalized medicine in Denmark. *Public Underst Sci* 2022; 31: 590–607.
  53. Vandeweerd CJ. Using Twitter bios to map Americans’ identities over time. SODAS Data Discussion 11 Feb 2022, <https://sodas.ku.dk/events/sodas-data-discussion-11-february-2022/>. 2022.
  54. Honnibal M and Montani I. Spacy 2: natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To Appear* 2017; 7: 411–420.
  55. Rehurek R and Sojka P. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University Brno, Czech Republic* 2011; 3: 2.
  56. CorText Network Analysis & Layout. In: *Cortext Manager Documentation*, <https://docs.cortext.net/analysis-mapping-heterogeneous-networks/mapping/> (2022, accessed 19 May 2022).
  57. Alothali E, Zaki N, Mohamed EA, et al. Detecting social bots on twitter: a literature review. In: 2018 International Conference on Innovations in Information Technology (IIT), 2018, pp.175–180. DOI: 10.1109/INNOVATIONS.2018.8605995.
  58. Marres N and Moats D. Mapping controversies with social media: the case for symmetry. *Soc Media + Soc* 2015; 1: 205630511560417.
  59. Kington T. Julian Assange humiliated and abused me, claims Swedish accuser Anna Ardin in new book. *The Times*, 2 August, <https://www.thetimes.co.uk/article/julian-assange-humiliated-and-abused-me-claims-swedish-accuser-anna-ardin-in-new-book-6dw9qsl8z> (2021, accessed 7 June 2022).
  60. Folkemødet. Om Folkemødet | Folkemødet, <https://folkemoedet.dk/om-folkemodet/om-folkemodet/> (2022, accessed 10 May 2022).
  61. Fazekas Z, Popa SA, Schmitt H, et al. Elite-public interaction on twitter: eU issue expansion in the campaign. *Eur J Polit Res* 2021; 60: 376–396.
  62. Mellon J and Prosser C. Twitter and Facebook are not representative of the general population: political attitudes and demographics of British social media users. *Res Politics* 2017; 4: 2053168017720008.
  63. Abettan C. Between hype and hope: what is really at stake with personalized medicine? *Med Health Care Philos* 2016; 19: 423–430.
  64. Hoeyer K. Data as promise: reconfiguring Danish public health through personalized medicine. *Soc Stud Sci* 2019; 49: 531–555.
  65. Stockdale J, Cassell J and Ford E. Giving something back’: a systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. *Wellcome Open Res* 2019; 3: 6.
  66. Bruns A. Filter bubble. *Internet Policy Rev* 2019; 8: 1–14. <https://policyreview.info/concepts/filter-bubble> (2019, accessed 7 September 2022).
-