

# Health research 2.0

*The use in research of personal fitness or health data shared on social network raises both scientific and ethical concerns*

Katrin Weigmann

Over the past decade, the Internet has allowed people to share and exchange increasing amounts of personal data. Much of it might seem trivial or of little consequence, but some of it, including information about health, is potentially both highly sensitive and useful to researchers. In support networks or communities of like-minded people, individuals voluntarily reveal and discuss health information they would usually have shared only with their physician or close family members. “In former times, the internet was used more passively,” said Barbara Prainsack, Associate Professor at the Department of Social Science, Health & Medicine at King’s College London (UK). “People would enter a search term like migraine and retrieve expert recommendations. This has now changed.”

*“The companies behind these networks also encourage participants to donate and share personal data in order to build large databases from self-reported data...”*

“By sharing our experiences, we can all contribute new data that can accelerate research and help create better treatments. Our experiences can actually change medicine... for good,” states the Web site of PatientsLikeMe ([www.patientslikeme.com](http://www.patientslikeme.com)), one of the largest online health networks, founded in 2004 by Benjamin and James Heywood after they learned that their brother had

amyotrophic lateral sclerosis (ALS). The DNA sequencing company 23andMe ([www.23andme.com](http://www.23andme.com)) is similarly interested in using personal health data for their “Research Revolution,” promising to “empower more people to jumpstart genetic research into the diseases that affect them and the people they love” (<http://blog.23andme.com/23andme-and-you/introducing-a-do-it-yourself-revolution-in-disease-research/>). The members of social health networks such as PatientsLikeMe and 23andMe engage in activities that traditionally belong in the hands of medical doctors or scientists. Members diagnose themselves, analyze their data, vote on research topics or even initiate research projects themselves. The companies behind these networks also encourage participants to donate and share personal data in order to build large databases from self-reported data for research purposes. This raises both ethical and regulatory concerns, as well as questions about the value of the results. Importantly, for most of their members, participating in research is not their main intention, though. “The incentive to participate can vary strongly from network to network,” said Prainsack. Some people simply like tracking their health-related data, while others join the network to buy genome-testing kits or look for medical information. “Often enough, you only get access to information if you reveal something about yourself. People participate in research without even being aware of it,” Prainsack explained.

PatientsLikeMe is based in Cambridge (MA, USA) and provides tools for its more than 220,000 members to monitor their condition, upload health data

and share their experiences. Membership is free because the company’s business model is based on selling anonymized user data to pharmaceutical companies and research institutions. PatientsLikeMe also conducts its own research. 23andMe, based in Mountain View (CA, USA), is the other big player of online patient networks. It was founded in 2006 by Linda Avey and Anne Wojcicki and is partly funded by Google, whose co-founder Sergey Brin is Wojcicki’s estranged husband. 23andMe, which has about 500,000 genotyped customers, is a combined direct-to-consumer (DTC) genetic testing service, research enterprise and social health network. In addition to buying genomic tests, customers can upload phenotypic and lifestyle information and share their data with others, thereby contributing to the company’s research database.

*“... it is not clear yet, how much crowdsourcing from social health networks will contribute to better health care”*

However, 23andMe has also attracted criticism from regulators, mainly in regard to its DTC genetic tests. In 2008, the State of California ordered 23andMe to stop offering tests to the public because State law requires that a licensed physician is involved. In 2010, the US Food and Drug Administration (FDA) announced plans to regulate commercial genetics companies, given that their products are diagnostics. In November 2013, the FDA sent 23andMe a warning letter because they were selling their kit without



marketing clearance or approval and told them to immediately stop marketing their genetic test kit. For the time being, 23andMe is selling its kit only for analyzing ancestral origins. But the company still provides the raw health-related data leaving customers to work out how to analyze their genomic data using free online software.

One of the first patient-organized studies based on a social health network was initiated by ALS patients through PatientsLikeMe and published in 2011 in *Nature Biotechnology* [1]. A group of ALS patients were self-experimenting with lithium carbonate in the hopes that this could slow the progression of their disease, as suggested by the results of a small trial from an Italian research group mainly from the University of Pisa [2]. The Research and Development team of PatientsLikeMe ran a more rigorous observational study that eventually refuted the Italian study. To date, the company has published several peer-reviewed papers based on patient-reported outcomes in major scientific journals.

23andMe has also published numerous research papers in high-impact journals [3]. One study, for example, replicated more than 180 previously reported genetic associations for type 2 diabetes, prostate cancer, cholesterol levels, multiple sclerosis and others. Another study, published in 2011, identified two novel loci linked to Parkinson's disease [4]—one of them was confirmed in collaboration with the International Parkinson's Disease Genomics Consortium [5]. Meanwhile, 23andMe has built a large Parkinson research community by enrolling more than 10,000 patients.

**H**ealth care will, in the view of some scholars, change dramatically in the near future. They foresee a more predictive, personalized medicine—a shift from a “one size fits all” strategy to a medicine that is tailored to each individual and focuses on disease prevention. As Leroy Hood, the inventor of automated DNA sequencing, and Stephen H. Friend,

co-founder and president of Sage Bionetworks, wrote: “Our vision is that, in the not too distant future, each patient will be surrounded by a ‘virtual cloud’ of billions of data points that will uniquely define their past medical history and current health status.” [6]. This vision will require a systems analysis approach and genotypic and phenotypic data from millions of individuals to develop models that can predict disease risk and suggest appropriate action.

Indeed, crowdsourcing is an efficient method for enrolling participants at very low cost. And the improvement in sequencing technologies and self-monitoring gadgets and smartphone applications will add even more health-related data to the mix. Yet, Cecile Janssens, Professor of Epidemiology at Emory University in Atlanta (GA, USA), remains skeptical that this will herald a new era for medicine. “We will get more personal applications here and there in future, but not to the extent that some people promise,” she said. In addition, it is not clear yet, how

much crowdsourcing from social health networks will contribute to better health care.

A number of systematic biases could hamper the interpretation of data from lay reports in social health networks [5]. The first caveat is “selection bias,” as patients in social health networks may not be representative for the disorder. “What you have is a selection of personality types that are more eager to share everything about themselves and their experiences,” Janssens explained. This could affect investigating psychiatric disorders in particular. The second problem, “information bias,” arises from self-reported data. For example, if participants learn about other people’s symptoms that come with a certain disease, they may be more inclined to experience these symptoms themselves. Networks such as PatientsLikeMe, which rely strongly on participants sharing information, may be particularly prone to this bias. “Participants that have been with PatientsLikeMe for a longer time tend to report more symptoms and diagnoses than people that just entered. There are people who self-report more than 30 diagnoses,” Janssens commented. Information bias, just like selection bias, presumably affects research into psychiatric diseases more than research focusing on biological questions. Most patients will have a formal clinical diagnosis of something such as Parkinson’s disease, whereas they are more likely to report anxiety disorders or depression without diagnosis. A third bias may come from confounding factors. For example, if coffee drinkers are more likely to also be smokers, a study that investigated the relationship between coffee drinking and the risk of lung cancer might not consider the confounding factor of smoking. It would thus overestimate the relationship between coffee and lung cancer. “When data are not collected for specific research questions, more often relevant factors are missing,” Janssens warned.

Simply increasing study size, according to Janssens, will not help. “The results are as valid as the data are. If you have flawed data, you will not improve them by looking at a lot of them,” she argued. The more data that are analyzed, the more correlations will be found that become statistically significant, which makes it increasingly difficult to filter out those that are meaningful. “We are entering an era where we find a lot of results that are statistically significant and in the

end, we do not know what is real and what is not,” Janssens said.

This does not mean that studies based on social health networks are meaningless. With their advantages in fast and easy recruitment and their easy access to data, such studies can raise important questions or make valuable contributions. For example, 23andMe only took a week to announce that they were not able to replicate results from a longevity study published in *Science*. The authors of the *Science* paper later retracted their study [5]. But data from social health networks need to be analyzed and interpreted with sufficient prudence, and results communicated carefully.

23andMe and PatientsLikeMe do part of that. “They do mention shortcomings of study design in their scientific papers. But when you look at their press releases, these references are gone,” Janssens commented.

Any research involving human health information raises ethical issues, such as informed consent and privacy protection, that are not unique to social health networks. Many existing governmental biobanks and data collections use a “broad consent” model whereby participants consent to future research of any kind; however, there is some debate over whether consent can be “informed” if the details of secondary research schemes are not yet known [7]. Data protection is also becoming more challenging as technology advances: It is possible to uncover the names and other private data of people whose genome profile was stored in a genomic database using publicly available information [8].

.....  
 “... the definition of “human subjects research” is narrow and vague enough to not include research based on social health networks...”  
 .....

But although consent and privacy protection are not unique issues for collecting data via the Internet, there is a difference between a publicly funded biobank and Web-based social health networks. “In governmental biobank projects, every care will be taken to inform the participants about the use of their data. There are ethics committees and

governance boards that monitor the procedure. For many social health networks, this is not the case,” commented Prainsack. In the USA, for instance, federally funded research projects involving human subjects require approval from research ethics committees, or institutional review boards (IRBs). However, projects that are not financed by federal money do not necessarily need IRB oversight. Moreover, the definition of “human subjects research” is narrow and vague enough to not include research based on social health networks, as researchers neither interact with participants nor obtain identifiable data. When 23andMe published their first paper in *PLoS Genetics*, they retrospectively obtained exemption from IRB on the grounds that it was not human subjects research. Although they accepted the paper, the editors of *PLoS Genetics* accompanied the publication with an editorial calling for an effort to standardize processes and procedures of consent and review of human genomic research [9].

“The fact that IRB is not required for certain types of research under US federal law does not mean that it does not take place,” said Pascal Borry, Assistant Professor of Bioethics at the Centre for Biomedical Ethics and Law at the University of Leuven in Belgium. Subsequent to their *PLoS Genetics* publication, 23andMe obtained IRB approval for their study protocol and revised consent procedure. But not every paper using anonymized data mentions IRB. In addition, discussions embark on when and how ethical oversight should be obtained. For example, uBiome (ubiome.com), a crowd-funded citizen science startup aiming to sequence the microbiome of its donors, raised controversy about ethical oversight earlier this year. uBiome obtained IRB approval only after securing the project’s funding and were heavily criticized for not undergoing IRB review from the beginning (<http://blogs.scientificamerican.com/molecules-to-medicine/2013/07/25/ubiome-ethical-lapse-or-not/>).

Informed consent is a central doctrine in clinical research. Subjects must be provided with all relevant information before any treatment starts or samples and data are collected, so that they are able to make an informed decision and participate voluntarily. “Making sure that people are really willing to participate in order to advance science



is, in general, not an easy task. It involves thorough communication and the development of adequate tools. Certainly, doing it via the Internet doesn't make it easier," Borry said. Obtaining consent from participants online is often reduced to clicking a checkbox. "This solves the legal aspect of consent, but not the ethical and social one," Prainsack said. It is doubtful whether people read or understand what they consent to.

.....  
**“Obtaining consent from participants online is often reduced to clicking a checkbox”**  
 .....

This became apparent when 23andMe obtained patents: the first in 2012 and then again in September 2013. "It became clear from customers' comments on the 23andMe website that they were not aware of consenting to their data being used for patents and they were not amused that this was happening," Borry said. Although 23andMe has the right to file patents, Borry explained that, "Patenting stands in opposition to what the company says about the democratization of research, open communication and open access." Prainsack takes a similar view: "What has been criticized by many people is that the company's rhetoric of common benefit contrasts with the attitude that of course, as a traditional for-profit company, they protect their intellectual property." Indeed, people may be less inclined to participate when they are aware of the company's intent to file patents [10]. As Borry explained: "People want scientific progress to happen, but they are not necessarily interested in advancing the economic development of one company." Not only does the company gain financial profit from a patent that is based on data they obtained for free, but it may thereby also impede research progress by charging licensing fees. "People are less inclined to participate in a research project if it is clearly designed for commercial purposes."

**S**ocial health networks that use participants' data for research are an unusual hybrid. On the one hand, they build on people's desire to have a voice in medical research and help advance science for the greater good. On the other hand, they collect data to generate income [11]. "It is important not to automatically equate

'commercial' with 'unethical' and 'non-commercial' with 'ethical,'" Prainsack said. She pointed out that participants should at least be well informed that they are part of a business: "We know from empirical studies that many people do not read the small print. Of course, companies might argue that this is their [the users] problem. But I think that Web-based platforms should adapt to reality. Information about the business model should be presented in a user-friendly way." A possible model, according to Prainsack, would be for a company to publish a visible "how do we make money?" section on their Web sites—a feature that is available on PatientsLikeMe, but is missing from most other Web sites.

.....  
**“If participants donate their personal data for research purposes, they should be treated with the same respect as participants in publicly funded research projects”**  
 .....

Generally, the way in which consumer-focused genomics companies and other social health networks have drawn their participants and customers into research activities has received little attention, but their practices should draw criticism. If participants donate their personal data for research purposes, they should be treated with the same respect as participants in publicly funded research projects [12]. Participants should be well informed so that they can proper consent whether or not to participate, and ethical oversight should be provided.

It is not yet clear how much of a contribution crowd-sourced research from social health networks will make to science. These networks might speed up the process of data gathering and improve cost efficiency, and they could shed some light on topics that are otherwise neglected. However, systematic biases present in the sampling might reduce the value of their findings, and the current lack of ethical oversight or even the appreciation of ethical issues needs to be overcome. As Janssens pointed out: "The problem with these studies is that people are disclosing a lot of information about themselves to advance science. And if you cannot really

promise to advance science considerably, how much can you ask from them?"

### Conflict of interest

The author declares that she has no conflict of interest.

### References

1. Wicks P, Vaughan TE, Massagli MP, Heywood J (2011) Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol* 29: 411–414
2. Fornai F, Longone P, Cafaro L, Kastsichukha O, Ferrucci M, Manca ML, Lazzeri G, Spalloni A, Bellio N, Lenzi P *et al* (2008) Lithium delays progression of amyotrophic lateral sclerosis. *PNAS* 105: 2052–2057
3. Swan M (2012) Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res* 14: e46
4. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM, Tanner CM, Langston JW *et al* (2011) Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinsons disease. *PLoS Genet* 7: e1002141
5. Janssens ACJW, Kraft P (2012) Research conducted using data obtained through online communities: ethical implications of methodological limitations. *PLoS Med* 9: e1001328
6. Hood L, Friend SH (2011) Predictive, personalized, preventive, participatory (P4) cancer medicine. *In Nat Rev Clin Oncol* 8: 184–187
7. Hayden EC (2012) Informed consent: a broken contract. *Nature* 486: 312–314
8. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2013) Identifying personal genomes by surname inference. *Science* 339: 321–324
9. Gibson G, Copenhaver GP (2010) Consent and internet-enabled human genomics. *PLoS Genet* 6: e1000965
10. Sterckx S, Cockbain J, Howard H, Huys I, Borry P (2012) Trust is not something you can reclaim easily: patenting in the field of direct-to-consumer genetic testing. *Genet Med* 15: 382–387
11. Prainsack B (2011) Voting with their Mice: Personal Genome Testing and the "Participatory Turn" in Disease Research. *Accountability in Res* 18: 132–147
12. Howard H, Knoppers BM, Borry P (2010) Blurring lines. *EMBO Rep* 11: 579–582