

## Big Data for Personalized and Persuasive Coaching via Self-monitoring Technology

Lisette van Gemert-Pijnen, Floor Sieverink, Liseth Siemons, Annemarie Braakman-Jansen  
 j.vangemert-pijnen@utwente.nl, f.sieverink@utwente.nl, l.siemons@utwente.nl, l.m.a.braakman-jansen@utwente.nl  
 Center for eHealth and Wellbeing Research: dept., Psychology, Health and Technology  
 University of Twente  
 Enschede, The Netherlands

**Abstract**—Big Data are the key towards personalized healthcare. However, this confronts us with new challenges. What are the hopes, challenges and dangers for using Big Data to develop personalized and persuasive coaching systems? The aim of this paper is to prompt a discussion on Big Data to personalize healthcare, using views and ideas from experts, and to develop an agenda for future research.

**Keywords**-big data; self-monitoring; personalized healthcare; coaching.

### I. INTRODUCTION

Many different definitions are available for “Big Data”. Where Mayer-Schönberger and Cukier [1] focus on the new insights and economic value that can be obtained from Big Data in contrast to traditional smaller settings, Wang & Krishnan [2] refer to Big Data as complex and large data sets that can no longer be processed using the traditional processing tools and methods. Yet another definition comes from Laney [3], who defines Big Data according to 3 assets (often referred to as the 3V-model) that require new, cost-effective forms of information processing to promote insight and decision making, including: 1) high-volume (i.e., the quantity of data), 2) high-velocity (i.e., the speed of data generation and processing), and 3) high-variety (i.e., the amount of different data types).

Marr [4] expanded this 3V model to the 5V model by adding 2 additional Vs: veracity (i.e., the accuracy or trustworthiness of the data) and maybe the most important asset: value (i.e., the ability to turn the data into value).

Though this is just a grasp out of all the definitions available, there is one thing they have in common: The use of Big Data for analysis and decision making requires a change of thought from knowing “why” to knowing “what”. We focused on small, exact datasets and causal connections (i.e., knowing “why”) in the past; now we focus on gathering or linking a large amount of (noisy) data with which we can demonstrate the presence of (unexpected) correlational connections (i.e., knowing “what”) [1]. As a result, we will obtain (and apply) new insights that we did not have before. Insights that can not only be lifesaving, but that can also open the door towards more personalized medicine [5-7] to tailor medical decisions, medications and/or products to the individual’s personal profile instead of to what is best for a group of patients. For example, the use of genetic biomarkers in pharmacogenetics can be used to determine the best medical treatment for a patient [5] or the analysis of data from thousands of patients that have been treated in the past

can be used to determine what treatment best fits the individual patient that is under treatment now, e.g., in terms of expected treatment effects and the risk for severe side-effects given the patient’s personal characteristics like age, gender, genetic features, etc. This shift towards more personalized healthcare is reflected in the change of focus from a disease-centered approach towards a patient-centered approach, empowering patients to take an active role in the decisions about their own health [7]. As a result, an increasing number of technologies (e.g., Personal Health Records) are being launched by (insurance) companies to support chronically ill people in the development of self-management skills. Furthermore, the past decades have shown a rapid growth in the amount of (personal) data that is digitally collected by individuals via wearable technologies for self-monitoring and that may or may not be stored on online platforms for remote control [1],[5-8], and shared via other online sources like social media. Social media have become socially accepted and used by a growing group of people [9]. They use it, for example, to share data collected by activity, mood, nutrition and sleep trackers on a variety of online platforms (such as Facebook, Twitter, blogs or forums). This data provides new opportunities for personalizing and improving healthcare [10-12]. Also, the data and messages shared via these tools provide insight in vast amounts of valuable information for scientific purposes. Nagar and colleagues [12], for example, used the data from Twitter to predict flu trends and De Choudhury, Counts, and Horvitz [13] used social media as a measurement tool for the identification of depression. The information gleaned from social media has the potential to complement traditional survey techniques in its ability to provide more fine-grained measurement over time while radically expanding population sample sizes [13]. The challenge is how to predict mobility (behaviors; infections) in a trustworthy and reliable way?

Furthermore, the combination of clinical data with personal data on, for instance, physical activities, eating and sleeping patterns, might be used to tailor the treatment and coaching purposes to the needs of patients even better and are therefore seen as the key towards a future with optimal medical help [5]. However, it also confronts us with new technological and societal challenges that require more sophisticated data collection and management tools and data-analysis techniques. Concerns have been raised about security, purpose limitation, liability, safety, profiling, and data ownership, to mention just a few [1], [5], [6] but perhaps the most well-known concern bears upon our privacy [5], [6]. For a great deal, these privacy concerns are

associated with potential misuse of data by, for instance, insurance companies [5], [8]. If these privacy concerns are not dealt with appropriately, the public's trust in technological applications might diminish severely [8].

In sum, we face new technological and societal challenges using Big Data:

- Amount of data is growing explosively. We have to create DATA wisdom, to enhance the value of data for personalized healthcare at individual, community and society level. This requires a new model for enhancing the value of data (in particular the social value of data),
- Transparency in decision making using large amounts of data from various sources and different qualities. How to understand automated decision rules (algorithms) to support people in making adequate choices?
- Trust, will become a key issue in data driven healthcare. The ultimate question is not privacy and security per se but how to create faith and trust in data management and data maintenance? Privacy issues become particularly relevant when the linkage of anonymous datasets leads to re-identification. Encryption of the data might prevent identification of individuals, but transparency is not always possible (e.g., when analyzing query logs with search terms). In the end it is all about creating trust to overcome uncertainty or anxiety for a digital world.
- Mobility is growing explosively, and health related issues threaten mobility (global health; infections/resistance; food, water, climate change etc. How to ensure a healthy life and to promote wellbeing?
- Our behavior models might not be up to the task to provide just-in-time, interactive and persuasive feedback, and to develop intuitive and adaptive technologies to improve self-care and to transform healthcare.
- Our business models should enable production of open source solutions in healthcare and technology

To better understand how Big Data impacts our life and healthcare this vision paper aims to generate new ideas and thoughts, to set a research agenda to address and solve the aforementioned problems. There are three sections. Section I describes the background of Big Data in literature. Section II describes the procedure of the meetings with experts (Focus group; individual meetings). Section III presents topics about Big Data that are based on the meetings with experts and literature, and Section IV describes the implications for research using Big Data in healthcare.

## II. METHODS: FOCUS GROUP MEETING

To gain a more in depth picture about the pros and cons of using big data in personalized healthcare, a focus group was organized with a multidisciplinary panel consisting of 6 experts in Big Data research and quantified self-monitoring from different scientific disciplines: psychology, philosophy,

computer science, business administration, law, and data science. Individual face to face meetings were conducted to validate the focus group results. Participants from the University of Twente were selected based on their societal impact, expertise and experiences with conducting Big Data research.

Many potential issues regarding the use of Big Data have already been mentioned in the literature, newspapers, social media, or debates and panel discussion websites. However, many of these media sources do not specifically address the healthcare setting and only focus on a limited set of issues at a time (e.g., the privacy and security issues). With the focus group meeting, we therefore aimed to gain more insight into the scientific and societal issues that play a role in using and managing Big Data to support the growing needs for personalized (and cost-effective) healthcare.

We used literature and information from (social)media to prepare the focus group discussions. During the discussion experts were asked to write down many issues as possible that might become relevant using Big Data for healthcare. Flap overs were used to express the issues and experts had to categorize these issues into concepts that cover the issues. These concepts are presented in this vision paper. The focus group was audio taped and transcripts were made by authors of this paper, and discussed with experts individually.

The intent of the focus group was to understand and go in-depth into issues that are related to the use of Big Data in healthcare and to provide insights into how these issues are approached by different scientific disciplines.

## III. TOPICS BIG DATA: RESULTS FOCUS GROUP

### A. Empowerment

What does it mean when you monitor your activities, food intake, or stress 24 hours a day using technologies, like smart wearables? What drives us to use these 24 hour monitoring devices and what do we need to understand the data generated by these systems? Do we understand the algorithms that are used to capture our behaviors and moods in pictures and graphs? Who owns the data and how to control the maintenance of that data? How to avoid a filtered scope on our lives ignoring others that are out of our affinity groups? The concept of empowerment captures topics as autonomy, freedom, having control.

Big data evokes a discussion about freedom and autonomy. Autonomy concerns our critical view on how to use technology, while freedom is more about our way of living and thinking. It might, therefore, be more important to focus on freedom instead of autonomy: understanding how you are being influenced and taking a stance against that instead of trying to keep everything away. The focus group made a distinction between positive freedom and negative freedom; two common concepts within the field of philosophy. Positive freedom is the freedom to do something yourself (e.g., to decide for yourself that you want to share your data), whilst negative freedom is the freedom to keep things away, protecting yourself (e.g., when you do not give permission to companies to link your data with other

sources). Not losing control, being able to use, share and understand your data will be one of the topics to discuss freedom, self-efficacy using self-monitoring technologies.

Empowerment forces us to think about having control, who has the power through the use of Big Data? There might be just a small elite that understands the algorithms and with the increasing complexity, this elite will become even smaller in the future. This can create a division between people who can access and understand the algorithms and people who do not.

Empowering by personalization is one of the aims of the participatory society. Big data can be a leverage to realize this by creating a personal profile, providing the right information, at right moments to enable just in time coaching. Though it can be useful to put people in a profile, the danger of profiling is that you can never leave the assigned group again; once assigned to a group means always assigned to that group. Profiling might be suffocating to people because it creates uncertainty about what people know about you, what data are being collected, and for what purposes. Also, it is often unclear how to determine the norm to which people are compared when assigning them to a group (i.e., standardization, losing freedom). Furthermore, being assigned to a profile might lead to discrimination and certain prejudices/biases. Questions that arise are: How can profiling be used in a sensible/sound way? And who is responsible when mistakes are being made based on a certain profile?

### B. Trust

Trust will become a key concept in a data driven society. This concept captures more than privacy and security issues. Trust refers to topics as how to create faith in data management and data maintenance, and how to make sense of these data for humans.

Privacy issues become particularly relevant when the linkage of anonymous datasets leads to re-identification. Encryption of the data might prevent identification of individuals, but transparency is not always possible (e.g., when analyzing query logs with search terms). In the end it is all about creating trust to overcome uncertainty or anxiety for a digital world.

People often give consent to institutions to use their data for certain purposes in return for the (free) use of the product or service. However, data can be (re)used for other purposes as well or can be sold to other interested parties, even though that is not always allowed. This leads to great concerns: e.g., healthcare insurance companies who use treatment data for other purposes on a more personal level (for instance, for determining a personalized health insurance premium based on your personal data about your health and lifestyle). It is not that people do not want to share data, they already do this using Facebook or Google services, but they want to understand what happens with the data, in particularly when it concerns the health domain.

Self-monitoring technologies, with no doctors or nurses involved in the caring process, are provided more and more by institutions. Smart algorithms can be applied to personalize data in such a way so you can manage your

health and wellbeing yourself. However, these algorithms decide what information you get to see, based on information about you as a user (e.g., search history, Facebook friends, location). This will influence trust in the healthcare system, using data from your device compared to personal advices given by your doctor or nurse.

### C. Data Wisdom

There is a rapid growth of self-monitoring technologies, but little is known about the reliability and validity of these systems. The lack of evidence for causality can lead to unreliability as well. Furthermore, how can you tell what you are actually measuring? How can we validate the correlations that we find? Does it really say what we think it says or are it just assumptions? How to match qualitative or experience data with quantitative data generated by your device?

Data wisdom is the concept that captures scientific and societal topics.

Scientific refers to how to create data wisdom, in several ways. Those who generate data are not the ones that have the knowledge to analyze, those who analyze lack domain insight (technologies, behaviors). Different kinds of expertise will be needed in the future to deal with Big Data. For instance, expertise to analyze Big Data, expertise to develop and understand the working of algorithms, or expertise in data interpretation and visualizations. The use of data to personalize healthcare demands for new knowledge to support critical and creative thinking to understand data driven decisions and to watch the impact on science, health and society. We all know the disaster with google flu trends, but we have to learn from these failures to set the agenda for future research in using several sources of data (geospatial data, medical data, technology device data) to develop predictive models about health and wellbeing. We have to search for new models, methods to deal with huge datasets, search for patterns rather than testing hypotheses based on small data. Results are not causal-driven but correlational-driven. This requires a change in thought. The golden rule for Randomized Clinical Trials will no longer be the ultimate format for health sciences. New methods are needed to get a grip on "big", how many data (critical mass) is needed and how rich and mature should data be to make meaningful decisions? How to add qualitative experiences and expertise to Big data? Numbers do not tell the whole story, and a clinical eye is important to interpret data in the context of individual health and wellbeing.

Societal refers to the implications for healthcare, addressing topics as ethics, values for a meaningful life. How to avoid a division between people who can access and understand the data and analytics that rule the decisions about treatments and lifestyle advices? Knowledge and skills are needed to empower people, and people should participate in debates about the values of data for self-regulations on the level of individuals, communities and society. Transparency and Trust are the key-topics in that debate. Digging into Data starts with a scientific and societal debate on the values of data for a smart and healthy society.

#### IV. DISCUSSION AND CONCLUSION

##### A. Discussion: Food for Thoughts

What wisdom is needed to keep control, to enable self-organization in a Data Driven Society?

We want to share some food for thoughts, based on the focus group results and current debates on digging into data. We think radically new concepts, models (ICT and behavior based) and products are needed to promote self-organization, and persuasive coaching to enable intuitive and empathic designs to support decisions without nudging people into predefined formats for improving health and wellbeing.

Redefining environment in research. Radically new concepts on technology support: community driven, non-obtrusive, intelligent and intuitive ICT environments.

Redefining evidence in research, introducing value based health concepts and ecological measurements. Redefining the collection and analysis of Big Data combining social sciences with geo-health informatics and engineering to enable accessing, integrating and analysing spatiotemporal data, patient data and behavioural data.

##### B. Conclusion: Future work

This vision paper is aimed to discuss Big Data topics for personalized healthcare. The topics will be further investigated in research to 1) develop new methods, models to better measure, aggregate, and make sense of previously hard-to-obtain or non-existent behavioral, psychosocial, and biometric data, taken into account the topics from this introduction paper, 2) to develop an agenda for Big Data research to transform and improve healthcare. For example topics below will be explored in current and future research projects:

- Health analytics; advanced methods (machine learning) and models to analyze Big Data.
- Predictive modelling; to set up smart models to predict behaviors, to prevent diseases and to personalize healthcare.
- Visualization of data; how to present data meaningful to support decision making?
- Integration of (mobile) tech with data-platforms to enable automated services, to tailor feedback.

- Disruptive models (new actors, role-players in data driven systems).

#### ACKNOWLEDGMENT

There are no parties or institutes involved in this publication that require acknowledgements.

#### REFERENCES

- [1] V. Mayer-Schonberger and K. Cukier: *Big Data: A Revolution That Will Transform How We Live, Work and Think*. New York, NY: Houghton Mifflin Harcourt; 2013.
- [2] W.Wang and E. Krishnan: *Big Data and Clinicians: A Review on the State of the Science*. *JMIR Med Inform.* 2014;2(1):e1. doi:[10.2196/medinform.2913](https://doi.org/10.2196/medinform.2913)
- [3] D. Laney: *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Stamford, CT: META Group Inc.; 2001.
- [4] B. Marr: *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. John Wiley & Sons; 2015.
- [5] S. Klous and N. Wielaard: *We are big data. The future of the information society [Wij zijn big data. De toekomst van de informatiesamenleving]*. Amsterdam: Business Contact; 2014.
- [6] T.B. Murdoch and A.S Detsky: *The inevitable application of big data to health care*. *JAMA*. 2013;309(13):1351-2.
- [7] N.V. Chawla and D.A. Davis: *Bringing big data to personalized healthcare: a patient-centered framework*. *J Gen Intern Med*. 2013;28(3):S660-5.
- [8] P. Kamakshi: *Survey on big data and related privacy issues*. *International Journal of Research in Engineering and Technology*. 2014;3(12):68-70.
- [9] A.M. Kaplan and M. Haenlein: *Users of the world, unite! The challenges and opportunities of Social Media*. *Business Horizons*. 2010;53(1):59-68.
- [10] R. Batool, W.A. Khan, M. Hussain, J. Maqbool, M. Afzal and S. Lee, editors: *Towards personalized health profiling in social network*. *Information Science and Service Science and Data Mining (ISSDM)*, 2012 6th International Conference on New Trends in; 2012: IEEE, 760-765.
- [11] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi and L. Donaldson: *Harnessing the cloud of patient experience: using social media to detect poor quality healthcare*. *BMJ quality & safety*. 2013;22(3):251-5.
- [12] R. Nagar, Q. Yuan, C.C. Freifeld, M. Santillana, A. Nojima, R. Chunara and J.S. Brownstein: *A Case Study of the New York City 2012-2013 Influenza Season With Daily Geocoded Twitter Data From Temporal and Spatiotemporal Perspectives*. *J Med Internet Res* 2014;16 (10):e236,. doi:[10.2196/jmir.3416](https://doi.org/10.2196/jmir.3416)
- [13] M. De Choudury, S. Counts, E. Horvitz, editors: *Social media as a measurement tool of depression in populations*. *Proceedings of the 5th Annual ACM Web Science Conference*; 2013:47-56