Special Issue Article

Achieving veracity: A study of the development and use of an information system for data analysis in preventive healthcare

Troels Mønsted
University of Oslo, Norway

Abstract

Within healthcare, information systems are increasingly developed to enable automatic analysis of the large amounts of data that are accumulated. A prerequisite for the practical use of such data analysis is the veracity of the output, that is, that the analysis is clinically valid. Whereas most research focuses on the technical configuration and clinical precision of data analysis systems, the purpose of this article is to investigate how veracity is achieved in practice. Based on a study of a project in Denmark aimed at developing an algorithm for stratification of citizens in preventive healthcare, this article confirms that achieving veracity requires close attention to the clinical validity of the algorithm. It also concludes, however, that the veracity in practice hinges critically on the citizens' ability to report high-quality data and the ability of the health professionals to interpret the outcome in the context of existing care practices.

Introduction

Within healthcare, large amounts of data are increasingly generated from various sources serving different purposes, which together form a substantial data trace from each patient. Reuse of this data offers great potential for the development of healthcare.[1] In Denmark, practically the entire population is covered by electronic medical records (EMRs) at hospitals and in general practice.[2] In addition, a vast number of specialized clinical information systems and databases exist in both secondary and primary care, together forming a massive data repository.[3] While these data are mainly created to document or support specific care activities and thus consist of highly specialized information such as lab test results, medicine prescriptions, x-rays, and progress notes that are shaped to support local practices, the wealth of information contained can potentially serve different, important purposes through secondary data analysis. For the individual patient, this may be to

Corresponding author:
Troels Mønsted, Department of Informatics, University of Oslo, P.O. Box 1080, 0316 Oslo, Norway.
Email: monsted@ifi.uio.no

identify patterns in a specific condition by analyzing data from different sources to enable more precise diagnoses, to detect exacerbation of a chronic condition, or to identify risk patterns for use in early detection and preventive care. At a population level, secondary analysis of health data can potentially be used to detect trends across patient cases, for example, for use in public health research, to detect outbreaks, and to assist the generalizability of clinical trials.[4,5]

The work involved in collecting and analyzing data is highly laborious, and therefore the practical feasibility of secondary data analysis hinges critically on the development of information systems to support this through data mining and analysis.[6] While there has been a strong interest for the development of decision-support and expert systems based on secondary data analysis since the 1990s and while these applications have shown some potential, for example, in terms of helping health professionals reduce medication errors,[7] this interest has lately been renewed. This is not least driven by recent developments within cognitive computing and machine learning. This is represented by the IBM Watson technology that supports analysis of more complex and heterogeneous data, which has been applied within life science research[8] and oncology[9] with some success.

Development of information systems for data analysis is not without its challenges. As argued by Raghupati and Raghupati,[5] the complexity of the healthcare domain poses four challenges in connection with development of systems for data analysis: They must be able to handle the *volume* (the sheer amount of data), the *velocity* (the fast pace by which new data are produced), the *variety* (the heterogeneity of the structured, semi-structured, and unstructured data, which often requires extraction from different information system), and the *veracity* (the need to ensure the clinical precision of the data input and output of the analysis). Later research confirms that the quality of the data is decisive for its reusability.[10] While the volume, velocity, and variety of health data explicate the need for development of robust information systems that are able to collect and handle data of multiple types and from various sources, the veracity of data analysis is particularly challenging to achieve in practice, as this ideally requires the development of algorithms that to a certain extent replicate the complex sensemaking that characterizes clinical assessment and decision-making.

Based on his studies of decision-support and expert systems, Berg[11] already in 1997 argued that the development of such systems tends to be based on a rationalistic approach heavily influenced by evidence-based medicine, which defines clinical decision-making as the *"[…] conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."*[12] Hereby, many decision-support systems inscribe an approach where clinical decisions are deduced based on statistical inference, and this does not take fully into account the situated, contingent, and interpretative elements of clinical decision-making.[13] This includes *value judgments* (e.g. weighing out risks and benefits), the *patient's perspective* (e.g. considerations to the subjective priorities of the patients that may influence on clinical judgment), and the usage of *subjective probabilities*.[11] These matters are difficult to account for algorithmically, as the information is hard to reach (e.g. as the priorities of a patient are sometimes contingent on the specific situation), and as this reasoning does not explicitly follow clinical guidelines but is also a product of the interpersonal relation between the health professional and the patient, through which the clinical decision is reached through discussion of, for example, the benefits and disadvantages of specific therapeutic actions.[14] This extends the challenge of achieving veracity in data analysis, as this does not only include the clinical precision but also the ability of the system to support health professionals and patients in reaching the right decision in the specific situation. In addition, this calls for an extended view on the role of data analysis. In line with Fitzpatrick,[15] who found that clinical information systems, here, the medical record, should not merely be considered passive repositories of information but *working records* that actively shape the interactions of which they are part, this calls attention to the agency of data analysis systems in decision-making processes. In

a similar discourse, on design of patient-centric information systems, it was found that new information systems significantly shape the roles of the actors by requiring patients to take a more active role[16] and by adding increased work to the patient in ensuring that their contribution to own care is both *meaningful* and *actionable* for the health professionals.[17]

Based on a study of the development and testing of an information system aimed at performing automatic analysis of patient data to enable early detection of citizens in need of preventive healthcare, the purpose of this article is to identify the challenges of achieving veracity of the analysis, both in the development of the algorithm and when it is used in practice.

The challenge of early detection in preventive healthcare

The study is focused on the development of an information system for early detection of citizens in risk of developing lifestyle-related disease. This term describes conditions that are primarily incurred by health-risk behaviors, such as a poor diet, smoking, harmful use of alcohol, and lack of regular physical activity and include cardio-vascular diseases, cancer, chronic respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), and type 2 diabetes.[18] Development of approaches for prevention of lifestyle-related disease is a timely matter, as they have a strong impact on the life expectancy and quality of life for citizens,[19–21] and account for a large fraction of the total costs at hospitals.[22] *Targeted health checks* (health assessment offered to citizens found with particular risk of developing lifestyle-related disease) have been found to be an effective approach,[23,24] in particular, if they enable an integration of preventive health offers in general practice and community health services.[25,26] Furthermore, there is increasing interest for the use of information systems as support for behavior change.[27] Yet, effective preventive healthcare requires that citizens are screened for risk at a population level, which involves highly laborious analysis of large amounts of health data. Based on a study of the project Tidlig Opsporing og Forebyggelse (TOF, Danish for "Early Detection and Prevention"), this article presents an analysis of how veracity was achieved in the data analysis through which the risk profile of the citizens was analyzed and what challenges remain.

Methods

Setting and case

This article reports from a case study of development and use of the stratification algorithm in the TOF project. TOF is a collaboration between several Danish partners: The Region of Southern Denmark, a university research unit on general practice, an IT vendor, the General Practitioners' Trade Union (GPTU), and 10 municipalities. The project was initiated in 2009 for the purpose of developing a health intervention for early detection of citizens in risk of developing lifestyle-related disease and initiation of preventive care for citizens in risk of developing lifestyle-related disease. The primary focus of the study is the development of a stratification algorithm and the trial of this in a pilot implementation that was carried out during 3 months in 2016 with participation of two municipal health centers and 47 general practitioners from 18 clinics. In total, 9400 citizens were invited, and 2661 completed the intervention.

Data collection

This research was conducted as a case study of the development of the preventive intervention and IT support in TOF. The author has not been actively or formally involved in the development or

decision activities, but has followed the project in the period 2013–2017 through participation in project meetings and in five 3-h workshops in which project stakeholders (citizens, GPs and nurses, municipal health professionals, and patient organization representatives) were involved in the design of the IT support and collection of project documentation, and through ongoing dialogue with project participants.

The data presented in this article were collected as a qualitative study conducted with the explicit purpose of documenting the experiences of citizens and GPs when using the IT support during the pilot implementation of TOF. The data collection is based on semi-structured, qualitative interviews.[28] A total of 13 citizens (age 46–58, eight females and five males) and five GPs were interviewed. Citizens were interviewed twice, the first time early in the intervention, immediately after reporting their health information through the IT support, the second time 1–2 months after the intervention. The GPs were interviewed immediately after their participation in the pilot implementation. In total, 31 interviews (duration of 15–30 min) were conducted, and documented using sound recording.

Analysis

For the analysis, all interviews have been transcribed and analyzed through an iterative process following the principles of the constant comparison method.[29] First, open coding was conducted through which first order codes (issues raised by the respondents) were identified. Through axial coding, links between the first order codes were found, through which second order themes were assembled. Finally, these were aggregated into overarching themes through constant comparison with the original data, through which the themes that guide the presentation of the findings were established.[30,31]

The need for data analysis in preventive healthcare

In the existing organization of preventive health in the region of southern Denmark, two challenges impaired their ability to efficiently target citizens in risk of developing lifestyle-related disease. First, the responsibility for providing preventive health offers was distributed across general practice and municipal health centers. Second, no efficient means for identifying citizens in risk of developing lifestyle-related disease existed, as the current approach presupposed that they would enter preventive care at their own initiative. To address this, the TOF project developed an IT-supported preventive intervention that at its core consisted of an information system that enabled automatic stratification of citizens in risk groups. This consisted of two main components—a *stratification model* and a digital *data collection tool* that were both developed for this specific occasion. The purpose of the stratification model was to stratify citizens in four risk groups (pre-existing diagnosis, high risk, moderate risk, and low risk) based on analysis of data from two sources—citizen-reported information of risk behaviors and data extracted from the GPs' medical records. This data collection for the stratification algorithm was facilitated by a digital infrastructure consisting of the *health folder* (a web-based system through which citizens could report information about risk behavior) and a *data capture tool* (a system used to extract data from the GPs medical record). Based on the risk assessment, the citizen would then either be referred to the GP (high risk) or the municipal health center (moderate risk), and the results of the analysis be presented to the citizen and the relevant health professional through the health folder.[32] This study overall shows that the veracity of the data analysis in TOF required two levels of development. First, a thorough development and validation of the algorithm itself. Second, organizational change to align the precise logic of the algorithm with the more ambiguous and interpretative practices of

reporting health information for the citizens, and the citizens' and health professionals' practices of understanding the outcome.

Achieving veracity in the algorithm

To ensure the clinical precision of the stratification algorithm, the TOF project conducted a thorough development and testing process in the period 2009 to 2012. To ensure clinical validity, the logic and the thresholds of the algorithm were based on clinically validated models for detection of risk of developing lifestyle-related disease. First, the citizen-reported information of risk behaviors was based on the Swedish National Guidelines for Disease Prevention (http://www.socialstyrelsen. se/nationalguidelines/nationalguidelinesformethodsofpreventingdisease). This covers specific information about the citizen's risk behavior, such as alcohol consumption, smoking, amount of physical exercise, diet (intake of candy, fruit, vegetables, and fish), observable symptoms (shortness of breath, coughing), and own experience of general health. Based on this, a 15-item questionnaire was developed. The data that were required to be extracted from the GPs' medical records was based on three internationally recognized models for identifying risk of developing hypertension, hyperlipidemia, COPD, type 2 diabetes, and cardio-vascular disease; the COPD-PS screener; the Danish Diabetes Risk Model; and the Heartscore body mass index (BMI) score. Based on these, a set of data consisting of prescription codes, National Health Service disbursement codes, and International Classification of Primary Care (ICPC-2) codes was defined. Finally, the analysis conducted by the algorithm was defined to take place in two steps: In the first step, the risk of developing lifestyle-related chronic disease is identified using the three validated risk scores (the COPD-PS screener, the Danish Diabetes Risk Model, and the Heartscore BMI score). In the second step, citizens with unhealthy lifestyles were identified as citizens with an alcohol intake exceeding 14 (females) or 21 (males) units per week, an unhealthy diet, a score of 4 out of 12 positives on the Swedish National Guidelines for Disease Prevention, a BMI exceeding 30, and/or less than 150 min of physical activity per week.[32]

The stratification model was tested through a feasibility study involving four GPs and 1400 patient cases.[32] This assessment confirmed the robustness of the information system and the veracity of the algorithm under ideal circumstances. In practice, however, the pilot implementation showed that this veracity is difficult to achieve.

Achieving veracity in practice

From a clinical perspective, the ability of the stratification algorithm to analyze precisely the risk profile of a citizen was well documented. The pilot implementation, however, showed that the algorithm in practice was not always experienced as precise by neither the citizens nor the GPs. In a typical case, a citizen who had been stratified as high risk reported that this category confused her:

> Actually, I couldn't understand it, and the nurse in the general practice clinic couldn't understand it either. I had answered that I was smoking 35 years ago or so, so it can't be a risk any longer. But it is because if you answer the same questions so many times, you may end up giving a wrong answer at some point. (Citizen, female, age 59)

In this case, the primary reason for being in high risk that immediately came to the citizen's mind was that she had previously been a regular smoker, which she had reported in the questionnaire. However, she had successfully quit smoking 35 years in advance of the intervention, and in

spite of being slightly overweight, she did not see herself as a relevant subject of preventive care. In her experience, veracity was therefore not satisfactory. When the citizen later was referred to her GP as the next step in the intervention, the confusion was at first shared by the GP, as he could not see any immediate reason to offer her preventive care considering her health profile, not least because she had already taken the necessary step toward improving her long-term health by quitting smoking a long time ago. He did, however, interview her more closely about her habits, including her diet. Here, it came to light that by making relevant adjustments, she could reduce her intake of fat. In line with this, later blood samples showed that she had slightly increased cholesterol numbers. This case demonstrated the veracity of the algorithm and its ability to detect a citizen who, on a long timescale, would benefit from adjusting habits, presumably by linking the previous status as smoker with light overweight. The case nevertheless also demonstrates two types of work that were necessary to achieve veracity in practice. First, a key prerequisite for the veracity of the stratification algorithm in TOF was the availability of precise health data. The data extracted from the GPs' medical records were by definition highly structured and compliant with the logic of the stratification algorithm, but the citizen-reported data proved to be a more critical challenge. While a high degree of structure was imposed by the questionnaire, some of the categories in the questionnaire were ambiguous for the citizens. For example, it required the citizens to report if they had a daily intake of candy. For the majority in this study, this was, however, not a sufficiently meaningful category, as most differentiated between a large and small intake. To represent their lifestyle precisely, some were therefore prone to interpreting the question and answer no when they assessed that they had a small and non-harmful intake.

> I know that large amounts of candy will make me ill, but that is not what I am eating. It doesn't ask how many bags of candy you eat, but if you eat candy every day. And I do that, but that is perhaps one piece of chocolate, not a large bar. (Citizen, female, age 46)

Likewise, the dichotomy between being either a smoker or non-smoker was sufficiently meaningful for all citizens. While even occasional use of tobacco is harmful in a strictly clinical logic as represented by the algorithm, some citizens did not consider themselves smokers, as they only smoked rarely and therefore felt in full control of this habit. For this reason, some were hesitant at categorizing themselves as smokers. This shows how the citizens' interpretation of the questions could critically affect the veracity of the data to be analyzed by the stratification algorithm. Another recurring issue for the citizens was that some of the questions in the questionnaire were experienced as redundant. For some, this caused complacency or indifference to the precision of the answers and consequentially the veracity of the analysis conducted by the algorithm.

> And there you begin to get annoyed, because it is the same things that are asked over and over again and at the end you begin just to tick boxes. "I don't want to do this anymore" I thought, because it is annoying when questions are very similar. (Citizen, female, age 59)

Overall, the reporting of the health information required some degree of interpretation and understanding of the importance of reporting precise information by the citizens. As a consequence, the pilot implementation showed that the majority of citizens who successfully completed the intervention had a high level of health competence.

Second, the output of the stratification algorithm required interpretation in order to align fully with the expectations of the GPs. As exemplified in the previous case, the algorithm detected risk behaviors that were only likely to cause lifestyle-related disease on a long-term scale. However, in the current preventive healthcare system, the offers in general practice are mainly geared toward

citizens with more acute risk profiles. In fact, the GPs in this study unanimously expressed that they, due to the heavy workload in general practice, felt obliged to prioritize acute illness, early detection of current undiagnosed diseases, and preventive care for citizens with very strong risk behaviors, rather than to spend resources on citizens with high health competence and long-term risk:

> We generally want to see those for whom we can make a difference here and now. Part of my motivation for going into this was that I could keep the "healthy marathon runner" out of the clinic, and get to see those who are otherwise just sitting down at the local pub, and only rarely come here in the clinic. (GP)

The algorithm created a new and laborious type of work for the GPs. Not only did they have to verify the details in the citizen-reported health information through the interview, but they also had to assess the citizens' fit with the specific preventive health offers they could provide, such as supervised and monitored weight loss. In addition, the GPs had to prioritize the citizens in relation to other patients through assessment of their motivation and willingness to engage in lifestyle change, as well as their ability to manage this on their own hand based on an individual assessment of their health competence. The stratification algorithm therefore did not take fully into account all of the contextual information required for the GPs to assess if a given citizen should be offered preventive care. Overall, the study therefore shows that the veracity of the algorithm in its current state of development is dependent on the health competence of the citizens and hereby their ability and willingness to report honest and precise health information, and the ability of the GP to interpret the outcome in context.

Concluding remarks

As argued in this article, a prerequisite for the successful development and implementation of data analysis in healthcare is not only that the information system and algorithm are sufficiently advanced and robust to handle large amounts of heterogeneous data but also that it produces an output with sufficient veracity.[5] This study shows that the veracity hinges critically on the organizational feasibility and a sufficient alignment with existing or emerging practices in the healthcare system. In the case of TOF, the high degree of precision in the algorithm, and hence its dependency on reliable, structured data, imposed work on the citizens in reporting this. However, since some of the very specific categories with which the algorithm operated were experienced as ambiguous to many citizens, the current use of the algorithm mostly appealed to those with a high level of health competence. Likewise, the output of the algorithm required interpretation, both to verify the health information that formed the foundation of the stratification of the individual citizen and to assess their relevance for the specific preventive health offers available. In the case of TOF, this in particular surfaced as a friction between the ability of the algorithm to detect risk of developing lifestyle-related disease on a very long-term scale and the need for the GPs to prioritize care for citizens with more acute diseases or health profiles. These findings draw some implications for the further development and implementation of data analysis algorithms in healthcare. First of all, experience from TOF confirms the argument of Raghupati and Raghupati[5] that the veracity of the algorithm must be ensured, and that this can be achieved through thorough systems development and testing, and by basing the logic of the algorithm on highly validated clinical evidence. However, the study also shows that veracity in practice only emerges through a fit with the enacted clinical practices, including the ways that citizens and GPs in practice make sense of a risk profile and how they prioritize their need for preventive care. This suggests two relevant future orientations for development of such algorithms. First, it is required to consider the specific thresholds in the

algorithm and to acknowledge that analysis criteria that are precise from a clinical perspective may not always integrate well in practice. In the case of TOF, a relevant response could be to redefine the high-risk category and only include citizens with more urgent care needs. Second, it is necessary to take into account that implementation of information systems typically also involves organizational change. In the case of data analysis algorithms, development also involves redesign of the diagnostic practices of which they become part.

Declaration of Conflicting Interests

Funding

References

1. Murdoch TB and Detsky A. The inevitable use of big data to health care. *JAMA* 2013; 309: 1351–1352.
2. Kierkegaard P. eHealth in Denmark: a case study. *J Med Syst* 2013; 37: 9991.
3. Ventrakaman V, Priya M and Ussing A. Mapping the healthcare data landscape in Denmark: Leapcraft, Copenhagen Health Clusters, 2015, http://healthtech.leapcraft.dk/
4. Cahan A, Cahan S and Cimino JJ. Computer-aided assessment of the generalizability of clinical trial results. *Int J Med Inf* 2017; 99: 60–66.
5. Raghupati W and Raghupati V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014; 2: 3.
6. Raghupathi W. Data mining in health care. In: Kudyba SP (ed) *Healthcare informatics: improving efficiency and productivity*. Abingdon: Taylor & Francis, 2010, pp. 211–223.
7. Kaushal R, Shojania KG and Bates DW. Effects of computerized physician order entry and clinical decision support system on medication safety. *Arch Intern Med* 2003; 163: 1409–1416.
8. Chen Y, Argentinis E and Weber E. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther* 2016; 38: 4688–4701.
9. Zauderer MG, Gucalp A, Epstein AS, et al. Piloting IBM Watson oncology within Memorial Sloan Kettering's regional network. *J Clin Oncol* 2014; 32: 15.
10. Sollie A, Sijmons RH, Helsper C, et al. Reusability of coded data in the primary care electronic medical record: a dynamic cohort study concerning cancer diagnoses. *Int J Med Inf* 2017; 99: 45–52.
11. Berg M. *Rationalizing medical work: decision-support techniques and medical practices*. Cambridge, MA: The MIT Press, 1997.
12. Sackett DL, Rosenberg WMC, Muir JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71.
13. Montgomery K. *How doctors think: clinical judgment and the practice of medicine*. Oxford: Oxford University Press, 2006.
14. Mønsted T. The use of narratives in medical work: a field study of physician-patient consultations. In: *Proceedings of the 12th European conference on computer supported cooperative work ECSCW 2011* (eds S Bødker, NO Bouvin, W Lutter, et al), Aarhus, 24–28 September 2011, pp. 81–100. New York: Springer.
15. Fitzpatrick G. Integrated care and the working record. *Health Inform J* 2004; 10: 291–304.
16. Nunes F, Andersen T and Fitzpatrick G. The agency of patients and carers in medical care and self-care technologies for interacting with doctors. *Health Inform J*. Epub ahead of print 1 June 2017. DOI: 10.1177/1460458217712054.
17. Andersen T, Bansler J, Kensing F, et al. Alignment of concerns: a design rationale for patient participation in eHealth. In: *Proceedings of the 47th Hawaii international conference on system science*, Waikoloa, HI, 6–9 January 2014, pp. 2587–2596. New York: IEEE.

18. World Health Organization. *Global status report on noncommunicable diseases*. Geneva: WHO, 2014.

19. World Health Organization. *Diet, nutrition, and the prevention of chronic diseases*. Geneva: WHO, 2003.

20. World Health Organization. *Global status on noncommunicable diseases*. Geneva: WHO, 2010.

21. Juel K, Sørensen J and Brønnum-Hansen H. Supplement: risk factors and public health in Denmark. *Scand J Public Health* 2008; 136(1suppl): 11–227.

22. Glümer C, Hilding-Nørkjær H, Jensen H, et al. *Sundhedsprofil for region og kommuner 2008*. Glostrup: Region Hovedstaden, 2008.

23. Si S, Moss JR, Sullivan TR, et al. Effectiveness of general practice-based health checks: a systematic review and meta-analysis. *Br J Gen Pract* 2014; 164(618): 47–53.

24. Engelsen C, Den Koekkoek PS, Godefrooij MB, et al. Screening for increased cardiometabolic risk in primary care: a systematic review. *Br J Gen Pract* 2014; 164(627): 616–626.

25. Andersen JH, Thorsen T and Kousgaard MB. *Samarbejde mellem kommuner og almen praksis om henvisning til kommunale sundhedstilbud for patienter med kroniske sygdomme*. Copenhagen: Forskningsenheden for Almen Praksis, 2014.

26. Helbech Hansen B. *Omfanget af henvisninger fra almet praktiserende læger til kommunale sundheds- og forebyggelsestilbud*. Copenhagen: Kommunernes Landsforening, 2015.

27. Kelders SM, Oinas-Kukkonen H, Oörini A, et al. Health behavior change support systems as a research discipline: a viewpoint. *Int J Med Inf* 2016; 96: 3–10.

28. DiCicco-Bloom B and Crabtree BF. The qualitative research interview. *Med Educ* 2006; 40: 4314–4321.

29. Glaser B and Strauss A. *The discovery of grounded theory: strategies in qualitative research*. London: Weidenfeld & Nicolson, 1967.

30. Eisenhardt KM. Building theories from case study research. *Acad Manag Rev* 1989; 14: 532–550.

31. Strauss A and Corbin J. *Basics of qualitative research: grounded theory procedures and techniques*. Newbury Park, CA: SAGE, 1989.

32. Larsen LB, Søndergaard J, Halling A, et al. A novel approach to population-based risk stratification, comprising individualized lifestyle intervention in Danish general practice to prevent chronic diseases: results from a feasibility study. *Health Informatics J* 2017; 23: 249–259.