IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 28, 2018, accepted November 28, 2018, date of publication December 7, 2018, date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885586

What Clinics Are Expecting From Data Scientists? A Review on Data Oriented Studies Through Qualitative and Quantitative Approaches

LINA XU[1], MONIKA SIMJANOSKA[2], BOJANA KOTESKA[2], VLADIMIR TRAJKOVIKJ[2], ANA MADEVSKA BOGDANOVA[2], KRISTINA DRUSANY STARIČ[3], AND FEDOR LEHOCKI[4]

[1]University College Dublin, Belfield, Dublin 4, Ireland
[2]Faculty of Computer Science and Engineering, Saints Cyril and Methodius University of Skopje, 1000 Skopje, Macedonia
[3]University Medical Centre Ljubljana, 1000 Ljubljana, Slovenia
[4]Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava, 811 07 Bratislava, Slovakia

Corresponding author: Lina Xu (lina.xu@ucd.ie)

This work was supported by COST (European Cooperation in Science and Technology), COST Action, under Grant ENJECT TD 1405.

ABSTRACT Ensuring healthy lives and promoting well-being for all, at all ages, is one main objective for sustainable development proposed by the United Nations. The concept of connected health (CH) has been proposed to achieve that goal by connecting all the stakeholders through enabling Telehealth technologies. This paper has first presented an overview of the whole picture of CH along with the data collection process in CH. In the whole picture of CH, translational medicine (TM), as a rapidly growing discipline in biomedical research, aims to expedite the discovery of new diagnostic tools and treatments by using a multi-disciplinary and highly collaborative approach. It has been introduced to bridge the technique gap between the clinics and data scientists, particularly targeting on health related data analysis and evidence medicine. What clinicians are expecting and what researchers can offer will/should all be defined and clarified through TM. To further facilitate the communication between the clinicians and the researchers, electronic health records (EHRs) are often applied in place. This paper first reviews the evolution history of EHR and its current status and standards. Then a detailed and comprehensive discussion on data analysis techniques applied in TM through both quantitative and qualitative approaches is elaborated. We reveal that future work in TM should put an emphasis on data oriented qualitative analysis, using advanced techniques from the artificial intelligence domain to predict health risk, such as heart attacks and early stages of cancers. Multidisciplinary research in the Internet of Medical Things across health science, data science, and engineering will be the main challenge in TM.

INDEX TERMS Connected health (CH), translational medicine (TM), electronic health record (EHR), qualitative and qualitative data analysis, Internet of Medical Things (IoMT).

I. INTRODUCTION

Connected Health (CH) aims to bring patients, clinicians and health science researchers all together to help the society to answer one question: how to connect patients, therapists and care-givers to deliver the optimum health results in an era of stretched resources and increasing demands [1]. Internet of Medical Things (IoMT), which is evolved from Internet of Things (IoT) has been introduced to improve the infrastructure for connection and communication in CH. Figure 1 has shown current stage of CH, specifying CH systems' scope and functionalities. As we can see, the end users normally include patients/outpatients and the elderly. The IoMT sensing devices, are utilised for the purpose of collecting health related data and providing real time monitoring results. Afterwards, the healthcare providers will pull the sensed data from the end users and then deliver it to the health science researchers for predictive analysis [2]. During this process (from left to right in Figure 1), the data converges from the end users to the researchers. The main focus of CH is the connection between the end users and the healthcare providers. Based on the provided data, the researchers either perform quantitative or qualitative analysis and then propose suggestions and summative conclusions to the clinicians/hospitals. Therefore, in general the GPs and healthcare

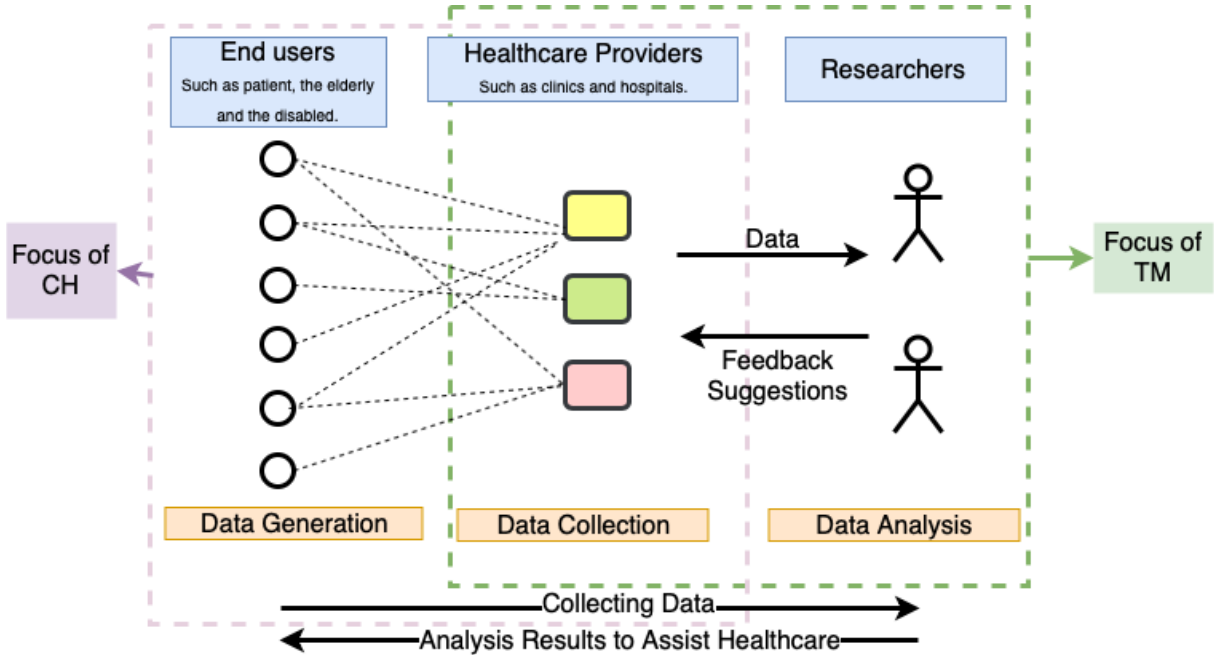

FIGURE 1. Overview of the data flow in connected health (CH).

professionals can give more accurate and personalised treatments for the end users based on the feedbacks and analysis results from the researchers. Through this way, the predictive analysis results are distributed to the end users, who at the end are benefitted from contributing their own personal data. To smooth the original data and the analysing results exchanging process between the clinicians/hospitals and the researchers is one of main tasks in CH, which has begun to draw people's attention recently. This part is also referred to as Translational Medicine (TM). Traditional CH is focusing on connecting patients, healthcare providers, researchers and other relevant stakeholders, with an aim to guarantee the data sharing and exchanging between different entities. The main purpose of CH is timely sharing, patients' condition monitoring and accurate information presenting through intelligently integrating and utilising of the health data, the devices and the communication software/hardware platforms provided by the IoMT systems. TM is introduced specifically to bridge the domain knowledge gap between the data scientists and the healthcare providers by enabling effective and efficient communication. There are two principle objectives in TM. One is to provide means to share the data collected by the clinicians with the researchers. The other one is to provide helpful suggestions and recommendations concluded by the researchers through predictive analysis to the doctors.

Data scientists need to perform analysis on the health data, which is normally abstracted from the health records provided by the clinics and hospitals. In order to make the health records more convenient to use and track and also easier to understand, Electronic Health Records (EHRs) is introduced. This format of information composes the majority of the data used in TM. For data scientists, being able to understand the needs from the clinics/hospitals is the most essential and critical step towards the goal of meeting the clinicians' requests. The success of this process determines the significance of the TM impact in the whole picture of CH. From this point of view, we will discuss later in this paper that qualitative analysis in TM should draw more attention and effort to fulfil the clinicians' needs.

In this paper, firstly we will discuss the health data collection from a general overview perspective. Then specifically, EHRs related issues are presented and its importance is revealed. At the end, the state-of-art data analysis technologies applied in TM are illustrated and the related work is analysed from two aspects: the quantitative ones and the qualitative ones Through the discussion and analysis, we indicate that the lack of data oriented qualitative research is the urgent problem that should be addressed in order to meet the physician's expectations. Hence, more studies and work are required in this direction, demanding close and intense collaborations between health care providers, health science researchers and data scientists.

Based on the above understanding of CH and TM, the rest of the paper is organised as follow: In Section II, we present the related work and the motivation for our work. Section III shows the background knowledge for data collection in CH. EHR and TM along with their related issues are presented in Section IV and Section V respectively. Then a comprehensive discussion and analysis from a quantitative and qualitative perspective is delivered in Section VI. Finally the conclusions are drawn in Section VII.

II. RELATED WORK AND MOTIVATION

A general overview on the technologies and strategies in Telemedicine and Telehealth aiming to improve patients' healthcare has been delivered in [3]. It has indicated that CH cannot only provide services in evidence-based medicine, but also can educate patients to gain health knowledge to achieve better self-management and self-care. As recently data analysis has been becoming more and more popular, this technology has been applied widely to assist health problem discovering and clinical decision making [4], [5].

A lot of novel studies and reviews in CH have been carried out from a quantitative analysis perspective [6]–[9]. However, in CH, especially in TM, qualitative analysis has been proved more useful and the results have also been well accepted in various domains. Nevertheless, limited reviews have been conducted from a qualitative perspective. Many qualitative analyses in TM area have been done on a high overview level [10]–[13]. A framework approach for qualitative data analysis has been introduced by [14], which enables the researchers to explore data in depth and meanwhile maintain an effective and transparent audit trail. A qualitative framework analysis on patient narratives has also been carried out in [15]. The study in [16] has provided an analysis of five main traditions of qualitative research and also detailed the study design, question development, data collection and analysis, result summarising and interpreting.

A review has been presented in [17] on the potential and practice in 763 published articles where Qualitative Data Analysis Software (QDAS) is used, including ATLAS.ti and NVivo. It is focusing on investigating who is using these tools and how they are being used to support the proposed research. Qualitative analysis is not only utilised for analysing patients' data in health area. A systematic review of qualitative studies on culturally and linguistically diverse healthcare students' experience of learning in a clinical environment has been delivered in [18]. The discussion on qualitative research ethics by making explicit conceptual and practical tensions that emerge at various stages of the research process and exploring issues associated with strategies for upholding confidentiality has been expanded in [19]. Limited studies have been focusing on data supported qualitative analysis. It has been indicated in [20] that qualitative research can generate large volumes of textual data in the form of transcripts and observational field notes. The work should be done by highly skilled researchers rather than novice. However, it is not common for a researcher to have both advanced health domain knowledge and advanced data analysis skills.

In [21], the authors aim to synthesise and identify important barriers to hypertension control as reported by patients and healthcare providers through a systematic review on qualitative and quantitative studies. However, limited research has been done for qualitative analysis from a technical perspective, focusing on content/script analysis and language processing. Qualitative methods can be used to explain the complex concepts within medical practice. Especially in clinical care, qualitative research generally involves analysing healthcare context, written texts, group conversations and seeks approaches to understand the meaning and underlying facts from the experimental results in a study case. The data gathered from EHRs is holding a huge amount of useful information that can be transformed into lifesaving actions. In addition, with different analytical approaches of the data, the costs of the diagnostics and treatments can be greatly reduced with more focused and predicable management. Qualitative research is an effective and efficient way to discover new knowledge and findings abstracted from EHRs to our existing healthcare systems, with a high possibility to improve the medical practice and therefore the entire healthcare system. To the best of our knowledge, currently no review papers have been conducted by comparing data analysis approaches from the quantitative and qualitative perspectives.

III. DATA COLLECTION IN CONNECTED HEALTH (CH)

CH is generally described as a paradigm with a purpose to improve the communication between patients and clinicians through all possible means. The patients, considered as the end users, are generating the data. This data is collected through many IoMT devices and equipment that are designed for monitoring health condition and physiological parameters. Generally, the clinicians are only collecting the data and using the knowledge that is concluded from the data to support their decisions. However, how to translate the collected data into a set of meaningful information is out of their expertise, but normally requires careful design and analysis. Therefore, the data scientists are involved to take over the job of data mining and predictive analysis. The existing approaches suffer problems related to either ethical or technological constraints. All the stakeholders, including the policy makers, the healthcare providers, the clinicians, the patients and the researchers are all facing the same barriers when processing the data, including security and data understanding [2]. To overcome those key barriers, two approaches have been proposed in [1]. The first one is to optimise and secure the procedure of obtaining the users' data. The second one is to analyse the data and then to understand the underlying relationship between the data and the patients' health conditions.

Massive volumes of health related data sets have been gathered from the patients and the general public. Advanced analytics on the data can provide meaningful information and knowledge, which will benefit all the stakeholders in the whole picture of CH. In order to adopt big data idea and next-generation analytics into health research and clinical practice, novel understandings of the existing systems, alone with new training tools are still missing in this field. If utilised appropriately, the health data pool can be an exceptional source of knowledge to conduct a smart and cognitive health care system [22]. Research on how big data analytics can

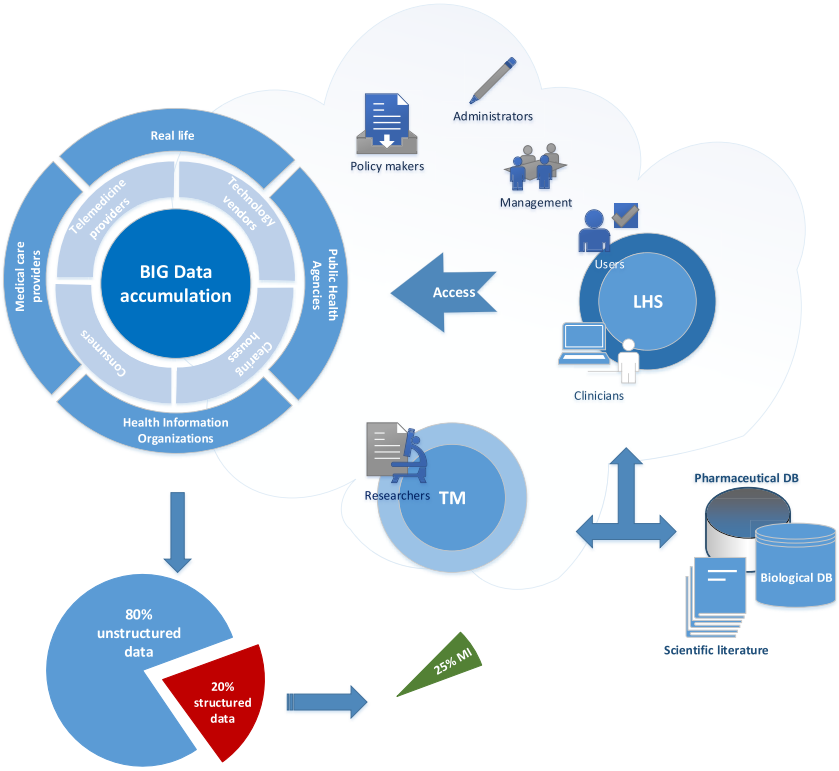

FIGURE 2. Connected health information exchange.

influence and transform traditional healthcare lies in many different fields:

1) Support research – Part of the research is to demonstrate how the CH data can be used in the development of evidence-based medicine.
2) Transform data into useful information – Enabling extracting knowledge from the EHRs is also in the key task in TM research (some of the approaches and techniques are reviewed in this paper).
3) Build an ecosystem – IBM's Watson [23] is an example, consisting of 21 supercomputers as the subsystems to collect cancer treatment information.
4) Others – Issues indirectly related to this research include supporting self-care and self-learning, motivating healthcare providers and arousing health problem awareness.

CH data is collected from many various sources. The amount of data will increase up to 25,000 PB by 2020 based on McKinsey Global Institute's prediction [24]. Existing report states that 80% of this data is unstructured. Accurately analysing the data can decrease the patients' mortality by 20% [25]. As it has already been approved, Big Data's importance cannot be neglected in the CH field and it has become the main trend since 2016 [26]. Specifically, there are urgent needs for new tools and skills to transform such a large amount of data into meaningful information. The predictive analytics are applied, aiming to increase the accuracy of the patients' diagnosis and deliver personalised healthcare. To fully utilise the biomedical data, the National Institutes of Health (NIH) has initiated a Big Data to Knowledge (BD2K) program [27], trying to extract of meaningful information from the biomedical clinical data. This health data is usually captured in hospitals and clinics without the instructions or advices from researchers and data scientists. Therefore, it must be appropriately annotated in order to be comprehensible for research purposes. Generally, the data can be either public or private, depending on the scenarios. Publicly available data includes the public resources, such as medical guidelines, scientific literatures and biomedical databases; On the other hand, the patients' personal data and the pharmaceutical records are usually considered as private.

Transforming data into useful information can inspire new theoretical models and also assist to conduct more effective and efficient healthcare systems [1]. However, how to measure the effectiveness and efficiency of the models is still an unsolved problem which requires comprehensive methodologies. In this paper, we will review the existing data analysis research carried out through both quantitative and qualitative approaches.

Figure 2 illustrates the relationship among all the stakeholders that have been previously discussed in the paper: the data flows and exchanges between them. The data can

be collected from routine practice, medical care providers, public health agencies, national health organisations, technology vendors, consumers, insurance claims, etc. Normally, a database is composed of 80% of unstructured data and only 20% structured data following specific EHR standard formats [25]. Even in a well-notated data pool, only 25% of the data can be used in research. In order to derive meaningful information (MI) from the data pool, additional analytical tools based on machine learning techniques are essential to build Learning Health Systems (LHS). The derived MI then can be transformed into domain knowledge based on experts' insight. The knowledge translation can be applied to clinical practice along with predictive analytics and therefore to improve personalised healthcare.

IV. ELECTRONIC HEALTH RECORD (EHR)

A. THE BIRTH OF EHR

The analysis on the combination of medical records and insurance claims can help to improve the healthcare management system and the claim process. It can also provide early identifications for certain health risks by comparing the doctor recommendations with the evidence-based practices [26] and therefore reduce hospitalisation. The desire to deliver the accurate information to the right hands at the right time has motivated the development of many standards for medical data recording. In this context, data can be mostly useful if gathered in a digital format. EHR is introduced to collect all kinds of patients' data which is obtained from resources such as demographics, clinicians' notes, medication records, science laboratory, etc. A complete documentation is also expected in order to provide an easy access for all medical institutions, deliver reliable healthcare to patients and support evidence-based practice for clinicians. However, fatal consequences can happen if the EHR systems are not appropriately implemented or they are misunderstood/misused by the clinicians and researchers. Several serious medical incidents that were caused by improper EHR implementation have been stated in [28]. The importance of EHR has been well accepted and understood. A lot of related studies have been done, aiming to provide regulatory requirements to evaluate the efficiency and safety of the EHR systems. Due to the lack of consistency in policies, feasibility in principles, high quality practices guidelines and regulations in electronic records, EHR systems might be suffering from the following weaknesses [28]:

- Design faults;
- Improper system usage;
- Inappropriate documentation capture;
- Copy/Paste risks;
- Templates misuse;
- Decision Support Systems complete reliance.

In order to overcome the above mentioned disadvantages, strategies and policies should be established by the healthcare providers. They should focus on providing healthcare training, prevention of treatment errors and standardising the presentation of safety alerts [28].

Currently in CH, most of the data is collected for the purpose of delivering healthcare evidence rather than facilitating scientific research. The raw routine data, also referred to as real-life data, cannot provide much useful information to clinical use without deep level analysis. To reveal the relationship between the patients' health condition and their raw real-life data, new strategies for data linkage and new analytical approaches specifically for healthcare are demanded [29]. The Finnish project MyData [30] is proposed to provide a human centric approach for organising the personal data. The services are individually controlled, enhanced with machine readable and open access features available for users. It relies on a set of open formats and standards and it also can support interoperability between various storages. The goal is to provide reusable resources (personal medical records) to different sectors.

Allowing interchange of data among several sectors requires high level of integrity and interoperability. However, in reality, the clinical and administrative data in EHR systems is normally stored in an unstructured format. The framework or format in which the data is stored has a huge influence on the efficiency of data analysis. The detailed study on the impact has been conducted in [31]. Each clinical system has a unique and dynamic work flow which leads to a customised EHR system in order to adapt to the local users' requirements. Special tools are needed for evaluating and tracking the data availability, quality and transformations. Natural language processing is essential to capture useful data out of the unstructured records. For example, the colorectal cancer screening EHRs collected during clinical practice would require careful interpretations and transformations before the data can be used for research [31].

The latest efforts on the EHRs integrity and interoperability have been discussed in an interview with the Health2047's CEO – Dr. Doug Given [32]. The interoperability was discussed not just for the purpose of transferring data between different systems, but also advancing data usage in later stage. Dr. Doug Given has indicated that the problem is not all about the technical implementation. It also needs to provide a competitive business strategy. An estimation of less than 25% of data collected by EHRs is useful, and all the CH stakeholders must work together to improve this figure.

A standardised EHR format is extremely hard to achieve, thus the interoperable standards are developed for data exchange between different EHR formats. In 2013 there has been a Structured Data Capture initiative [33] by the Standards & Interoperability (S&I) Framework with an aim to establish EHR interoperability. The goal is to use interoperability infrastructure to:

- access a template that contains structured data,
- automatically populate the template from existing EHR data, and
- store or transmit the completed template to another appropriate organisation or researcher.

In a word, standards should be placed when designing the structure of the EHR template and the desired electronic

content. There also should be standards defining how EHRs should interact with those standards and how one can efficiently associate these EHR templates with the data extracted [34].

In 2015, Office of the National Coordinator for Health Information Technology from the U.S. Department of Health and Human Services released an interoperability roadmap intended for those who will build the interoperability infrastructure and for those who will use that infrastructure. The roadmap is composed of three main parts [35]:

1) Drivers – a supportive payment and regulatory environment;
2) Policy and technical components
3) Outcomes – individuals will have access and can contribute to the infrastructure; providers can share and use patient information from all sources.

Achieving interoperability will enable true LHS. LHS is a cycle that describes the use of advanced technology infrastructure for analysing healthcare and research data (extracted from EHR), which is then translated into clinical knowledge and applied into clinical practice [36], [37]. Health Evaluation Through Logical Programming (HELP and HELP2) presented in [38] is one of the most original medical systems that can support electronic medical records. The importance of allowing easy access to patient data, supporting continuous revision of the strategies, and enabling clinical reports to assist decision making in CH is described in [38]. In [39], the authors have indicated that the medical information in computable/digital format can hugely accelerate the knowledge transformation process from the laboratory bench to the patients' bedside, which in turn proves the importance of EHR. It also states that high quality LHS can be developed based on the proper use and analysis of EHRs. The benefits of LHS are also evidenced and presented in [40].

B. EHR STANDARDS

Different formats of EHRs indicate various levels of interoperability. Therefore, it is essential to regulate EHRs through standards. Herein several existing EHR standards are discussed.

1) VIRTUAL DATA WAREHOUSE (VDW)

The VDW is a virtual set of parallel databases that are constructed by extracting data directly from the local EHR systems. VDW can transform the data to align with the standard variable names and the coded values. Since the data comes from various sources, VDW may include different coding patterns and specifics. Therefore, it can be seen as a pool of information where analytics datasets can be constructed rather than an analytic dataset directly. VDW is a solution for standardising and pooling EHR data for multi-site research [41].

2) HEALTH LEVEL SEVEN (HL7)

HL7 [42] is a non-profit, ANSI (American National Standards Institute) accredited standards, which is developed to providing a comprehensive framework and a set of related standards for information exchange, integration, sharing and retrieval from EHRs. It aims to support the clinical practice and management and delivery and evaluate healthcare services. It can enable the exchange of demographics and other textual information. It is designed to provide a list of specifications for making the existing systems interoperable rather than to provide a real software system. In the strategy descriptions, the interoperability is defined in three different contexts. It can have influences on how to design the software, storage the data and use the information:

- Technical interoperability – enable the ability to accept data from each other in a multiple system context without intervention.
- Semantic interoperability – need domain-specific learning to understand, interpret and utilise the data.
- Process interoperability – enable process coordination and allow human understanding.

HL7 standards are divided into categories depending on their usages, including primary standards, foundational standards, clinical and administrative domains, EHR profiles, implementation guidelines, healthcare rules and references, and education and public awareness.

3) SNOMED CT

It is extremely hard to form a standard that can maintain the semantic meaning. Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT) has been developed into a logic-based healthcare terminology over a period of 40 years [43]. SNOMED's ability has been evaluated through experiments by Elkin et al. [44]. The results have shown that SNOMED can accurately represent over 90% of the medical terms commonly used in medical reports.

V. TRANSLATIONAL MEDICINE

A. CLINICIANS EXPECTATION

Scientists and researchers try to fulfil the doctors' expectations based on the understanding of the EHRs collected in the GPs. The involved work and technologies can all be considered in the scope of TM research. TM can be seen as a bridge between the researchers and the clinical practices. There are three main steps that need to be carried out in TM, as specified in Figure 4. Public healthcare data from the EHRs including pharmaceutical and genomic data should be used. When including genomic data, standards for data messaging, annotation and representation have to be considered. The road map to include genomic data into EHRs is described in Figure 5. We have reviewed the existing methodologies and approaches that have been published in the recent literature. Through analysing and discussing the related issues and existing limitations (both ethical and technological), we have proposed a road map that can be used to overcome the problems, as described in Figure 3. Following the proposed road map, we can show that TM is a good evaluator of the overall data

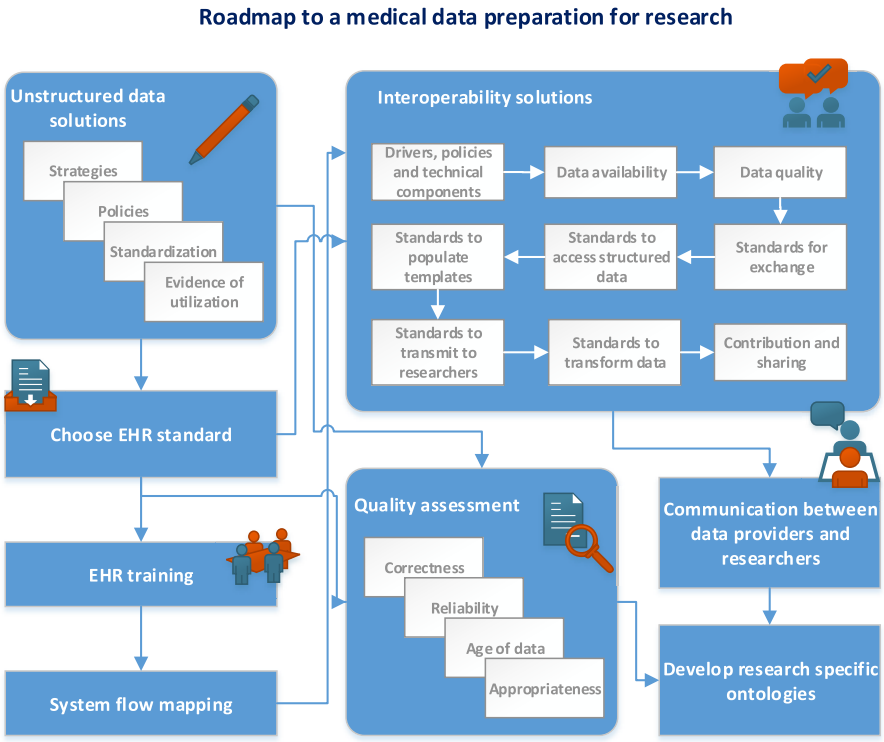

FIGURE 3. Roadmap to overcome existing data-related issues.

quality and there is a need for standards for the development of better decision making systems.

B. SUCCESSFUL STORIES

The data from the hospital is regularly analysed and used to improve clinical practice. The outcomes of the treatments vary as the practice changes. The analysed results are used as feedbacks to suggest that certain proper changes should be made. [45] has shown that head circumference of the baby is one of the factors that influences the incidence of anal sphincter injury through analysing the related EHRs. Bases on this result, if the head circumference of the baby for the second delivery is larger than that for the first deliver, it is advised that the woman should have birth with caesarean section instead of vaginally to protect anal sphincter from injury.

VI. DISCUSSION: A QUANTITATIVE AND QUALITATIVE ANGLE

A. MEANINGFUL INFORMATION (MI) DERIVATION

The process of MI derivation is a matter of research objectives that highly depends on the data sources (nature of the problem), i.e. what is measured and what information do we want to extract. Questionnaires, surveys and systematic measurements, are examples of a quantitative research design; whereas the deep interviews, the textual data and sometimes graphical data are examples of a qualitative research design. In quantitative research, the response categories are predefined and the direction of the study is based on a large body of literature. On the other hand, the qualitative research is composed of non-deterministic categories (i.e. the participants speak of their own experience and provide personal feedbacks) [46], [47]. The choice of methodologies for data analysis highly depends on the researchers' own knowledge and personal experience. The researchers with a technical background, would rather choose quantitative approach. The researchers that have good insight in this area will be able to conduct an overall viewing and hence usually choose the qualitative approach [47]. However, the overwhelming volumes of studies in the qualitative research can cause serious problems, such as leaking the information and data that the participants are not willing to share, which in turn can affect the quality of the interview answers, the accuracy of the collected information, and then the final research findings. A study on the reflexivity in qualitative analysis has confirmed that the level of researchers' involvement into the research can influence all the parts of the process, from choosing the participants, collecting and analysing the data, to concluding the results [48].

An overview on the applications with the purpose to derive MI from EHRs from both qualitative and quantitative

TABLE 1. Application of research design approaches for MI derivation.

Authors	Research design	Methods	EHR Data
Alicia L. Nobles et al. [49]	Quantitative	Statistical (t-test)	patient demographics, diagnoses, and procedures
A Rosemary Tate et al. [50]	Quantitative	Statistical (distributions, correlations)	demographics, clinical data
Taxiarchis Botsis et al. [51]	Quantitative	Descriptive statistics	patient demographics, diagnoses
Sheena Dungey et al. [52]	Quantitative	Statistical (means, medians, correlations)	patient demographics, diagnoses
Max J. Romano et al. [53]	Quantitative	Statistical (multivariable logistic regression)	patient demographics
Danielle G. T. Arts [54]	Qualitative	Interview	patient health and demographics
Jimeng Sun et al. [55]	Quantitative	Statistical (5-fold cross-validation)	clinical
Lindsey A. Knake et al. [56]	Quantitative	Statistical (Chi-square test)	patient demographics, diagnoses
Craig D. Newgard et al. [57]	Quantitative	Statistical (kappa, ICC and Bland-Altman plots)	demographics, clinical
Charlene R Weir et al. [58]	Qualitative	Interviews	clinical
Gloria Ser et al. [59]	Qualitative	Interviews	demographics

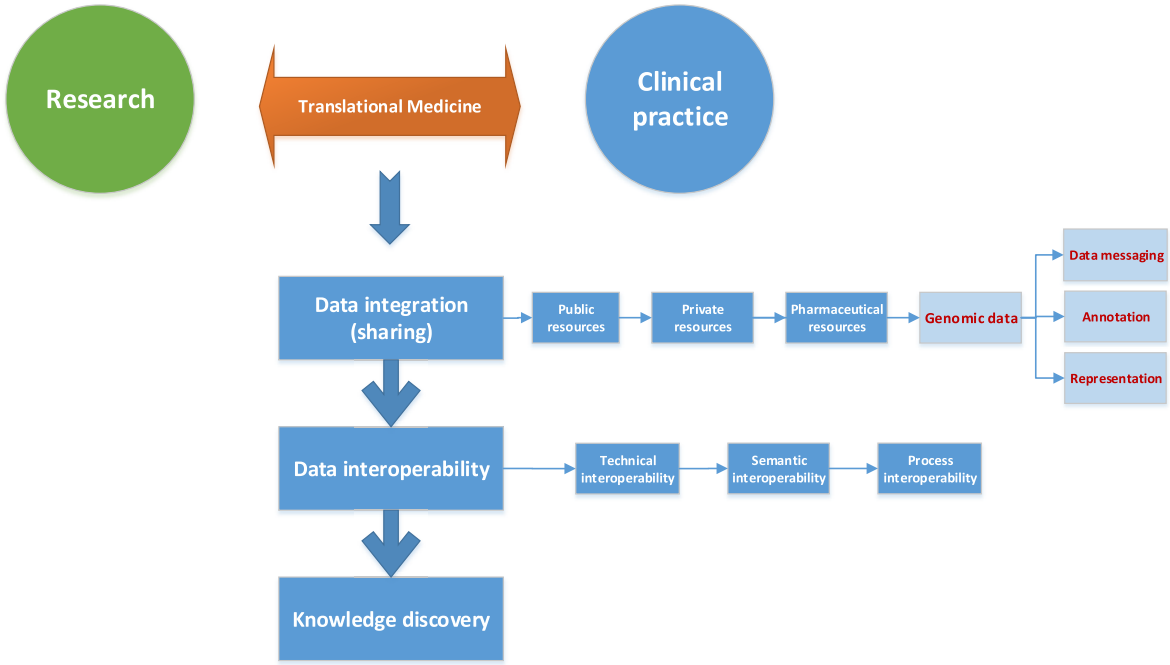

FIGURE 4. Translational medicine bridge.

research design approaches is given in Table 1. Most of the used methods are from quantitative perspective, which is as expected since the data is already well structured in the EHRs.

B. QUANTITATIVE RESEARCH DESIGN

The phenomena of interest in quantitative research design can be explained by the large volume of numerical data which can be easily and timely analysed by mathematical methods [46] including statistics and machine learning techniques. Quantitative research has empirical nature (relies on the experiments) and stays independent from the researchers. It makes the reality measurable in an objective manner. The idea is to discover or validate existing relationships and then to derive generalisations that will contribute to the theory of interest [60]. It involves the application of various machine learning methods for data analysis. Some of the most commonly used in human behaviour analysis research area is Self-Organised Maps [61], K-means [62], multivariate-time series clustering [63], bottom-up and top-down hierarchical and non-hierarchical clustering algorithms (suitable for open-ended questionnaire [64]).

Quantitative research can be broadly classified as either descriptive experimental approaches or causal comparative approaches. The descriptive research is based on the observations and investigations on the correlation between two or more circumstances. The researchers can explore and apply pre-experiments (variables remain the same and the control group is not randomly selected), true experiments (mathematical models are used in the analyses), or quasi-experiments (the study group is not randomly selected and the results may not be comparable) [60]. Support Vector Regression, Partial Least Squares, Random Forest, K-nearest neighbours, Petri nets and Markov models are some of the preferred analysis techniques after the researchers apply designed treatments to the target groups and measure the results/outcomes.

The differences between causal comparative design and descriptive experimental design have been presented in [65]. In the causal comparative research, the emphasis has been put

on how those dependent variables can affect the independent variables and what are the cause and also the relationships between those variables [60]. For example, the results can be used to inspect the influences on the recent information and communication technologies [66]. Analysis techniques such as Hierarchical clustering, K-means, mapping, ordinary least squares regression and spatial autocorrelation are commonly used methods [67], [68]. The factorial design focuses on the comparison on the categories of independent variables with the categories of dependent variables [69]. Support Vector Machines are generally used to identify the patterns that can be used to describe individuals' personality. Therefore, if someone wants to inspect the interaction between the independent variables and their influence on the depended variables, then he/she should use causal comparative research design [60]. It can well reveal the causations of the differences in the outcomes from the experimental groups or the individuals. The analysis of causal comparative design cannot certainly determine the research outcome and it highly relies on the probabilistic methods. Data analysis often includes the use of statistical tests for comparisons, analysis of covariance, measures of central tendency, descriptive and inferential statistics [70]–[72].

C. QUALITATIVE RESEARCH DESIGN

Qualitative research is normally considered harder to define and conduct. Qualitative research is comprised of several simplistic definitions. It is a naturalistic way to study the phenomena in order to discover descriptive terms rather than variables [46]. One of the key characteristics of a qualitative research is that the social phenomena is explored from participants' point of view. The use of different techniques can result in significant changes on the qualitative research strategies. Since the qualitative research is used to discover new theories and findings, the initial questions and the conclusions are normally in an unstructured format [60].

There are variety of strategies to conduct a qualitative research such as: case study, ethnography study, phenomenological study, grounded theory study, and content analysis [46].

Case studies and the grounded theory research are used to investigate experimental processes, activities, and events. They are helpful to study the characteristics of individuals since they have a well defined time and location frame. For case studies, researchers normally attempt to learn from sample experiments and results. The data arrives from multiple sources in the format of interviews, observations, documents, and audiovisual materials [60].

Ethnographic research is used to analyse the cultural-sharing behaviours of whole groups rather than just individuals. It can take a long period of time, since the researchers must observe the group's daily life constantly in order to identify the research structures and needed factors. The procedure is to gain access to the group and build trust, and thereafter to start the communication. These research tasks include: the justification for the study, the description of the group and method used, the collection of the evidence, and conclude the research findings [60].

Grounded theory can generalise theories based on the understandings of the participants in the study. The theory must be revealed and discovered from the health data rather than just from the literature. The process is expected to be repetitive, including collecting and analysing the data in cycle. The data analysis process includes open coding, axial coding, selective coding, and then developing a theory. The key approach when conducting this type of research is to define the research question, review the literature, determine the methodology, discuss the implications and draw conclusions [60].

Phenomenological study is introduced to to understand an experience from the participants point of view. The drawback of this type of studies is that the researchers' involvement might be bias [48]. The method of collecting the data is similar as that in grounded theory, by performing interviews. A procedural format is suggested when initialising the research questions, performing the interviews, analysing the data and writing scientific reports. Those activities should lead to the identification of common themes in participants' perceptions [60].

Content analysis study is to find the detailed and systematic examination of the contents from a particular body of materials for the purpose of identifying patterns (verbal, visual, behavioural). It is supposed to be highly objective. The data analysis consists of putting the materials in a frequency table and performing statistical analysis to derive results in a quantitative form. The report is constructed to include the description of the studied materials, a description of the used methodology, the statistical analysis reflecting the frequency table, and the conclusions on the data patterns, themes, or biases derived from the research [60].

When choosing an appropriate method for qualitative analysis, the researchers must consider the type of the data being analysed: whether the data is numerical (continuous or discrete) or categorical; the dimensionality of the data pool and the dimensionality of the data set that will actually be evaluated. Some of the currently most popular algorithms used are clustering-based, such as Algorithm Clope, Algorithm Large Item, Kohonen maps and Hierarchical clustering [73]. Qualitative studies often require a large body of health data. Since most of the time the sample size is not sufficient enough for obtaining trustworthy cluster analysis and conducting reliable conclusions, Hierarchical clustering, K-means clustering, and Latent Class Analysis techniques then are applied to analyse binary data produced form coded qualitative interviews.

D. TRANSLATION OF QUALITATIVE DATA INTO QUANTITATIVE DATA

As a response to the lack of qualitative analyses, there is a possibility to translate the qualitative data into quantitative. There are different methods for translating qualitative into quantitative data [74]. Bayesian statistical reasoning is a statistical method where translation is made by specifying the prior

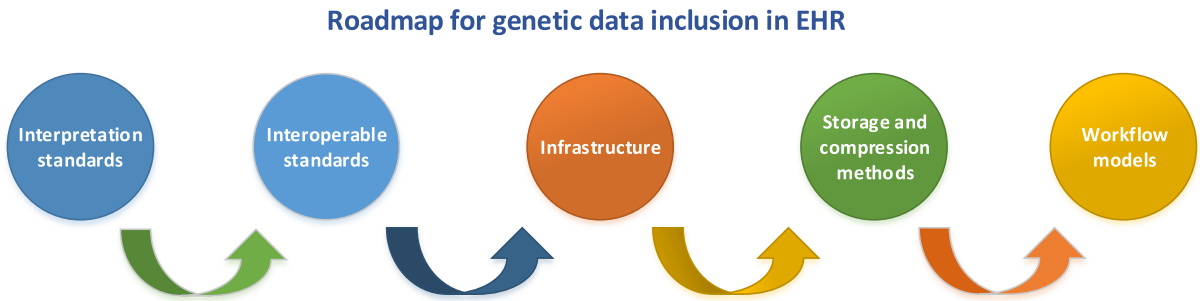

FIGURE 5. Genomic data inclusion in EHR.

distributions for each parameter in the reference case (ex. historical data). The evaluation of the conditional probability is based on the experts' knowledge and then the posterior distribution is evaluated [75]. In a case of Bayesian networks, the translation is made when graphical influence diagrams are populated with a set of conditional probability tables which define the quantitative relationships between each variable [74]. The pairwise comparison method allows qualitative inputs from experts to be translated onto a numerical scale. The process is composed of four steps: pairwise comparison for each parameter across all scenarios, translation of subjective measures into numerical values, evaluation of the translation and consistent derivation of eigenvectors [76]. The fuzzy set translation method is used for translating narratives to quantitative inputs to a simulation model. The translation process involves five steps: making statements about rates of change of all important drivers; deriving a translation key used to translate statements into numbers; combining the result set of numbers into a membership function; using the membership functions by linking with probability density functions or demystifying the output to obtain inputs for simulation models; and using the simulation models outputs to enrich storylines [74], [77]. Agent-based modelling method (ABM) represents a discrete-event simulation composed of heterogeneous and autonomous decision making entities to interact with each other, which is called "agents". To construct an ABM, a qualitative content analysis step that includes the identification of agents is required. Then, the translation step is done by construction of a causal diagram that depicts the major agent interactions and system processes. There also is a possibility to make the translation by combining one or more translation methods to improve the performance (e.g. pairwise comparison, and/or questionnaire surveys and interviews) [74], [78].

E. FUTURE RESEARCH DIRECTION

In this paper, we have presented a large body of studies related to health data analysis and medical system standardisation. It is still challenging that how can we guarantee that the CH paradigm is appropriately introduced and the proposed mechanism is suitable for evaluating the efficiency and effectiveness of CH. In order to consider the impact from not only technological but also human aspects, we need to establish an evaluation framework. The high level of diversity between the two aspects leads to an approach to include of both quantitative and qualitative metrics. Therefore, we indicate that a hybrid research design should draw more attentions in the future work.

Quantitative research is objectively to measure the phenomena through statistic approaches. It abstracts variables from data and reveals relationship between them. Qualitative research is much more flexible, allowing sensitive and descriptive context when creating the research framework. There are four elements that are considered when both research approaches are discussed [46]:

1) the study design paradigm – define what we try to measure; whether we want to compare groups or study individuals; whether we want to investigate new theories or claim the old ones;
2) the target population of the study – who is to be assessed, the infrastructure or the people involved;
3) the research strategies – choose one of the quantitative or qualitative research design, *and*
4) the methods that will be used for data analysis.

Following those steps, the researchers will be able to apply proper strategies for different purposes in the CH concept. For example, if one wants to evaluate the technical aspects of a system or the level of data interoperability, it can be done by selecting certain quantitative metrics and performing statistical analysis. If the users need to develop new IoMT devices or standards, then qualitative research design is preferred, as indicated in Section VI-C.

Considering the current data analysis techniques and following our understanding, we would conclude that data clustering is expected to be more suitable for qualitative research, where no initial categories are made and no predefined outcomes are expected. However, the reported papers in the literature [60], [67], [68] have presented various techniques (mostly clustering and regression) that can be applied to both descriptive and causal-comparative quantitative approaches. When the quantitative research question is described in nature languages, the researchers are advised to select measures from descriptive statistics. If the quantitative research question is causal-comparative, then an analysis tool that allows

direct comparisons between groups is needed. All of the mentioned procedures are reliable to be used, including: t-test, ANOVA, analysis of covariance (ANCOVA), MANOVA and multiple analysis of covariance (MANCOVA), all with different criteria of variable dependence/independence [79]. Clustering techniques are found to be suitable for a large amount of data with high dimensions, where there is a need for investigating the existing relationship between different subsets. If the aim of the research is to predict the influence of some factors in a given problem, then usually the regression techniques are applied.

The data predictive analytics methods can assist to process the patients' data, analyse the clinical notes, find the underlying correlations among symptoms, summarise habits and discover the risk for certain diseases. All the analysis results can help to predict the patients' future health conditions and make proper suggestions for treatments. By combing the data from multiple sources, the impacts of certain biomedical factors such as genome structure or clinical variables can be also taken into the account in order to evaluate the medical care. For example, data intensive algorithms can be used to detect the slightest changes in the patient's health indicators and predict possible disorders or alarm physicians to further check the patient condition through analysing the heart rate and breathing patterns Another research example is chronic disease management. By combining various sources of data (e.g. EHR and genome sequencing), physicians can identify consistent patterns in symptoms and create accurate patient profiles, which can greatly improve the possibility for personalised and effective treatment.

The process of intelligently combining the data science and medical practice workflow is the future research direction. It is important to offer a certain degree of transparency to the clinicians and physicians about the machine learning models and prediction results. A suggestion for diagnosis to a physician made by an algorithm without any justification will be impossible to put into practice due to trustworthy issues. Then the physicians will be forced to do a full chart review and physical examination themselves. Information sharing and cross-domain collaboration are the keys for future research in order to meet the clinics' expectations and the doctors' needs.

VII. CONCLUSION

TM significantly increases the possibilities for discovering of new diagnostic tools and treatments by utilising multi-disciplinary expertise and highly collaborative approaches. The importance of the CH paradigm is already proved by the necessity of continuous collaborations between the clinicians and the data researchers. The development of EHRs has allowed a complete overview and a convenient access over the patients' medical history. It also can bridge and improve the cross-domain communication between the clinicians and the data scientists. We conclude that a medical system based on well-defined EHRs can accelerate the diagnostic procedure and therefore improve the accuracy of the treatments by obtaining better analysis results and data understanding. Following the introduction of EHRs, we have delivered an in-depth review of the data analysis approaches applied in TM from two angles: quantitative and qualitative approaches. It has shown that quantitative analysis is not sufficient to deliver accurate and useful conclusions for clinicians. Qualitative analysis should be involved and conducted in many cases. The lack of data oriented qualitative research is the urgent problem that should be addressed in order to meet the doctors' needs. The high degree of the diversities between the two types of approaches inspired the introduction of solutions with hybrid (combining quantitative and qualitative metrics) research design. We indicate that by utilising data analysis techniques on patient's data, meaningful information, such as human behaviours, can be extracted to support specific health risk identifications and health treatment improvements in a wider scope. Extracting meaningful information from the collected data and creating new healthcare models require careful design and analysis. The paper has also concluded the elements and the methods that should be taken into consideration for health problems discovery and addressing.

REFERENCES

[1] B. M. Caulfield and S. C. Donnelly, "What is Connected Health and why will it change your practice?" *QJM, Int. J. Med.*, vol. 106, no. 8, pp. 703–707, 2013.

[2] Accenture. (2012). *Connected Health: The Drive to Integrated Healthcare Delivery*. [Online]. Available: http://www.accenture.com/us-en/Pages/service-connected-health-study.aspx

[3] J. Kvedar, M. J. Coye, and W. Everett, "Connected health: A review of technologies and strategies to improve patient care with telemedicine and telehealth," *Health Affairs*, vol. 33, no. 2, pp. 194–199, 2014. [Online]. Available: https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2013.0992

[4] D. Jiang, C. Tang, and A. Zhang, "Cluster analysis for gene expression data: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1370–1386, Nov. 2004.

[5] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, Mar. 2016.

[6] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, "The quantitative analysis of mammographic densities," *Phys. Med. Biol.*, vol. 39, no. 10, p. 1629, 1994. [Online]. Available: http://stacks.iop.org/0031-9155/39/i=10/a=008

[7] M. A. Steinman, S. R. Ranji, K. G. Shojania, and R. Gonzales, "Improving antibiotic selection: A systematic review and quantitative analysis of quality improvement strategies," *Med. Care*, vol. 44, no. 7, pp. 617–628, 2006. [Online]. Available: http://www.jstor.org/stable/40221329

[8] M. Ryan *et al.*, "Eliciting public preferences for healthcare: A systematic review of techniques," *Health Technol. Assessment*, vol. 5, no. 5, pp. 1–186, 2001.

[9] V. P. da Silva, B. R. R. Oliveira, R. G. T. Mello, H. Moraes, A. C. Deslandes, and J. Laks, "Heart rate variability indexes in dementia: A systematic review with a quantitative analysis," *Current Alzheimer Res.*, vol. 15, no. 1, pp. 80–88, 05 2017.

[10] L. Leung, "Validity, reliability, and generalizability in qualitative research," *J. Family Med. Primary Care*, vol. 4, no. 3, pp. 324–327, 2015.

[11] S. Elo, M. Kääriäinen, O. Kanste, T. Pölkki, K. Utriainen, and H. Kyngäs, "Qualitative content analysis: A focus on trustworthiness," *SAGE Open*, vol. 4, no. 1, p. 2158244014522633, 2014, doi: 10.1177/2158244014522633.

[12] J. M. Morse, "Critical analysis of strategies for determining rigor in qualitative inquiry," *Qualitative Health Res.*, vol. 25, no. 9, pp. 1212–1222, 2015, doi: 10.1177/1049732315588501.

[13] A. J. Wright, S. Sutton, D. Armstrong, P. Aveyard, A. L. Kinmonth, and T. M. Marteau, "Factors influencing the impact of pharmacogenomic prescribing on adherence to nicotine replacement therapy: A qualitative study of participants from a randomized controlled trial," *Transl. Behav. Med.*, vol. 8, no. 1, pp. 18–28, 2018, doi: 10.1093/tbm/ibx008.

[14] J. Smith and J. Firth, "Qualitative data analysis: The framework approach," *Nurse Res.*, vol. 18, no. 2, pp. 52–62, Jan. 2011. [Online]. Available: http://eprints.hud.ac.uk/id/eprint/18884/

[15] J. C. Pickup, M. F. Holloway, and K. Samsi, "Real-time continuous glucose monitoring in type 1 diabetes: A qualitative framework analysis of patient narratives," *Diabetes Care*, vol. 38, no. 4, pp. 544–550, 2015. [Online]. Available: http://care.diabetesjournals.org/content/38/4/544

[16] S. Lewis, "Qualitative inquiry and research design: Choosing among five approaches," *Health Promotion Pract.*, vol. 16, no. 4, pp. 473–475, 2015, doi: 10.1177/1524839915580941.

[17] M. Woods, T. Paulus, D. P. Atkins, and R. Macklin, "Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS.ti and NVivo, 1994–2013," *Social Sci. Comput. Rev.*, vol. 34, no. 5, pp. 597–617, 2016, doi: 10.1177/0894439315596311.

[18] K. E. A. Mikkonen, S. Elo, H.-M. Kuivila, A.-M. Tuomikoski, and M. Kääriäinen, "Culturally and linguistically diverse healthcare students' experiences of learning in a clinical environment: A systematic review of qualitative studies," *Int. J. Nursing Stud.*, vol. 54, pp. 173–187, 2015.

[19] T. Damianakis and M. R. Woodford, "Qualitative research with small connected communities: Generating new knowledge while upholding research ethics," *Qualitative Health Res.*, vol. 22, no. 5, pp. 708–718, 2012, doi: 10.1177/1049732311431444.

[20] C. Pope, S. Ziebland, and N. Mays, "Analysing qualitative data," *BMJ, Brit. Med. J.*, vol. 320, no. 7227, pp. 114–116, 2000.

[21] R. Khatib *et al.*, "Patient and healthcare provider barriers to hypertension awareness, treatment and follow up: A systematic review and meta-analysis of qualitative and quantitative studies," *PLOS ONE*, vol. 9, no. 1, pp. 1–12, Jan. 2014, doi: 10.1371/journal.pone.0084238.

[22] H. M. Krumholz, "Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system," *Health Affairs*, vol. 33, no. 7, pp. 1163–1170, 2014.

[23] IBM's Watson. (2016). *Watson for Oncology*. [Online]. Available: http://www.ibm.com/smarterplanet/us/en/ibmwatson/watson-oncology.html

[24] Healthcare IT Connect (HITC). (2012). *Big Data Infographic*. [Online]. Available: http://www.healthcareitconnect.com/

[25] F. Pennic. (2012). *Understanding How Big Data Flows in Healthcare Infographic*. [Online]. Available: http://hitconsultant.net/2012/12/12/understanding-how-big-data-flows-in-healthcare-infographic

[26] HIT Consultant. (2016). *10 Healthcare Analytics Trends to Watch (Infographic)*. [Online]. Available: http://hitconsultant.net/2016/03/14/10-healthcare-analytics-trends-to-watch

[27] R. Margolis *et al.*, "The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: Capitalizing on biomedical big data," *J. Amer. Med. Informat. Assoc.*, vol. 21, no. 6, pp. 957–958, 2014.

[28] S. Bowman, "Impact of electronic health record systems on information integrity: Quality and safety implications," *Perspect. Health Inf. Manage.*, vol. 10, no. 1, Oct. 2013.

[29] S. R. Deeny and A. Steventon, "Making sense of the shadows: Priorities for creating a learning healthcare system based on routinely collected data," *BMJ Quality Safety*, vol. 24, no. 8, pp. 505–515, 2015.

[30] Ministry of Transport and Communication. (2015). *MyData—A Nordic Model for Human-Centered Personal Data Management and Processing*. [Online]. Available: http://www.lvm.fi/documents/20181/859937/MyData-nordic-model/2e9b4eb0-68d7-463b-9460-821493449a63?version=1.0

[31] K. E. Johnson, A. Kamineni, S. Fuller, D. Olmstead, and K. J. Wernli, "How the provenance of electronic health record data matters for research: A case example using system mapping," *Proc. EGEMS*, vol. 2, no. 1, p. 1058, 2014.

[32] E. Garvin. (2016). *Physician Perspective: Solving EHR Integration & Interoperability*. [Online]. Available: http://hitconsultant.net/2016/04/26/solving-ehr-integration-interoperability

[33] D. Fridsma. (2013). *EHR Interoperability: The Structured Data Capture Initiative*. [Online]. Available: https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/ehr-interoperability-structured-data-capture-initiative

[34] (Dec. 13, 2018). [Online]. Available: https://wiki.ihe.net/

[35] Office of the National Coordinator for Health IT. (2015). *Connecting Health and Care for the Nation—A Shared Nationwide Interoperability Roadmap*. [Online]. Available: https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf

[36] D. Aisner, L. Olsen, J. M. McGinnis, *The Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine)*. Washington, DC, USA: National Academies Press, 2007.

[37] B. C. Delaney, K. A. Peterson, S. Speedie, A. Taweel, T. N. Arvanitis, and F. R. Hobbs, "Envisioning a learning health care system: The electronic primary care research network, a case study," *Ann. Family Med.*, vol. 10, no. 1, pp. 54–59, 2012.

[38] P. J. Haug, B. H. S. C. Rocha, and R. S. Evans, "Decision support in medicine: Lessons from the help system," *Int. J. Med. Informat.*, vol. 69, nos. 2–3, pp. 273–284, 2003.

[39] C. P. Friedman, A. K. Wong, and D. Blumenthal, "Achieving a nationwide learning health system," *Sci. Transl. Med.*, vol. 2, no. 57, p. 57cm29, 2010.

[40] P. P. Yu, M. A. Hoffman, and D. F. Hayes, "Biomarkers and oncology: The path forward to a learning health system," *Archives Pathol. Lab. Med.*, vol. 139, no. 4, pp. 451–456, 2014.

[41] T. R. Ross *et al.*, "The HMO research network virtual data warehouse: A public data model to support collaboration," *EGEMS*, vol. 2, no. 1, p. 1049, 2014.

[42] HL7. (2016). *Health Level Seven International*. [Online]. Available: http://www.hl7.org

[43] R. Cornet and N. de Keizer, "Forty years of SNOMED: A literature review," *BMC Med. Informat. Decision Making*, vol. 8, no. 1, p. S2, 2008.

[44] P. L. Elkin *et al.*, "Evaluation of the content coverage of SNOMED CT: Ability of SNOMED clinical terms to represent clinical problem lists," *Mayo Clin. Proc.*, vol. 81, no. 6, pp. 741–748, 2006.

[45] K. D. Starič, P. Bukovec, K. Jakopič, E. Zdravevski, V. Trajkovik, and A. Lukanović, "Can we predict obstetric anal sphincter injury?" *Eur. J. Obstetrics Gynecol. Reproductiv. Biol.*, vol. 210, pp. 196–200, Mar. 2017.

[46] K. Yilmaz, "Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences," *Eur.J. Edu.*, vol. 48, no. 2, pp. 311–325, 2013.

[47] J. W. Creswell, *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. Newbury Park, CA, USA: Sage, 2013.

[48] R. Berger, "Now I see it, now I don't: Researcher's position and reflexivity in qualitative research," *Qualitative Res.*, vol. 15, no. 2, pp. 219–234, 2015.

[49] A. L. Nobles, K. Vilankar, H. Wu, and L. E. Barnes, "Evaluation of data quality of multisite electronic health record data for secondary analysis," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct./Nov. 2015, pp. 2612–2620.

[50] A. R. Tate *et al.*, "Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface," *J. Amer. Med. Inform. Assoc.*, vol. 21, no. 2, pp. 292–298, 2013.

[51] T. Botsis, G. Hartvigsen, F. Chen, and C. Weng, "Secondary use of EHR: Data quality issues and informatics opportunities," *Summit Transl. Bioinform.*, vol. 2010, pp. 1–5, Mar. 2010.

[52] S. Dungey, N. Beloff, S. Puri, R. Boggon, T. Williams, and A. R. Tate, "A pragmatic approach for measuring data quality in primary care databases," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Jun. 2014, pp. 797–800.

[53] M. J. Romano and R. S. Stafford, "Electronic health records and clinical decision support systems: Impact on national ambulatory care quality," *Archives Internal Med.*, vol. 171, no. 10, pp. 897–903, 2011.

[54] D. G. T. Arts, N. F. de Keizer, and G.-J. Scheffer, "Defining and improving data quality in medical registries: A literature review, case study, and generic framework," *J. Amer. Med. Inform. Assoc.*, vol. 9, no. 6, pp. 600–611, 2002.

[55] J. Sun, J. Hu, D. Luo, M. Markatou, F. Wang, and S. Edabollahi, "Combining knowledge and data driven insights for identifying risk factors using electronic health records," in *Proc. AMIA Annu. Symp.* Bethesda, MD, USA: American Medical Informatics Association, 2012, p. 901.

[56] L. A. Knake *et al.*, "Quality of EHR data extractions for studies of preterm birth in a tertiary care center: Guidelines for obtaining reliable data," *BMC Pediatrics*, vol. 16, no. 1, p. 59, 2016.

[57] C. D. Newgard, D. Zive, J. Jui, C. Weathers, and M. Daya, "Electronic versus manual data processing: Evaluating the use of electronic health records in out-of-hospital clinical research," *Acad. Emergency Med.*, vol. 19, no. 2, pp. 217–227, 2012, doi: 10.1111/j.1553-2712.2011.01275.x.

[58] C. R. Weir, N. Staggers, B. Gibson, K. Doing-Harris, R. Barrus, and R. Dunlea, "A qualitative evaluation of the crucial attributes of contextual Information necessary in EHR design to support patient-centered medical home care," *BMC Med. Informat. Decision Making*, vol. 15, no. 1, p. 30, 2015.

[59] G. Ser, A. Robertson, and A. Sheikh, "A qualitative exploration of workarounds related to the implementation of national electronic health records in early adopter mental health hospitals," *PLoS ONE*, vol. 9, no. 1, p. e77669, 2014.

[60] C. Williams, "Research methods," *J. Bus. Econ. Res.*, vol. 5, no. 3, pp. 65–72, 2011.

[61] C. Laspidou, E. Papageorgiou, K. Kokkinos, S. Sahu, A. Gupta, and L. Tassiulas, "Exploring patterns in water consumption by clustering," *Procedia Eng.*, vol. 119, pp. 1439–1446, 2015.

[62] A. Seret, S. K. L. M. vanden Broucke, B. Baesens, and J. Vanthienen, "A dynamic understanding of customer behavior processes based on clustering and sequence mining," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4648–4657, 2014.

[63] V. Rodríguez-Fernández, H. D. Menéndez, and D. Camacho, "Analysing temporal performance profiles of UAV operators using time series clustering," *Expert Syst. Appl.*, vol. 70, pp. 103–118, Mar. 2017.

[64] O. R. Battaglia, B. Di Paola, and C. Fazio. (2015). "Cluster analysis of educational data: An example of quantitative study on the answers to an open-ended questionnaire." [Online]. Available: https://arxiv.org/abs/1512.08998

[65] B. Johnson and L. B. Christensen, *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. Newbury Park, CA, USA: Sage, 2008.

[66] J. B. Pick, A. Sarkar, and J. Johnson, "United States digital divide: State level analysis of spatial clustering and multivariate determinants of ICT utilization," *Socio-Econ. Planning Sci.*, vol. 49, pp. 16–32, Mar. 2015.

[67] G. Paneiro, F. Durão, M. C. E. Silva, and P. F. Neves, "Prediction of ground vibration amplitudes due to urban railway traffic using quantitative and qualitative field data," *Transp. Res. D, Transp. Environ.*, vol. 40, pp. 1–13, Oct. 2015.

[68] C. Kinsella, C. McTague, and K. N. Raleigh, "Unmasking geographic polarization and clustering: A micro-scalar analysis of partisan voting behavior," *Appl. Geogr.*, vol. 62, pp. 404–419, Aug. 2015.

[69] Y.-H. Yang and J.-Y. Liu, "Quantitative study of music listening behavior in a social and affective context," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1304–1315, Oct. 2013.

[70] A. M. Penner and M. Paret, "Gender differences in mathematics achievement: Exploring the early grades and the extremes," *Social Sci. Res.*, vol. 37, no. 1, pp. 239–253, Mar. 2008.

[71] J. Lee, "Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies," *Rev. Edu. Res.*, vol. 78, no. 3, pp. 608–644, 2008.

[72] J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study," *BMJ*, vol. 337, Dec. 2008, Art. no. a2338.

[73] S. Cherevko and A. Malikov, "Review of modern techniques of qualitative data clustering," in *Proc. Young Sci. Int. Workshop Trends Inf. Process. (YSIP)*, 2014, p. 7.

[74] V. R. Mallampalli *et al.*, "Methods for translating narrative scenarios into quantitative assessments of land use change," *Environ. Model. Softw.*, vol. 82, pp. 7–20, Aug. 2016.

[75] E. Kemp-Benedict, "Converting qualitative assessments to quantitative assumptions: Bayes' rule and the pundit's wager," *Technol. Forecasting Social Change*, vol. 77, no. 1, pp. 167–171, 2010.

[76] J. Abildtrup *et al.*, "Socio-economic scenario development for the assessment of climate change impacts on agricultural land use: A pairwise comparison approach," *Environ. Sci. Policy*, vol. 9, no. 2, pp. 101–115, 2006.

[77] J. Alcamo, "Chapter six the SAS approach: Combining qualitative and quantitative knowledge in environmental scenarios," *Develop. Integr. Environ. Assessment*, vol. 2, pp. 123–150, 2008.

[78] A. Ligmann-Zielinska and P. Jankowski, "Agent-based models as laboratories for spatially explicit planning policies," *Environ. Planning B, Planning Des.*, vol. 34, no. 2, pp. 316–335, 2007.

[79] A. J. Onwuegbuzie and N. L. Leech, "Linking research questions to mixed methods data analysis procedures 1," *Qualitative Rep.*, vol. 11, no. 3, pp. 474–498, 2006.

LINA XU received the B.E. degree from the Software Engineering School, Fudan University, China, and the B.Sc. degree in computer science from the University College Dublin (UCD), Ireland, through a Fudan and UCD joint Program, and the Ph.D. degree in computer science in UCD, in 2014. She was with HP Labs as a Research Scientist, from 2014 to 2016. She is currently an Assistant Professor with UCD. Her research interests include, Internet of Things (IoT), 5G network, smart networking, and machine learning. She is currently working on a smart living project aiming to apply IoT technologies to smart living.

MONIKA SIMJANOSKA received the bachelor's in computer science from the Faculty of Natural Science and Mathematics, Saints Cyril and Methodius University of Skopje, in 2012, and master's degrees in computer science from the Faculty of Computer Science and Engineering, Saints Cyril and Methodius University of Skopje, in 2013, where she is currently pursuing the Ph.D. degree. Her Ph.D. thesis is in the field of biomedical signal processing. She is currently a Teaching Assistant with the Faculty of Computer Science and Engineering. She is an author or co-author of more than 30 conference and journal papers in the field of bioinformatics, biomedical signal processing, and cloud computing. She has been involved in many national and international projects, including NATO Science for Peace, as a young researcher.

BOJANA KOTESKA received the bachelor's and master's degrees in informatics and the Ph.D. degree in computer science from the Faculty of Computer Sciences and Engineering, Saints Cyril and Methodius University of Skopje, in 2011, 2012, and 2018, respectively. As the best student of her generation, she received awards from the faculty and other organizations, including the Organization of Engineers of Macedonia that awarded her the Golden Engineering Ring. In 2010, she started working at the Faculty of Natural Science and Mathematics and later at the Faculty of Computer Science and Engineering as a Demonstrator. She is also working as a Teaching and Research Assistant with the Faculty of Computer Science and Engineering at the moment. She has been active an participant in 12 national and international projects in the field of computer science. She has participated in over 25 scientific conferences and has published 40 scientific papers, of which four papers are in international journals with impact factor.

VLADIMIR TRAJKOVIK received the Ph.D. degree, in 2003. He is currently a Professor with the Faculty of Computer Science and Engineering, Saints Cyril and Methodius University of Skopje, Macedonia. He has authored or co-authored over 150 refereed journal and conference papers, 12 book chapters, and four books. His current research interests are in the field of distributed and mobile systems, connected health, recommender systems, and various applications of ICT in the assisted living and education. He has been involved in many international and national scientific and applicative projects, including HORIZON 2020, FP7, EU-TEMPUS, EU-COST, ERASMUS, and NATO Science for Peace, as a Project Coordinator and a Participant. He is a member of several professional societies. He has done management consultancy in five ICT companies. He has been a Mentor of five ICT companies in the last 15 years. He is a Co-Founder and a Partner in four ICT companies.

ANA MADEVSKA BOGDANOVA received the Ph.D. degree, in 2003. She is currently a Full Professor with the Faculty of Computer Science and Engineering, Saints Cyril and Methodius University of Skopje, Macedonia. She has authored or co-authored over 100 refereed journal and conference papers, eight book chapters and books (Editor and author). Her current research interests are within the domain of data science, intelligent systems, connected health, biosensors, telemedicine systems, and ICT in education. She has been involved in international and national projects, including NATO SPS and ERASMUS+, as a Project Coordinator, EUCOST, HORIZON 2020, EU-TEMPUS and DAAD, as a Participant. She is a member of several scientific and professional, international and national societies.

KRISTINA DRUSANY STARIČ graduated from the Faculty of Medicine, University of Ljubljana, in 2002. In 2011, she passed a board exam. She received the Ph.D. degree, in 2017. Her doctoral thesis was on the effects of the first birth and episiotomy on anal sphincter. She was with the University Clinical Center, Ljubljana, where she has been with the Division for Obstetrics and Gynecology, since 2005, where she also specialized in gynecology and obstetrics. Her work consists mainly of surgical reconstruction of the pelvic floor and malignant diseases of the female genital tract. She actively participates and organizes Slovenian and international gynecological professional meetings, where she also received several awards for the best contribution and participates in domestic and international research projects, such as TASI2. She has written numerous scientific, professional, and popular publications and has been a member in domestic and international associations of gynecologists and obstetricians, such as the International Urogynaecological Society, the Mediterian Pelvic Floor Society, and ENJECT.

FEDOR LEHOCKI has been a member of the Institute of Robotics and Cybernetics, Slovak University of Technology, since 2014. He is currently the Director of the National Centre of Telemedicine Services. He co-authored the strategic documents for the Government of Slovak Republic related to national eHealth program. His current research interests include mHealth, telemedicine technology, systems and services, clinical decision support systems, knowledge representation, and remote monitoring and treatment. He is a member of the IEEE EMBS and served as a member of the Medical Terminology and Health Informatics Standards Center, advisory body for the Ministry of Health of the Slovak Republic. He served as an Associate Editor for the IEEE Journal of Biomedical and Health Informatics and was a reviewer for *Biomedical Engineering* Book Series published by the Wiley and IEEE Press.