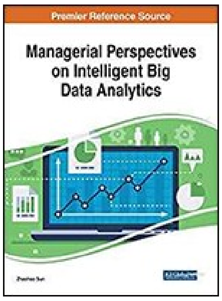

Managerial Perspectives on Intelligent Big Data Analytics

by Zhaohao Sun

Chapter 9: Remote Patient Monitoring for Healthcare: A Big Challenge for Big Data

Andrew Stranieri,
Federation University, Australia

Venki Balasubramanian,
Federation University, Australia

ABSTRACT

Remote patient monitoring involves the collection of data from wearable sensors that typically requires analysis in real time. The real-time analysis of data streaming continuously to a server challenges data mining algorithms that have mostly been developed for static data residing in central repositories. Remote patient monitoring also generates huge data sets that present storage and management problems. Although virtual records of every health event throughout an individual's lifespan known as the electronic health record are rapidly emerging, few electronic records accommodate data from continuous remote patient monitoring. These factors combine to make data analytics with continuous patient data very challenging. In this chapter, benefits for data analytics inherent in the use of standards for clinical concepts for remote patient monitoring is presented. The openEHR standard that describes the way in which concepts are used in clinical practice is well suited to be adopted as the standard required to record meta-data about remote monitoring. The claim is advanced that this is likely to facilitate meaningful real time analyses with big remote patient monitoring data. The point is made by drawing on a case study involving the transmission of patient vital sign data collected from wearable sensors in an Indian hospital.

INTRODUCTION

Continuous remote monitoring of patients using wearable sensors and Cloud processing is emerging as a technology that promises to lead to new ways to realize early detection of conditions and increased efficiency and safety in health care systems (Chan, Estève, Fourniols, Escriba, & Campo, 2012). The approach combines body area wireless sensor networks (BSN) with systems that are designed to process and store the data for the purpose of raising alarms immediately or for data analytics exercises at a later point in time (Balasubramanian, Stranieri, & Kaur, 2015). Real time remote monitoring systems have been described for a number of remote monitoring applications including: continuous vital signs monitoring (Balasubramanian & Stranieri, 2014; Catley, Smith, McGregor, & Tracy, 2009), arrhythmia detection (Kakria, Tripathi, & Kitipawang, 2015), regulating oxygen therapy (Zhu et al., 2005), monitoring of pregnant women (Balasubramanian, Hoang, & Ahmad, 2008), fall detection (Thilo et al., 2016), chemotherapy reaction (Breen et al., 2017) and glucose monitoring (Klonoff, Ahn, & Drincic, 2017). Ultimately, a multitude of condition specific applications, each using different subsets of each patient's health data commissioned by diverse healthcare practices can be expected to emerge in the near future. For instance, a rehabilitation clinic may be interested in tracking a patient's gait, while a counselling service may be interested in tracking heart rate variability to detect suicidal depression (Carta & Angst, 2016) and a hospital may be interested in detecting post-operative sepsis (Brown et al., 2016).

Remote patient monitoring (RPM) applications often generate high volumes of data with great velocity and variety to produce valuable diagnostic information. For instance an ECG wearable sensor alone can produce 125 to 8000 samples per second (Shimmer, 2018), that can be used to predict various heart conditions in real time. In many occasions, a RPM application uses more than one wearable sensor to monitor vital signs, such as ECG, body temperature, blood pressure, oxygen saturation (SpO_2) and respiratory rate, to analyze and predict the health condition of the patient. This leads to large data repositories that present serious challenges for Big Data analytics algorithms (Kalid et al., 2018). A review by (Mikalef, Pappas, Krogstie, & Giannakos, 2017) reveals that Big Data is characterized in terms of the five main 'Vs:' volume, velocity, variety, veracity and value. Although a great deal has been written about the Big Data explosion, little is known of the conditions under which Big Data Analysis (BDA) leads to the generation of value for an organization (Wang, Kung, & Byrd, 2018).

In this chapter, the observation is first made that BDA for remote patient monitoring is difficult to perform due to the volume, velocity, veracity and diversity of data. Consequently, few electronic health records include RPM data despite the increasing prevalence of data from continuous monitors because electronic health records were designed for structured and less variable health data. In addition, explicit decisions about the way in which RPM data is collected, processed and interpreted in practice are rarely made by analysts acting in isolation in health care, but by diverse stakeholders working in teams in sociopolitical contexts. For instance, in the data analytics exercise with an Australian hospital described by (Sharma, Stranieri, Ugon, Vamplew, & Martin, 2017), the problem, and interpretation of analytics results depended on stakeholder priorities at the executive, management and operational levels of the hospital. The data analytics process model CRISP-DM (Shearer, 2000) cannot readily accommodate diverse stakeholder priorities and also cannot easily be adapted for continuous analytics with

RPM data.

The openEHR (open Electronic Health Record pronounced open A'yr) standard that depicts the pragmatics of health care concepts described by (Kalra, Beale, & Heard, 2005) provides an important precursor to facilitate the application of Big Data analytics for RPM data. The use of openEHR has the potential to ensure data is correctly interpreted in analytics exercises and facilitate diverse stakeholder priorities and views. The next section in this chapter outlines the background literature, describes RPM research and provides an overview of the openEHR standard. Following that, the way in which openEHR can facilitate RPM Big Data analytics is described.

BACKGROUND

In general, an application system consists of a group of related application programs designed to perform certain functions. The RPM is an application system made up of two related applications, the healthcare application (HA) and the body area wireless sensor network (BAWSN), the monitoring application component. A BAWSN consists of a number of wireless sensors located on or in close proximity to the human body, such as on the clothing. The low-power sensors, such as medical sensors, wearable sensors, mobile sensors and fixed sensors, depending on the disease or needs of those aged and other patients, are equipped with a wireless interface and are capable of sensing the required intrinsic health data of that person over an extended period of time. In addition, these sensors can transmit the data to a monitoring application in a Local Processing Unit (LPU), generally a smartphone, for pre-processing. The distinct functions of a BAWSN are to authenticate the patient for continuous monitoring, to sense the vital health data from the patient, to pre-process the health data of the patient for sending any alert messages in the case of an emergency and to send the pre-processed data to the HA for further medical diagnoses. The HA is a sophisticated application assisting the doctors/care staff to monitor the patients' health condition and consult with the patient 'on the fly', regardless of where they are located. It is evident from early work by (Soini, Nummela, Oksa, Ukkonen, & Sydänheimo, 2008; Van Halteren et al., 2004; Van Laerhoven et al., 2004; Venkatasubramanian et al., 2005), the HA depends heavily on its monitoring component, the BAWSN, for the continuous generation of the health data. One of the pioneer RPM applications by (Balasubramanian et al., 2008), Active Care Loop Framework consist of Assistive Maternity Care application and a BAWSN capable of continuously monitor the Blood Pressure of a pregnant women and raise alarms using an SMS gateway is shown in Figure 1.

Although the BAWSN achieves the critical function of gathering trustworthy health data from the patient, the HA provides the visualisation of the patients' progress to the doctor and can have many functionalities. Examples are maintaining the electronic health records (EHR) in the database, alerting the concerned clinicians about the condition of the patients, the ability to provide a common ground for the patients and the care staff to discuss their needs in detail and in private; it can have an intelligent algorithm to predict any forthcoming emergency.

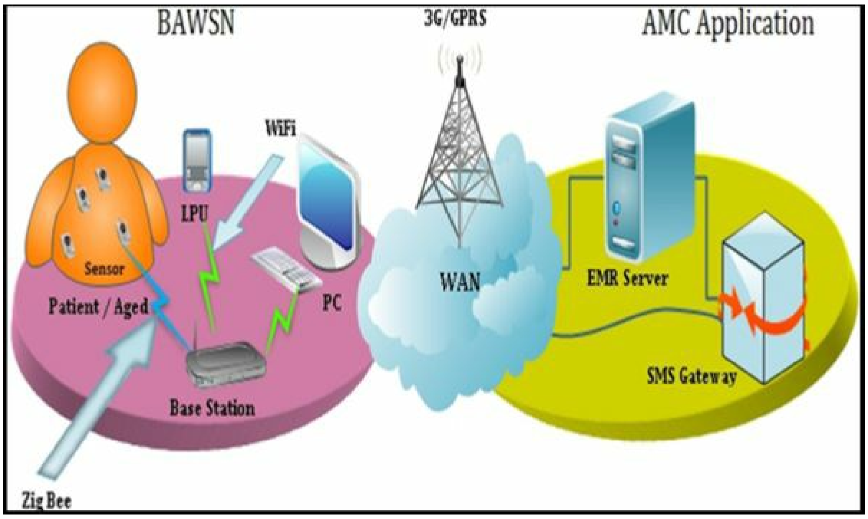

Figure 1: Active care loop framework for monitoring pregnant women

The general functionalities mentioned above are under the perspective of the user of this application. However, from the application developers' perspective, the specific implementation of these functionalities differs considerably depending on the health care requirements. For instance, the design of the electronic health records differs considerably for patients who are suffering from lymphoma and heart disease, and for those with other functionalities associated with an intelligent algorithm to predict any situation (Balasubramanian Appiah, 2012).

Therefore, the development of electronic health records requires a very high level of interoperability between diverse computer systems and extensive use of standards (Sitton & Reich, 2016). Government led electronic health record systems development tends to be enormously expensive and few countries have successfully implemented EHR systems despite the promise of potential efficiency gains and healthcare improvements that arise from access to so much data (Séroussi & Bouaud, 2016) (Garavand, Samadbeik, Kafashi, & Abhari, 2016). (Allen-Graham et al., 2018) outline the benefits and deficiencies of an electronic health record system introduced by the Australian government for a cost of well over $AUD 1 billion.

The standards essential for electronic health records include Open Systems Interconnection (OSI) network communication standards, messaging standards such as HL7 (Schloeffel, Beale, Hayworth, Heard, & Leslie, 2006) and medical vocabulary standards such as SNOMET-CT ((IHTSDO), 2018). Each standard maintained and kept up to date by worldwide communities engaging thousands of contributors. A great deal of importance on the benefits of having standardized terminologies for data mining exercises some years ago was emphasized by (Ramakrishnan, Hanauer, & Keller, 2010), when SNOMED-CT and Big Data were in their infancy. However, perhaps contrary to early expectations, the emergence of SNOMED-CT has not automatically facilitated Big Data analytics (BDa) (Benson & Grieve, 2016).

Reasons for this include the observation that coding of conditions, events, and test results to the appropriate SNOMED-CT code requires expertise and, in practice is often not done precisely or consistently, resulting in ambiguous data. For example: a variable "systolic blood pressure" may appear in a dataset with no indication of whether this refers to inter-arterial blood pressure measured with an intravenous device or the more common, around the cuff blood pressure. Relating blood pressures measures over time for the same patient is likely to result in very misleading analyses if the different kinds of blood pressure measures are confused. In addition, as (Matney et al., 2017) found, physiological variables used by diverse providers needed to be manually mapped to SNOMED-CT concepts in order to create a minimum data set of variables that could be used for data mining exercises. The concept of "patient height" may appear to be terminologically unique and well defined as the distance between the top of the head and the bottom of the feet, however this concept is inappropriate if the patient cannot stand straight or is an infant. Height data collected inappropriately is likely to hamper analytics exercises. Issues related to understanding the data is recognized as critical for BDa or Data mining exercise and is a key phase of the CRISP-DM reference model used to guide the execution of Data Mining exercises (SmartVision, Accessed 2017).

The CRISP-DM standard sets out six phases illustrated in *Figure 2*.

The first CRSIP-DM phase, *business understanding* focuses on understanding organizational objectives and identifying a data mining problem that is in alignment with the business objectives. The outcome is a preliminary plan designed to achieve project objectives. The next phase is the *data understanding* phase which provides understanding of the data that needs to be analysed. In the understanding phase, the data mining expert becomes familiar with the meaning and quality of the data. Following that phase, data needs to be prepared for modelling. The *data preparation* phase includes deciding what needs to be included in the dataset, cleaning the data and all other activities that need to be done to process the data, which serves as an input to the modelling step. In the *modelling* phase, a classification, prediction, association or clustering technique is applied on the data set and a model is generated. In the evaluation phase, the model is evaluated and results are analysed in relation to the business success criterion. If the model and the results are not in alignment with the organisational objectives, a new cycle of CRISP-DM is initiated otherwise, the model is deployed.

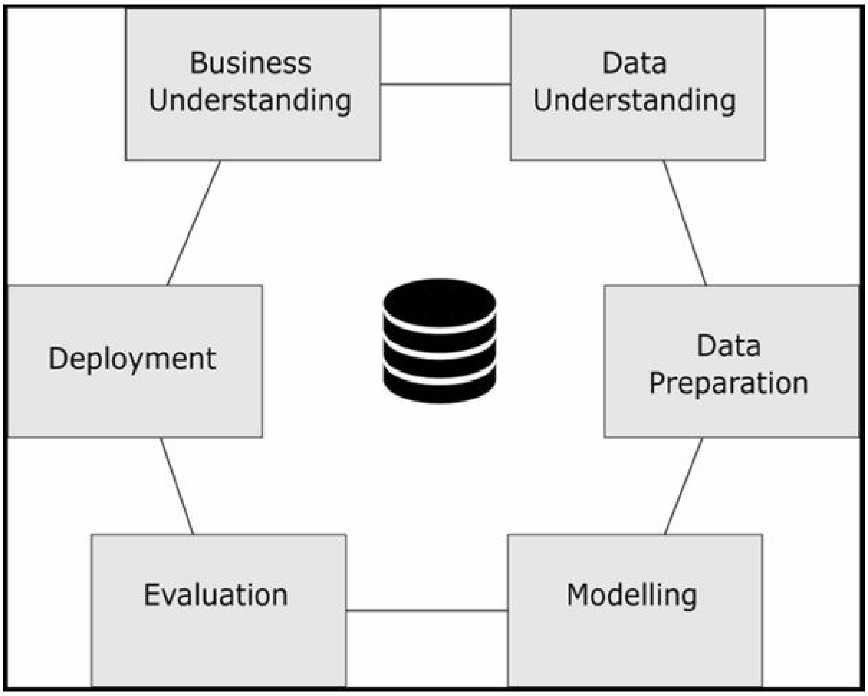

Figure 2: The CRISP-DM process model (adapted from (Chapman et al., 2000)

CRISP-DM has limited applicability for remote patient monitoring data because the temporal nature of RPM data requires additional abstraction as noted in (Catley et al., 2009). A patient's blood pressure measured continuously every 20 minutes over 24 hours may fluctuate between 140/70 mmHg and 110/90 mmHg for a particular patient. This level of fluctuation is not usually clinically significant so can be abstracted to a label like "Normal blood pressure". Conversely, a sudden drop in blood pressure from 150/80 mmHg to 90/60 mmHg in minutes warrants concern even if both measures are not clinically concerning in their own right.

Another limitation inherent in the CRISP-DM approach was raised in (Sharma et al., 2017). In their case study of a data analytics exercise in a hospital setting, they report that every aspect of the exercise required decisions and collaboration amongst groups of stakeholders within the organization. However, the way in which groups reason toward making decisions in an analytics exercise is not described or prescribed in the CRISP-DM process. This is paradoxical because major decisions including the specification of business objectives, the selection of a problem to focus on, the identification of relevant variables, and the ultimate interpretation of Big Data Analysis findings are rarely made by a single decision maker but involve a complex interplay between and within staff at operational, management and executive levels.

This chapter outlines challenges for Big Data Analytics that specifically arise in the presence of Remote Patient Monitoring data. *Table 1* provides an outline of the main challenge which are elaborated on in subsequent sections. A key feature of the chapter involves the assertion that many Big Data challenges inherent in remote patient monitoring can be reduced with the use of standards. The openEHR standard outlined in the next section, is sufficiently expressive for this.

Table 1: Overview of big data analysis challenges for remote patient monitoring

Phenomena	Challenge
The interpretation of Big Data Analyses in health requires input from many individuals across multiple disciplines	Decision making techniques designed to support groups to reach decisions are rarely used to facilitate consensus between stakeholders interpreting Big Data Analytics exercises in healthcare
The extent to which Big Data Analyses realise business objectives is a recommended evaluation criteria by CRISP-DM, however this is not an effective criteria to evaluate BDA using RPM data.	The challenge is to discover criteria for the evaluation of BDA exercises on RPM data that do not rely on abstract statements of business objectives.
BDA is difficult to perform live on RPM because few analytics algorithms can operate on streams of data in real time	The challenge is to discover analytics algorithms that can operate quickly on partial data, then revise analyses as new data streams in.
RPM data cannot easily be integrated into electronic health records.	The challenge is for RPM developers and EHR developers to use common standards to encode meta-data that will facilitate inter-operability of health information.
Existing standards for inter-operability of health data including HL7 are not well suited to encoding meta-data from RPM datasets.	The challenge is to adapt the openEHR standard so that RPM data can be readily encoded. The openEHR standard requires minor modification to existing archetypes compared with HL7 which would need major adjustments.
The meaning of each variable in an RPM exercise is often difficult to ascertain	The challenge is to collect meta-data that describes each variable with sufficient detail to enable Analysts to correctly interpret RPM data.

Outline of openEHR

The demands of interoperability between health care provider computer systems has driven the development of standards in addition to OSI network standards. The openEHR standard (Kalra et al., 2005) (www.openehr.org) was proposed over a decade ago as an attempt to model the pragmatics of health care knowledge. This was considered to be critical for the design of electronic health records systems and the achievement of the interoperability required.

The archetype is a core primitive in the openEHR standard. An archetype models a concept in use within health care with the following elements: concept name, description, purpose, use and misuse. For example, the archetype named "Blood pressure" listed in the openEHR clinical knowledge base (http://openehr.org/ckm/) is linked to SNOMED-CT Concept 16307200007. In the "Blood pressure" openEHR archetype, the blood pressure concept is described as the local measurement of arterial pressure as a surrogate for pressure in systemic circulation. The purpose of the concept is to record an individual's blood pressure. The appropriate use and misuse are listed. For instance, the concept is not to be used to refer to intravenous blood pressure. The openEHR archetype for "Blood pressure" also includes a description of the data associated with the measurement of blood pressure. This includes definitions and units of measure for systolic, diastolic, mean arterial pressure and pulse pressure. The state of the individual when the "Blood pressure" is measures is also specified in the archetype; for instance the assumed position is sitting but "Blood pressure" can also be taken standing or reclining. Descriptors relevant for a protocol for the measure of "Blood pressure" including cuff size, location on the body, various formulas and the type of device used are also specified in the archetype. Version 1.1.1 of the "Blood pressure" archetype was attributed to an originator, Sam Heard in 2006 (openEHR, 2018). A community comprised of over 30 contributors whose names are listed in the archetype is also included.

In the next section of this chapter, the way in which the openEHR standard can facilitate data analytics with Big Data that derives from remote patient monitoring will be outlined. The approach accommodates the group reasoning amongst diverse stakeholders inherent in most data mining exercises despite not being made explicit in the CRSIP-DM process model. The innovations will be described with reference to a case study involving remote monitoring of vital signs amongst patients in an Indian Hospital.

OPENEHR FOR RPM ANALYTICS

RPM Trial

An architecture was designed and implemented by (Balasubramanian & Stranieri, 2014) that enables the transmission of patient data to Cloud-based repositories, as shown in Figure 3, where software services invoked by health care providers can be instantiated, executed and terminated readily to securely and efficiently process all or part of a patient's data collected continuously with wearable sensors. The prototype of the architecture was trialed in a Medical College Hospital, Coimbatore, India in 2016. Consenting patients in a general medical ward at the hospital were fitted with wearable sensors capable of monitoring ECG, blood pressure, temperature, respiratory rate and heart rate. The sensors were configured to continuously transmit data to a nearby Android Tablet running prototype software developed by (Balasubramanian & Stranieri, 2014).

Twelve patients were selected to participate in the trial over a three-week period. The data was streamed by room and bed number only so that the patient privacy was maintained. Nurses who volunteered for the trial were trained to recharge sensors, locate them on the patients, and check that data transmission had commenced. *Table 2* illustrates sample raw data for blood oxygen (SPO2), diastolic blood pressure (DiaPress), systolic blood pressure (SysPress), pulse rate and respiratory rate. *Table 1* presents data collected from one patient for 27 seconds during the trial. A "0" was entered into the array between sensed episodes. The device did not broadcast any meta-data so the units of measurement for each variable were only understood from the manufacturer's technical manuals. A value for the body temperature seemed to be included in every second's transmission except when the SPO2 and pulse was recorded.

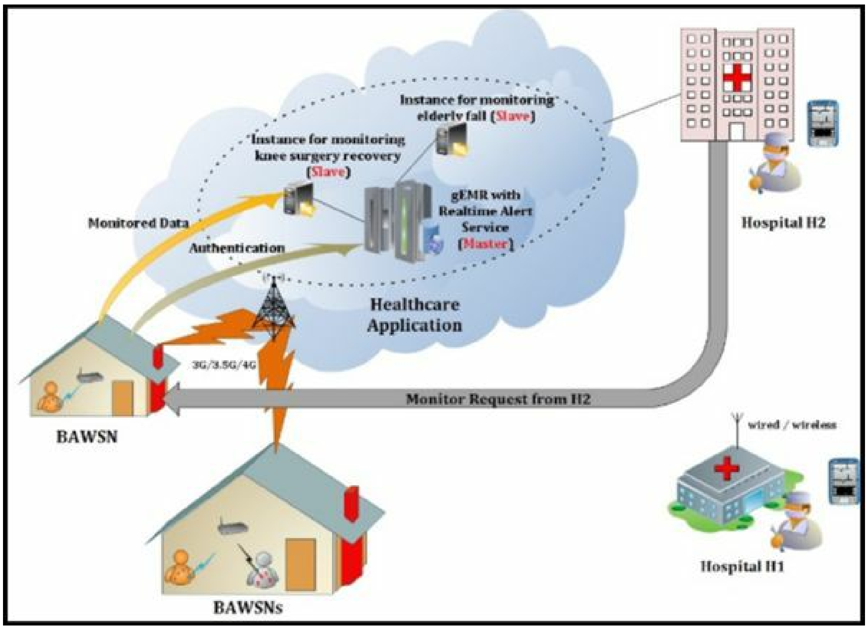

Figure 3: Architecture Design for Assistive Patient monitoring cloud Platform for active healthcare Applications (AppA)

During the trial, the nurses had to remove the sensors an average of two times per day per patient, to help the patient to use the restroom or to have lunch. States such as active, paused, disconnected, are not identified in a standard way and rarely captured in RPM devices but become very important for RPM Analytics exercises.

The transfer of data from the tablet to the cloud server used a TCP/IP connection with the total payload of 26 Bytes, the TCP packet sent every second from a sensor will be less than 100 Bytes which includes the maximum TCP header size of 60 Bytes. Therefore, the total amount of data produced for one patient will not be more than 300 Megabytes every month. However, one hundred patients monitored in this way generates 360 Gbytes of data per year. The vast majority of this data is not of direct clinical interest for treating physicians, however once collected in digital form, health record legislation in most jurisdictions mandate that digital health data be stored and only deleted following onerous procedures. Most hospital information systems are not designed to store RPM data so storage must be done outside these systems with safeguards in place to ensure privacy and security.

The data was processed in real time by software executing on the Tablet and in the Cloud to raise SMS alarms to nurses and doctors mobile phones during the trial. Data regarding the status of the messages was not captured for the trial however this can be regarded as useful data for future data analysis exercises as discussed further below. Data relating to remote physician login such as the login duration, delays, and outcomes was not collected but can also be expected to be useful for future analyses.

The Trial illustrated that remote and continuous patient monitoring can be seen as technology that has recently arisen that enables a great deal of data to be generated continuously. However, as RPM continues to be adopted by healthcare systems, problems for data analytics exercises can be expected to emerge that dramatically reduce the utility of the data. Two challenges include:

- **Meta-Data:** Describing the data generated, on the fly, that includes units of measure needs to be associated with each bucket of data collected. For interoperability with other systems such as hospital information systems, the meta-data needs to be expressed using the openEHR standard though some extensions are required to accommodate RPM.
- **Real Time Analytics:** The incremental acquisition of data from real time sensors raises the possibility of real time analytics. Automated raising of alarms is an obvious application of real time analytics for RPM.

Table 2: Sample data for 27 seconds RPM

SPO2	SysPress	DiaPress	PulseRate	RespRate	Temp1	SensorId	PatientId	CreatedDate
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
95	0	0	80	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
95	0	0	81	0	0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
0	0	0	0	0	37.0	5	4	17/10/2016 4:22
95	0	0	81	0	0	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	0	0	0	0	37.1	5	4	17/10/2016 4:22
0	140	90	0	0	37.1	5	4	17/10/2016 4:22

SOLUTIONS AND RECOMMENDATIONS

Meta-Data

The sample data illustrated in *Table 2* can be imagined to be useful for real time processing to raise alarms that the patient is deteriorating. If the vitals signs are not alarming, the data does not need to be stored beyond its collection. Indeed, the vast majority of static patient monitoring devices currently in use in hospitals do not store data after its collection and simply provide an optional print out facility. However, RPM data is collected digitally so there is a legislative requirement in most jurisdictions to store the data (Swire, 2013). For instance, the Health Records Act in Victoria, Australia prescribes detailed processes for the archiving of health records and deletion is only permissible to correct errors.

The storage of RPM data such as that in *Table 2* can be seen to be of use for future, off-line analytics purposes in addition to real time purposes. For example, associations between features can be used for predictions, particularly if linked with other data. Lee et al (2010) present an example of the potential that data mining can bring to monitoring by integrating environmental data including weather information with patient bio-data to predict an asthma attack. Data such as that presented in *Table 2* can be linked to medication and demographic information about patients across multiple hospitals for outcomes such as the early detection of adverse reactions to medications. If the data was linked to staffing data in hospital information systems, a raft of analyses to do with workplace efficiencies can be readily imagined.

Linking data collected at different times and places across diverse repositories becomes extremely labor intensive unless meta-data describing the data collected is included with the data. Without the meta-data, the "Understanding Data" phase in the CRISP-DM process involves discovering what each feature is, the unit of measure, the collection method and many more meta-data concepts. This kind of exercise with RPM data cannot be imagined to be successfully performed by data analysts without clinical knowledge. For example, in *Table 2* the temperature occasionally drops from 37.1 degrees to 0 degrees for a second then returns to 37.0 degrees. Clinical knowledge is required to understand that this is not a real drop in body temperature but an indication that the temperature was not sensed at that instance. Similarly, clinical knowledge is required to know that the temperature feature may refer to surface temperature, core temperature, temperature at extremities or other commonly used temperature measures and that each measure of temperature ought not be confused with other measures.

Meta-data stored with the RPM data enables the "Understanding Data" phase to be performed far more easily. However, this is likely to be the case only to the extent that the Meta-data conforms to a common standard. The contention advanced here is that the openEHR standard is sufficiently expressive to capture the semantics and pragmatics of RPM data with some expansion. An archetype currently exists for each feature listed in *Table 2* such as the blood pressure archetype described above. That archetype clearly distinguishes between systolic and diastolic blood pressure and defines units of measure. An archetype can be used as a rich template whenever RPM data is initialized for storage to enable clear and comparable descriptions of the data later when RPM Analytics exercises are performed.

Additional fields that are required for RPM include the status of the monitoring (e.g. active, off, paused), the status and history of the remote access, and feature based interpretation (e.g. 0 means no data is sensed). (Robles-Bykbaev, Quisi-Peralta, López-Nores, Gil-Solla, & García-Duque, 2016) demonstrate that meta-data expressed as openEHR archetypes can lead to a great deal of automation in the development of decision support system knowledge bases from data mining.

Real Time Analytics

Data streams pose unique space and time constraints on the computation process (Aggarwal & Philip, 2007). Unlike conventional data mining, stream mining approaches must occur in real time, which challenges computational processing efficiency. Further, most machine learning algorithms have been developed assuming all data is available to the algorithm. (Sanila, Subramanian, & Sathyalakshmi, 2018) reviews real time mining techniques to reveal the adaptation of algorithms for use when data streams incrementally in, is a pressing research problem. One approach to deal with this involves data summarisation where data is typically segmented into windows and reduced using filters. For instance, (Allami, Stranieri, Balasubramanian, & Jelinek, 2016) presents a low computational resource algorithm for reducing ECG data without losing key data points critical for diagnoses. Another approach involves incrementing subsequence counts (Abadia, Stranieri, Quinn, & Seifollahi, 2011) as each new data streams in. Sub-sequence counts can be directly used in classification algorithms advanced by (Quinn, Stranieri, Yearwood, Hafen, & Jelinek, 2008). Frequency counts profiles of interbeat heart rate variability has been shown by (Allami, Stranieri, Balasubramanian, & Jelinek, 2017) to predict future heart rate variability and some heart conditions.

Real time analytics with data streams is enhanced when it is linked with other data. The data stream exemplified in *Table 2* can be enhanced with links to the patient's conditions, medication, demographic or other relevant data. However, accessing static data stored in other repositories during a real time analysis of the stream presents severe computational complexity challenges. Setting up processes to perform analyses that link stream data with static data requires a great deal of labor intensive work and challenges in the "Understanding Data" phase as alluded to above.

Analytics with Big Data in practice requires that many decisions in the analytics process be made by a group of stakeholders who often have competing priorities. This is particularly the case in the business understanding, data understanding and evaluation phases of a CRISP-DM process. The pilot study (Sharma et al., 2017) revealed three main categories of decisions that were made by groups in a hospital data mining exercise; decisions about what problem to focus on, how to interpret data, and how to resource the data mining exercise.

Decision made by groups involve characteristics that render the decision making process quite different than decisions made by individuals. A group, deliberating toward a decision has been called a Reasoning Community by (Yearwood & Stranieri, 2010) who advance a process for ensuring that some of the deficiencies of group reasoning including groupthink (Janis, 1972), shared information bias (Wittenbaum, 2000) and argument fallacies are contained. A Reasoning Community refers to any group of participants that reason individually, communicate with each other, and attempt to coalesce their reasoning in order to reason collectively to perform an action or solve a problem. Reasoning communities are viewed as broader and more encompassing than communities of practice(Lave & Wenger, 1998).

The process for ideal deliberation advanced by (Yearwood & Stranieri, 2010) includes tasks to be performed in four key phases illustrated in *Figure 4* and include: Engagement, Individual reasoning, Coalesced Reasoning and Decision making described below.

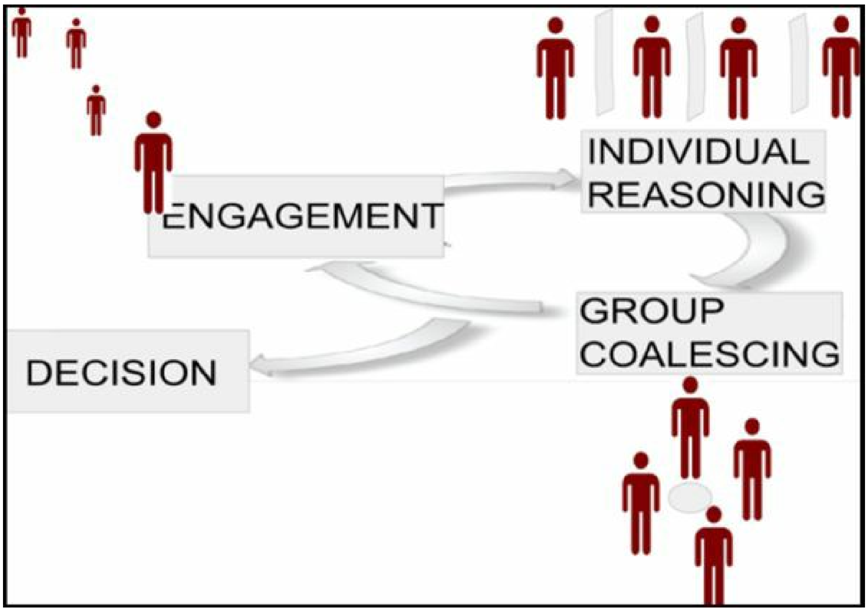

Figure 4: Group reasoning in the CRISP-DM process

The Engagement phase for an RPM Analytics exercise involves the selection and recruitment of the people who participate in the decision making process. An RPM Analytics process can be expected to involve clinicians along with data analysts but may also require Internet of Things (IOT) experts. This phase also involves the articulation of the issue to be resolved. RPM Analytics is new and emerging, so articulating the issue that could be resolved except at a very high and abstract level such as "increase efficiency" may be difficult.

In the individual reasoning phase of a Reasoning Community process, each participant establishes facts, makes inferences from facts to draw conclusions and, by so doing contributes reasons to a pool of reasons for the community. A key part of individual reasoning involves an individual's coalescing of reasoning. This is the process of juxtaposing background knowledge with reasons advanced by other participants in order to understand the issue and position his or her claims amidst the others. A participant's coalescing of reasoning involves making sense of reasons in order to assert their own claims or to understand the claims of others. Following the Reasoning Community model, each participant of an RPM Analytics process will initially analyse the data independently from others. Initially, there is no interaction or exchange of thoughts/ideas between the group members in order to avoid negative consequences of group interactions such as groupthink (Janis, 1972).

Group coalescing of reasoning involves organizing the analyses advanced by each participant into an explicit, coherent representation. This is important for shared and democratic decision-making where decisions are made on the basis of reasoned debate. Further, group coalescing enables communities in the future to adopt coalesced reasoning as a starting point for their own deliberations in what Stranieri and Yearwood (Stranieri & Yearwood, 2012) call re-use of reasoning. Most analytics exercises perform individual reasoning but do not systematically perform group coalescing.

Finally, the decision making phase of a Reasoning Community depicts the stage when participants must decide on an ultimate interpretation from the RPM Analyses. Many patterns noticed in an analytics exercise are deemed to be spurious or uninteresting. Conclusions that are reached as a result of the RPM Analytics are typically determined to be worthwhile by the entire team including clinicians, data analyst and management.

The recognition that any Analytics exercise is not performed in isolation by an analyst but occurs in a socio-political context where a group of interacting individuals performs at each phase of the CRISP-DM process. RPM Analytics exercises, currently in their infancy, are likely to require a more diverse group of stakeholders that each initially arrive at diverse conclusions that must be assimilated into an agreed analytics outcome.

FUTURE RESEARCH DIRECTIONS

Future research is required to explore how RPM endeavors are actually established and conducted as their prevalence emerges. Work is also required to validate the difficulties inherent in the use of the openEHR standard to specify meta-data needs of RPM data.

CONCLUSION

Remote patient monitoring systems collect data from patients, typically with wearable sensors, and transfer the data to servers so that health care professionals can remotely log in to view the data in real time. Although these systems are emerging, to date little attention has been placed on the challenges inherent in analyzing data collected from remote patient monitoring systems. In this chapter, the openEHR standard was presented as an important standard for specifying the variables and context for the data collected so that data collected from diverse RPM systems or at different times can be more readily compared and analysed. The view that openEHR can be used to describe meta-data inherent in collecting RPM data. The chapter also advanced the notion that RPM analytics exercises, like any analytics process is not performed by an analyst in isolation but involves a group of stakeholders who have diverse interests, expertise and background. Real time analytics is challenging because data streams need to be analysed in real time and ideally linked with static data stored in electronic health records. Analyses reached by individuals need to be validated and confirmed by stakeholders for analyses to be accepted. Remote patient monitoring in a Big Data era has the potential to add another dimension to health care, however many technical, organizational and clinical challenges need to be addressed before useful outcomes of analyses emerge.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Abadia, R., Stranieri, A., Quinn, A., & Seifollahi, S. (2011). *Real time processing of data from patient biodevices*. *Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management-Volume 120*.

Aggarwal, C. C., & Philip, S. Y. (2007). *A survey of synopsis construction in data streams*. In *Data Streams* (pp. 169–207). Springer.

Allami, R., Stranieri, A., Balasubramanian, V., & Jelinek, H. F. (2016). *ECG Reduction for Wearable Sensor*. Paper presented at the Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on. 10.1109/SITIS.2016.88

Allami, R., Stranieri, A., Balasubramanian, V., & Jelinek, H. F. (2017). *A count data model for heart rate variability forecasting and premature ventricular contraction detection*. *Signal, Image and Video Processing*, 1–9.

Balasubramanian, V., Hoang, D. B., & Ahmad, N. (2008). *SOAP based Assistive Care Loop using wireless sensor networks*. Paper presented at the IT in Medicine and Education, 2008. ITME 2008. IEEE International Symposium on. 10.1109/ITME.2008.4743897

Balasubramanian, V., & Stranieri, A. (2014). *A scalable cloud Platform for Active healthcare monitoring applications*. Paper presented at the e-Learning, e-Management and e-Services (IC3e), 2014 IEEE Conference on. 10.1109/IC3e.2014.7081248

Balasubramanian, V., Stranieri, A., & Kaur, R. (2015). *AppA: assistive patient monitoring cloud platform for active healthcare applications*. *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*. 10.1145/2701126.2701224

Balasubramanian Appiah, V. (2012). *Dependability of body area wireless sensor networks in assistive care loop framework (PhD)*. University Technology Sydney.

Breen, S., Kofoed, S., Ritchie, D., Dryden, T., Maguire, R., Kearney, N., & Aranda, S. (2017). *Remote real-time monitoring for chemotherapy side-effects in patients with blood cancers*. *Collegian (Royal College of Nursing, Australia)*, 24(6), 541–549. doi:10.1016/j.colegn.2016.10.009

Brown, S. M., Jones, J., Kuttler, K. G., Keddington, R. K., Allen, T. L., & Haug, P. (2016). *Prospective evaluation of an* *automated method to identify patients with severe sepsis or septic shock in the emergency department*. *BMC Emergency Medicine*, 16(1), 31. doi:10.118612873-016-0095-0 PMID:27549755

Carta, M., & Angst, J. (2016). *Screening for bipolar disorders: A public health issue*. *Journal of Affective Disorders*, 205, 139–143. doi:10.1016/j.jad.2016.03.072 PMID:27442457

Catley, C., Smith, K., McGregor, C., & Tracy, M. (2009). *Extending CRISP-DM to incorporate temporal data mining of multidimensional medical data streams: A neonatal intensive care unit case study*. Paper presented at the Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on. 10.1109/CBMS.2009.5255394

Chan, M., Estève, D., Fourniols, J.-Y., Escriba, C., & Campo, E. (2012). *Smart wearable systems: Current status and future challenges*. *Artificial Intelligence in Medicine*, 56(3), 137–156. doi:10.1016/j.artmed.2012.09.003 PMID:23122689

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. Academic Press.

IHTSDO. (2018). *SNOMED-CT*. Retrieved from *https://www.snomed.org*

Janis, I. L. (1972). *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes*. Academic Press.

Kakria, P., Tripathi, N., & Kitipawang, P. (2015). *A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors*. *International Journal of Telemedicine and Applications*, 2015, 8. doi:10.1155/2015/373474 PMID:26788055

Kalid, N., Zaidan, A. A., Zaidan, B. B., Salman, O. H., Hashim, M., & Muzammil, H. (2018). *Based Real Time Remote Health Monitoring Systems: A Review on Patients Prioritization and Related "Big Data" Using Body Sensors information and Communication Technology*. *Journal of Medical Systems*, 42(2), 30. doi:10.100710916-017-0883-4 PMID:29288419

Kalra, D., Beale, T., & Heard, S. (2005). *The openEHR foundation*. *Studies in Health Technology and Informatics*, 115, 153–173. PMID:16160223

Klonoff, D. C., Ahn, D., & Drincic, A. (2017). *Continuous glucose monitoring: A review of the technology and clinical use*. *Diabetes Research and Clinical Practice*, 133, 178–192. doi:10.1016/j.diabres.2017.08.005 PMID:28965029

Lave, J., & Wenger, E. (1998). *Communities of practice*. Academic Press.

Matney, S. A., Settergren, T., Carrington, J. M., Richesson, R. L., Sheide, A., & Westra, B. L. (2017). *Standardizing Physiologic Assessment Data to Enable Big Data Analytics*. *Western Journal of Nursing Research*, 39(1), 63–77. doi:10.1177/0193945916659471 PMID:27435084

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). *Big data analytics capabilities: A systematic literature review and research agenda*. *Information Systems and e-Business Management*, 1–32. doi:10.100710257-017-0362-y

openEHR. (2018). *Clinical Knowledge Manager*. Retrieved from *http://openehr.org/ckm/*

Quinn, A., Stranieri, A., Yearwood, J., Hafen, G., & Jelinek, H. F. (2008). *AWSum-Combining Classification with Knowledge Aquisition*. *Int. J. Software and Informatics*, 2(2), 199–214.

Robles-Bykbaev, V., Quisi-Peralta, D., López-Nores, M., Gil-Solla, A., & García-Duque, J. (2016). *SPELTA-Miner: An expert system based on data mining and multilabel classification to design therapy plans for communication disorders*. Paper presented at the Control, Decision and Information Technologies (CoDIT), 2016 International Conference on.

Sanila, S., Subramanian, D. V., & Sathyalakshmi, S. (2018). *Real-Time Mining Techniques: A Big Data Perspective for a Smart Future*. *Indian Journal of Science and Technology*, 10(42), 1–7. doi:10.17485/ijst/2017/v10i42/120344

Schloeffel, P., Beale, T., Hayworth, G., Heard, S., & Leslie, H. (2006). *The relationship between CEN 13606, HL7, and openEHR*. *HIC 2006 and HINZ 2006: Proceedings, 24*.

Sharma, V., Stranieri, A., Ugon, J., Vamplew, P., & Martin, L. (2017). *An Agile Group Aware Process beyond CRISP-DM: A Hospital Data Mining Case Study*. *Proceedings of the International Conference on Compute and Data Analysis*. 10.1145/3093241.3093273

Shearer, C. (2000). *The CRISP-DM model: The new blueprint for data mining*. *Journal of Data Warehousing*, 5(4), 13–22.

Shimmer. (2018). Retrieved from *http://www.shimmersensing.com/products/shimmer3-ecg-sensor#specifications-tab*

Sitton, M., & Reich, Y. (2016). *Enterprise Systems Engineering for Better Operational Interoperability*. *Systems Engineering*.

SmartVision. (n.d.). *CRSP-DM*. Retrieved from *http://crisp-dm.eu*

Soini, M., Nummela, J., Oksa, P., Ukkonen, L., & Sydänheimo, L. (2008). *Wireless body area network for hip rehabilitation system*. *Ubiquitous Computing and Communication Journal*, 3(5), 42–48.

Stranieri, A., & Yearwood, J. (2012). *The Case for a Re-Use of Community Reasoning*. In J. Yearwood & A. Stranieri (Eds.), *Approaches for Community Decision Making and Collective Reasoning* (pp. 237–249). IGI Global.

Swire, P. (2013). *Finding the best of the imperfect alternatives for privacy, health it, and cybersecurity*. *Wisconsin Law Review*, (2): 649–669.

Thilo, F. J., Hürlimann, B., Hahn, S., Bilger, S., Schols, J. M., & Halfens, R. J. (2016). *Involvement of older people in the development of fall detection systems: A scoping review*. *BMC Geriatrics*, 16(1), 42. doi:10.118612877-016-0216-3 PMID:26869259

Van Halteren, A., Bults, R., Wac, K., Konstantas, D., Widya, I., Dokovsky, N., … Herzog, R. (2004). *Mobile patient monitoring: The mobihealth system*. *The Journal on Information Technology in Healthcare*, 2(5), 365–373.

Van Laerhoven, K., Lo, B. P., Ng, J. W., Thiemjarus, S., King, R., Kwan, S., … Needham, P. (2004). *Medical healthcare monitoring with wearable and implantable sensors*. *Proc. of the 3rd International Workshop on Ubiquitous Computing for Healthcare Applications*.

Venkatasubramanian, K., Deng, G., Mukherjee, T., Quintero, J., Annamalai, V., & Gupta, S. K. (2005). *Ayushman: A wireless sensor network based health monitoring infrastructure and testbed*. *Proceedings of the First IEEE international conference on Distributed Computing in Sensor Systems*. 10.1007/11502593_39

Wang, Y., Kung, L., & Byrd, T. A. (2018). *Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations*. *Technological Forecasting and Social Change*, 126, 3–13. doi:10.1016/j.techfore.2015.12.019

Wittenbaum, G. M. (2000). *The bias toward discussing shared information: Why are high-status group members immune*. *Communication Research*, 27(3), 379–401. doi:10.1177/009365000027003005

Yearwood, J., & Stranieri, A. (2010). *Technologies for Supporting Reasoning Communities and Collaborative Decision Making: Cooperative Approaches: Cooperative Approaches*. IGI Global.

Zhu, Z., Barnette, R. K., Fussell, K. M., Michael Rodriguez, R., Canonico, A., & Light, R. W. (2005). *Continuous oxygen monitoring—A better way to prescribe long-term oxygen therapy*. *Respiratory Medicine*, 99(11), 1386–1392. doi:10.1016/j.rmed.2005.03.010 PMID:15878655

ADDITIONAL READING

Stranieri, A., & Yearwood, J. (2012). *The Case for a Re-Use of Community Reasoning*. In J. Yearwood & A. Stranieri (Eds.), *Approaches for Community Decision Making and Collective Reasoning* (pp. 237–249). IGI Global Press.

KEY TERMS AND DEFINITIONS

CRISP-DM:
A process model for performing data mining exercises.

Electronic Health Record:
A virtual record of major health related events for an individual from before birth to after death.

openEHR:
A standard that describes clinical concepts and their use.

Reasoning Community:
A model of how individuals reason together to solve a problem. This model can be applied to describe how analysts and other stakeholders interact to analyze data from remote patient monitoring systems.

Remote Patient Monitoring:

Monitoring patients physiological signs. This is typically performed with wearable, implantable or digestible sensors but may be done at a distance with camera surveillance.