Portuguese hospitals' main challenges in implementing Big Data projects for early detection of adverse events

Ana Oliveira Bianchi de Aguiar

Advisor

Professor Susana Frazão Pinheiro

Dissertation submitted in partial fulfilment of requirements for the Degree of MSc in International Business Administration, at Universidade Católica Portuguesa

May 2015

CATÓLICA LISBON
BUSINESS & ECONOMICS

Abstract

Title: Portuguese hospitals' main challenges in implementing Big Data projects for early detection of adverse events
Author: Ana Oliveira Bianchi de Aguiar

Big Data has been creating much excitement and promises to solve many of the current health systems' challenges. A specific application allows predicting adverse events, such as nosocomial infections, 24-48 hours earlier than traditional methods, by analysing in real-time physiological data allied with clinical information, and by extracting knowledge from this stored data. However, the implementation of this kind of projects is not without challenges. Hence, the objective of this thesis is to understand the main barriers in implementing Big Data projects for early detection of adverse events in the specific case of Portuguese hospitals.

The collection of primary data, through surveys and interviews, allowed identifying three main barriers. Firstly, there is a generalized low knowledge regarding Big Data, which can hinder the consideration of these projects in the yearly budget and create difficulties in understanding how it can be applied and benefit the hospital. Secondly, a shortage of "Data Scientists" in Portuguese hospitals was reported, being this skilled labour crucial to creatively look at the data and understand how it generates value. Finally, an initial high investment with still undiscovered business value is a true barrier, reflecting the hospitals' budget constraints.

However, two initially identified obstacles were not validated by this analysis. Firstly, being an organizational change necessary to adapt to this new paradigm, resistance from managers and caregivers is not expected. Furthermore, data security and privacy were not considered true impediments but rather a requirement of the technology.

"Big Data" tem vindo a despertar muita atenção e promete resolver os principais desafios que os sistemas de saúde hoje enfrentam. Uma aplicação específica permite prever intercorrências, como infeções adquiridas no hospital, 24-48 horas mais cedo do que os métodos tradicionais, através de uma análise em tempo real de fluxos fisiológicos e informação complementar, tal como da extração de novos algoritmos integrados nos dados armazenados. Contudo, a implementação destes projectos tem associada desafios e dificuldades. Assim, o objetivo desta tese é compreender quais as principais barreiras à implementação de projectos de "Big Data" para deteção precoce de intercorrências, no caso específico dos hospitais portugueses.

Dados recolhidos através de inquéritos e entrevistas, permitiram identificar três barreiras principais. Primeiramente, o nível de conhecimento sobre "Big Data" é baixo, o que poderá impedir a inclusão deste tipo de projetos no orçamento e dificultar o entendimento relativamente à sua aplicação no meio hospitalar. Seguidamente, foi reportada uma carência generalizada de "Data Scientists", sendo estes cruciais para olhar de forma criativa para os dados, compreendendo como podem gerar valor. Finalmente, a necessidade de existir um elevado investimento inicial, associada à falta de evidência relativamente aos benefícios, foi considerada uma barreira, refletida nas restrições orçamentais dos hospitais.

Contudo, dois obstáculos inicialmente identificados, não foram validados pela análise. Primeiro, sendo necessária uma transformação organizacional, não é esperada resistência por parte dos gestores ou médicos e enfermeiros. Por outro lado, segurança e privacidade dos dados não foram consideradas uma barreira, mas algo que a tecnologia teria que garantir.

Table of Contents

Table of Illustrations

Preface - Acknowledgements

This work was not an individual effort – many intervened to guarantee its completion and success.

It started with all the teachers that contributed to my growth and learning process, which culminated in this final work. Particularly, my tutor, Susana Frazão Pinheiro, to whom I thank all the great insights, enthusiasm and collaboration.

Moreover, this work's merit much depended on the enormous time effort of those who kindly answered and distributed my surveys and, especially, of those who undertook a person-to-person discussion with me - I thank them all.

However, this was not only a professional endeavour, but embedded all aspects of my life. My family played an important role when I faced obstacles, giving me the strength to believe and work towards this goal.

And I thank to my beloved one, who always believes in my skills and pushes me forward.

I thank everyone that contributed to this work and who made it so exciting.

1. Introduction

Currently, health systems are under extreme pressure, facing challenges such as population ageing and chronic diseases, the rising of costs without the correspondent quality improvement and an uneven access to care (Deloitte, 2014; My Health London, 2015). For example, in 2011, health expenses reached $6.9 trillion (WHO, 2014b). Indeed, the Portuguese health system shares these challenges, aggravated by the recent debt crisis and consequent austerity measures, as well as the systems' poor governance (Sakellarides *et al*., 2005).

Nevertheless, Big Data may be an important catalyst in solving these challenges. Despite the excitement around it, Big Data holds a fast-evolving definition (Gandomi & Haider, 2015), which has been grasped by the concept of "Vs" – the data is high "volume" with a "variety" of sources and formats, flowing and analysed at high "velocity" (McAfee & Brynjolfsson, 2012; Gandomi & Haider, 2015), hence generating economic "value" (Gantz & Reinsel, 2012), with "veracity" (White, 2012). In this context, Mckinsey (2013) identifies five main pathways in which Big Data may create value in healthcare: right living, right care, right provider, right value and right innovation.

Inside the right care, it is possible to emphasize a particular Big Data application for prediction of adverse events, such as nosocomial infections or post-surgical complications. Briefly, Big Data allows to predict with 24-48 hours in advance the occurrence of such events not only through the analysis in real-time of physiological data combined with complementary clinical information, but also through knowledge extraction of this stored data (Kohn *et al.*, 2014). Indeed, this application could enormously prevent morbidity and mortality (Khazaei *et al.*, 2014), being able to drive healthcare efficiency and dramatically enhance patient outcome as well as reduce their length of stay (Blount *et al*., 2010).

However, the implementation of Big Data projects is not without barriers and challenges. In this context, the literature emphasizes data privacy and security concerns (Feldman *et al*., 2012), cultural and organizational inertia (McAfee & Brynjolfsson, 2012), skilled labor constraints (Chen *et al.*, 2012) and the high initial investment and unclear benefits (Zillner *et al.*, 2014) as main barriers.

Therefore, this thesis aims to understand **Portuguese hospitals' main challenges in implementing Big Data projects for early detection of adverse events.**

Hence, based on the Literature Review, Hypotheses were formulated for the specific case of Big Data for early detection of adverse events in Portuguese hospitals. In fact, 5 main Hypotheses were suggested and analysed based on primary data collected by the researcher through surveys and interviews to both caregivers and managers.

H1: Security and Privacy are a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

H2: Lack of IT skilled labour is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

H3: Cultural and organizational rigidity is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

H4: Budget constraints and undiscovered business value is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

H5: The absence of knowledge regarding Big Data and its potential benefits is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

In this context, this study is relevant at an academic and managerial level.

On the one hand, Big Data is extremely underdeveloped from an academic point of view, with only 44 articles in 2012 (Wamba *et al.*, 2015) - a comprehensive Literature Review, particularly for the case of Portugal, is a significant contribution. Besides, it is relevant to provide insights on the barriers of this specific case, as they may differ among geographies and applications.

On the other hand, understanding these challenges is extremely important for Portuguese providers (hospitals), the Government and third party suppliers/partners. This insight will allow them to act upon the true constraints, thus promoting this kind of projects, which have been proved to play a role in solving the health system current issues.

Therefore, the thesis is composed by 6 main chapters. Chapter 2 is a comprehensive Literature Review on subjects such as health systems' challenges, Big Data definition and opportunities, Big Data state-of-the-art in Portugal and main barriers in implementing Big Data projects. Chapter 3 presents the Hypotheses to test and the utilized methodology. Chapter 4 is

composed by the results with a quantitative and qualitative analysis. Chapter 5 discusses the results in a critical point of view, followed by recommendations. Finally, a Chapter 6 concludes the thesis with a wrap-up, limitations and further research topics.

2. Literature Review

2.1. Health systems overview

2.1.1 Health systems

Health systems have been defined as "all the organizations, institutions and resources that are devoted to producing health actions" (WHO, 2000, p.11), undertaking the functions of delivering services, producing resources, financing and preserving. Hence, their objectives include improving the populations' health, fulfilling their expectations and protecting against ill-expenses (WHO, 2000).

Nevertheless, health systems present diverse designs. In this context, OECD typology has been considered widely influential (Burau & Blank, 2006), distinguishing three main models, based on level of coverage, the financing method and the delivery of healthcare (OECD, 1987). Firstly, the National Health Service (NHS), or Beveridge, offers universal coverage, while healthcare delivery is publicly owned and primarily funded by general tax revenues. On the other hand, the Social Health Insurance model (SHI), or Bismarck, is a social security system, where healthcare is delivered by both public and private providers and financed by a non-profit insurance fund, supplied by employers' and employee's contributions. Finally, the Private Health Insurance (PHI) is solely based on private insurance, being characterized by private finance, provision and ownership of facilities.

However, health systems historically associated with the classifications above, namely the U.K. and Sweden with the NHS, are not necessarily pure models, but mostly variations (Burau & Blank, 2006). For instance, changes in the U.K. policy have been eroding the free, state-owned access to health (Propper, 2000). Moreover, critics emphasize the emergence of health systems that fail to be integrated into this 3-model typology, namely the new concept of National Health Insurance (NHI) implemented by South Korea and Taiwan (Lee *et al.*, 2008).

Despite this discussion and the systems' differences, the main challenges are mostly shared.

2.1.2 Health systems' current shared challenges

Population aging and chronic diseases, the rising of costs without the correspondent quality improvement and an uneven access to care, were recognized among 2014's healthcare shared challenges (Deloitte, 2014; My Health London, 2015)

2.1.2.1 Population aging and chronic diseases

Together, the trends of population aging and proliferation of chronic diseases will be the main drivers of healthcare demand growth (Deloitte, 2014).

A combination of three factors has been driving the aging population growth for the last decades. On the one hand, in only half century, the average life span saw a 20-year increase (CDC, 2003), with the number of people reaching 60 years or above, more than tripling in the past 50 years (UN, 2012). On the other hand, fertility rate has been heavily declining, while the post-war "Baby Boom" children will attain above 65 in the period of 2010-2030 (CDC, 2003).

Simultaneously, this age group exhibits greater risk of developing chronic diseases (non-communicable diseases), namely circulatory problems, heart disease and diabetes (Hofmarcher *et al*., 2007). Therefore, population ageing, combined with population growth, is expected to be the catalyst of the increasingly number of deaths by chronic diseases (Abegunde *et al.*, 2007). In fact, Global Burden of Disease Study 2010 (Horton, 2012) concluded that 1.3 million deaths were attributed to diabetes, while WHO (2014a) reported that non-communicable diseases cause more deaths than all other causes combined.

Taking this trend into account, by 2014, chronic diseases were considered one of the main health and development challenges of the 21st century, both due to the human and economic harm (WHO, 2014). The latter accounts for two major factors: the direct cost of care and the morbidity and mortality of labour units (Abegunde *et al.*, 2007). An example would be Liu *et al.* (2002) which determined that coronary heart diseases, in 1999, cost £1.73 billion to the U.K. health system, £2.42 billion in informal care and £2.91 billion in productivity loss.

2.1.2.2 Expenses and Quality

Related to the ageing population and increasing number of chronic disease patients, is the rise of healthcare spending.

In 2011, health expenses reached the $6.9 trillion (WHO, 2014b), absorbing on average 7,4% of the OECD countries' GDP in 2000 and 9,1% in 2011 (OECD, 2015). In this scenario, U.S. leads with 16,3% of GDP dedicated to healthcare (OECD, 2015). In light of this, cutting costs was considered by Mckinsey (2008) the great healthcare challenge of the century.

Nevertheless, this growth did not necessarily lead to greater levels of quality in health, with 20-40% of this spending considered waste (WHO, 2014b). Taking U.S. as an example, despite being the country dedicating more resources to healthcare, it ranks last in overall performance when compared with OECD countries (Davis *et al.*, 2014). Another illustrative example is the estimation that half of the patients do not obtain the necessary care (Asch *et al.*, 2006).

This combined problem reflects the systems' inefficiencies and undermines their sustainability. Although some observers agree that this rise is not a critical issue (Pauly, 1993; Chernew *et al.*, 2003), Bodenheimer (2005) concludes that most researchers argue against, emphasizing the negative effect to employers, employees, governments, and patients.

2.1.2.3 Access to Health

Still today the "inverse care law" (Hart, 1971) is applicable to health systems as they are considered inequitable, offering less access and quality to those who need them more - the poor (Gwatkin *et al.*, 2004). In fact, either rich or poor, no country is reported to have been able to provide immediate access to their population (WHO, 2010), although this phenomenon is more prevalent in low-income countries. As an example, out-of-the-pocket expenses are considerably higher in lower-income countries (WHO, 2012). However, there are also significant differences in access inside high-income countries, with, for example in the U.S., infant mortality rates being more than twice in non-Hispanic blacks than in non-Hispanic whites (CDC, 2013).

In this context, WHO established the progress towards Universal Health Coverage (UHC) as a major priority in the international health agenda. UHC is defined as "all people receiving quality health services that meet their needs without exposing them to financial hardship in paying for them" (WHO, 2013, p.3). Therefore, this view embraces three main dimensions which must be worked on: the population - who is covered by the pooled funds; the services - which services are covered; and the cost - how much cost is covered (WHO, 2010). Based on

the evidence, this movement is expected to lead to improved access to the necessary care and higher levels of population health, especially for the poor (Moreno-Serra & Smith, 2012).

All in all, today, this combination of factors is testing health systems sustainability, that have to deal with the constant trade-off of cost-quality-reach.

2.1.3 Portuguese Health System

The Portuguese health system is not pure, with three parallel systems. These comprise the NHS, private or public sub-systems associated to certain occupations, and private Voluntary Health Insurance. Regarding healthcare delivery, the system is composed by both public and private providers, funded through numerous forms - from historically and activity-based budget to out-of-pocket payments (Barros *et al.*, 2011).

Portugal has made exceptional advances in terms of health, with the life expectancy at birth doubling in the 20th century and the mortality rate being reduced from 55.5 to 3.3 in only 38 years (Barros *et al.*, 2011). However, in the same line of other health systems, the Portuguese system sustainability is being pressured, facing similar challenges.

Firstly, the fertility index has been declining, reaching 1,4 in 2011 (OECD, 2014) and leaving an elderly index of 1,29 (OECD, 2015). This supports the argument that Portugal is facing the same aging population trend. Concurrently, obesity has been increasing (OECD, 2015) and, by 2008, 50% of deaths in Portugal were imputed to either circulatory system's diseases or malignant neoplasms, both chronic diseases (Barros *et al.*, 2011).

In what concerns healthcare costs, despite the mentioned rise, due to the recent debt crisis and consequent austerity measures, Portugal is allocating less financial resources to this sector (Sakellarides *et al.*, 2005). For instance, an 11% reduction in the NHS budget for 2012 was declared (Morgan & Astolfi, 2013). This gains particular importance since the percentage of the spending funded by the government is of 65% (OECD, 2015), which means an immediate rationalisation in state hospitals and health centers.

Simultaneously, in 2014, only 15,6% classified their health status as "very good" and 30% were extremely unsatisfied with the Portuguese Emergencies (IMS, 2015). In fact, concerning quality, several deficiencies in Portuguese health institutions were identified, among which,

the absence of performance indicators for decision-support and an insufficient quality-conscience culture (Mendes, 2012).

On the other hand, the percentage of out-of-pocket payments in the total health expenditure has increased from 24,3% towards 27,3%, ranking as one of the highest (OECD, 2015). This leads to the matter of access to health. Indeed, disparities have been found between both regions and social classes. For instance, in 1999-2003, infant mortality rate (per 1000) was 2.3 points higher in Alentejo region than in Lisbon region (Barros *et al.*, 2011).

Adding to these shared challenges, the Portuguese system exhibits poor levels of governance. Among European countries, Portugal scores extremely low in Government effectiveness and efficiency, policy fit to the level of development and, finally, parallel economy's weight and political influence in decision-making (Sakellarides *et al.*, 2005). This hinders even further the efficiency and effectiveness of the system.

Concluding, in line with the global scenario, ageing population and increased patients of chronic diseases, combined with low efficiency, transparency and access are risking the sustainability of the Portuguese health system.

2.2 Big Data, big opportunities

2.2.1 Big Data

Big Data has been generating excitement, becoming a buzzword worldwide, the "Management revolution" (McAfee & Brynjolfsson, 2012). Nevertheless, much confusion has been created around a fast-evolving definition (Gandomi & Haider, 2015). Most researchers use the idea of "Vs" to grasp the concept, varying between 3 and 5 "Vs". As described below, the data is high "volume" with a "variety" of sources and formats, flowing and analysed at high "velocity", hence generating economic "value", with "veracity".

McAfee & Brynjolfsson (2012) and Gandomi & Haider (2015) attribute three main features to Big Data. "Volume" is related with the high size of the data, which could entail multiple terabyte or petabyte (Gandomi & Haider, 2015). "Variety" is linked not only with the variety of sources and data formats but also to stress the heterogeneity of the structured, semi-

structured and unstructured data. The final "V" corresponds to the "Velocity" at which data is generated and delivered (Russom, 2011).

Later on, IDC included "Value" into the definition, highlighting the economic benefits extracted through Big Data (Gantz & Reinsel, 2012). Finally, "Veracity" stresses the importance of data and source quality (White, 2012).

Overall, Wamba *et al.* (2015, p.2) systematic review sees Big Data "as a holistic approach to manage, process and analyze 5 Vs…in order to create actionable insights for sustained value delivery, measuring performance and establishing competitive advantages.".

2.2.2.2 Healthcare Big Data

Healthcare Big Data presents no standard definition and has been associated with other subjects, namely Electronic Healthcare Records (EHR) and databases' pooling (Velthuis *et al.*, 2013). Nevertheless, the industry is believed to have reached a point at which Big Data may play a major role (Mckinsey, 2013).

Firstly, the healthcare industry is considered to have one of the biggest and fastest growing datasets, in terms of size and extent of coverage (Kambatla *et al.*, 2014). Indeed, in 2011, clinical data alone is estimated to have reached the 150 exabytes, presenting an increase between 1.2-2.4 exabytes per year (Hughes, 2011). This translates into high "Volume" of data.

Secondly, this increased "Volume" has been attributed to the increased "Variety" of health data sources. In other words, the heavy adoption of EHR by care providers (Chen *et al.*, 2012), the development of new medical instruments, patient sensors, in-home care devices, mobile devices and health communities (Kambatla *et al.*, 2014), as well as the emergence of genetic-related data, are feeding the flow of health data (Crown, 2015).

Finally, EHR data is accessible in almost real-time (Crown, 2015), giving in the "Velocity" dimension.

Therefore, the healthcare industry has reached a point where Big Data presents a great potential.

2.2.2 Big Data in answering Healthcare systems' challenges

The application of Big Data in Healthcare can create value in diverse strands. Mckinsey (2013) identifies five main value pathways: right living, right care, right provider, right value and right innovation.

In the first pathway, Big Data will allow improving the consumers' ability to actively promote their well-being, namely through effective targeting of high risk patients for disease prevention. For instance, Asthmapolis improves the self-management of asthma patients, by providing feedback based on data attained from an inhaler-sensor (Feldman *et al*., 2012).

Secondly, Big Data can prove a progress in evidence-based medicine, making sure all providers are able to come up with the best possible treatment, at the right time. For example, Premier, based on compiled data from its 2,700-member network, is able to provide clinical outcomes comparisons, resource utilization and costs information, having, so far, prevented more than 29,000 deaths and reduced expenses in $7 billion (IBM, 2013).

Moreover, Big Data's advancements in resource optimization and performance measurement will enhance the decision-making over the caregiver with the most suitable skills and best-proven outcomes.

The right value pathway is related with the reduction of costs while providing the same or higher levels of quality, maximizing health value. An example is the reduction of waste and abuse by predicting health fraud through claims' real-time analysis (Raghupathi & Raghupathi, 2014).

Finally, Big Data can have a major role in innovation, not only by contributing directly to new advancements but also by expanding the innovation process itself. For instance, clinical trial errors and duration could be reduced by an improved trial method and targeting of the right patients for recruitment (Mckinsey, 2011a).

Overall, by generating value through these pathways, Big Data will enable the evolution towards a learning health system where continuously exchange of feedback between patients and providers will drive treatment optimization (Velthuis *et al*., 2013). Therefore, Big Data can play a major role in responding to the challenges identified in Section 2.1. In fact, were Big Data effectively applied in the U.S., the impact in efficiency and quality could generate a value higher than $300 billion yearly, reducing expenditure in roughly 8% (McKinsey,

2011a). A similar impact would be expected in Europe, with an estimation of $149 billion (Kambatla *et al.*, 2014).

2.3. Big Data in Portugal

2.3.1 Information Systems in Portuguese hospitals

Several efforts have been being undertaken to implement Information Systems (IS) more effectively in the Portuguese health system (Espanha, 2010). In fact, these have already produced positive results, namely in the patients' involvement with the caregiver (Espanha, 2010). An example is the Electronic Prescription that has been widely integrated in Portuguese hospitals at the several levels of care (Portugal. Alto Comissariado da Saúde, 2010). Evermore adopted by caregivers, this tool has been reinforcing patients' security and outcomes by reducing medications' reading errors and by providing clinical support through warning signs (Espanha, 2010).

However, simultaneously, Portuguese hospital's IS have been considered inadequate. In fact, fragilities have been pointed out, such as the inexistence of an integrated datacenter, capable of aggregating all the necessary information (Espanha, 2010). Moreover, Portugal is still lagging behind regarding the usage of data for clinical decision support, namely in the development of guidelines based on scientific evidence for the main disease groups (Sakellarides *et al.*, 2005). Indeed, this is considered a driver in Portuguese's overconsumption of antibiotics and the resulting complications (Portugal. Alto Comissariado da Saúde, 2010).

All in all, IS are ever more important and, despite many efforts, several fragilities have been found in Portuguese hospitals, transforming them in one of today's main problems (Lapão, 2010).

2.3.2 Big Data Overview in Portugal

Overall, Big Data is still in its infancy in Portugal. Nevertheless, this platform grew in 2014 – 9,2% - undertaking an important part of the Information Technology (IT) market (IDC, 2014). Demonstrating its importance in managers' agenda, 24% of the businesses intend to invest in

Big Data & Analytics, being healthcare one of the sectors with highest IT estimated growth for 2015 (IDC, 2014).

At the same time, Start-ups in the area of Big Data have been emerging in Portugal. For example, Feedzai, Stevie Award winner for Innovation, is able to detect financial fraud 30% earlier than previous models, through machine learning methods (Feedzai, 2015). A second example is Vitalidi, which incorporates into daily-life objects sensors capable of measuring and recording Electrocardiographic signals (ECG), applying an off-the-person approach (Vitalidi, 2015).

Illustrative of Big Data's state-of-the-art in Portugal are projects such as "Máquina do Tempo" and VITAL. Firstly, "Máquina do Tempo", developed by SAPO, allows viewers to explore networks and connections between celebrities, based on a 25-year archive of news (SAPO, 2015). Furthermore, VITAL by Centro Hospitalar São João and winner of Microsoft Health Innovation Awards 2014, analyses the hospital's information in order to instantly detect possible patients' anomalies. Partially similar with the Big Data application focused in this work, VITAL allows an earlier and proactive intervention in infections, antibiotics consumption as well as in health deterioration of admitted patients (CHSJ, 2015).

From an academic perspective, Data Science academia is starting to emerge. Firstly, academic research has been conducted in this field, with the International Journal on Multidisciplinary Approaches on Innovation, co-edited by a Portuguese, undertaking a Call for Papers on the topic "Boosting Innovation with Big Data" (JIM, 2015). Moreover, the Lisbon Machine Learning School is already in its 5th edition, covering theory and practice in this field of study (LXMLS, 2015). Illustrating these important first steps in Data Science education, a Portuguese team won the Filtering's and Opinion Mining's awards in the Data Science International Competition of RepLab. (Peixoto, 2013).

2.4. Big Data in early detection of adverse events

Adverse events and complications, such as nosocomial infections and sepsis, cause extensive morbidity and mortality, being extremely expensive to the system. Nevertheless, many are preventable and Big Data can play a major role in predicting these adverse events (Bates *et al.*, 2014), not only through the analysis in real-time of physiological data combined with EHR, but also through knowledge extraction of this stored data (Kohn *et al.*, 2014).

In a hospital, high volumes of physiological data, streaming from multiple sources are generated every second. Indeed, the sequences of vital signs are considered to be multidimensional, extremely connected between each other, with high velocity and non-stationary (Lehman *et al*., 2015). Therefore, as an example, in a Neonatal Intensive Care Unit (NICU), newborns are monitored through several sensors, recording functioning of the heart, respiration rate, neurological function as well as drug and nutrition infusion data. Hence, the amount of data produced per second from these devices is extremely high, with the ECG alone, sampling 1,000 readings in a second – 86,4 million a day for one patient (Khazaei *et al*., 2014).

Concurrently, the utility of vital signs in revealing prognosis of adverse events has been demonstrated, such as in late-onset neonatal sepsis (McGregor *et al*., 2012). In another example, Goldberg (1981) concluded that negligible vital signs variations, such as increased systolic arterial blood pressure and heart rate and pulse pressure, may be an early warning of pneumothorax.

Nevertheless, traditionally, caregivers' diagnoses are mainly based on manual interpretation of physiological data streams (Sow *et al*., 2012). As an example, in a NICU, medical records derive from the hand annotations of vital signs readings, summarizing a 30-60 minute period (Catley *et al*., 2008). On the other hand, alarms triggered when a defined threshold is reached are considered to be based on limited processed data, being its deficiencies widely accepted (Stacey & McGregor, 2007). Hence, it is believed that physiological streams' monitoring is mostly left to a "black box regulatory body approved medical devices" (Khazaei *et al*., 2014, p.225), with all the data being posteriorly deleted (Kohn *et al.*, 2014). Therefore, especially in an ICU, the adverse event is only recognized after the appearance of symptoms or interpretation of an exam, leading to a reactive response (Kohn *et al.*, 2014).

All things considered, despite the high amount of data generated at the point of care and the proved relation between the data and complications, physicians are yet unable to extract relevant information in real-time (Sow *et al*., 2012) – this Big Data approach could drive medical research and improve quality and efficiency (McGregor, 2013).

In this context, Kohn *et al.* (2014) distinguishes between two Big Data applications.

Firstly, systems capable of analysing in real-time structured and unstructured high volumes of constantly flowing data points from vital signs and EHR, will allow the detection of already

known patterns in physiological data, predicting adverse events. For example, a decrease of variability in heart rate, which is normally connected with initial stage of sepsis, will be immediately observed and warned by the system (Kohn *et al.*, 2014).

Nevertheless, these patterns are solely pre-defined by physicians, leading to the second application. There may be rules yet to be discovered that could show the onset of an adverse event - rules still hidden in historical data (Catley *et al*., 2008). In this context, researchers can apply data mining, machine learning and statistical modeling to the combination of stored data, discovering new patterns intrinsic to onset complications (Khazaei *et al*., 2014).

Please notice that these new findings would be fed into the first application, providing new and more accurate algorithms, therefore, forming a continuous learning loop. Hence, this analysis will allow caregivers to base part of their decisions on the past experience of similar patients (Kohn *et al*., 2014).

An illustrative example of both applications, the Artemis project, was deployed in Sick Kids Toronto's NICU. The process starts with the data acquisition element which continuously collects physiological data streams as well as complementary clinical information. This data is then input into an online analysis component which is able to process it in real-time and output early warning signs into the caregivers' interface. However, simultaneously, both the data collected and the analytics' results are stored in a particular component, from which data will be mined and knowledge extracted for clinical research support. Finally, these new validated algorithms are redeployed into the online analysis, improving its efficacy with the number of monitored patients (McGregor, 2013). This application has been allowing the hospital to predict signs of infections 24 hours earlier than traditional methods (IBM, 2013).

All in all, this multiple, high-speed vital signs flowing from numerous patients in several places, combined with EHR, is a still untapped Big Data problem which could enormously prevent morbidity and mortality by prematurely detect onset conditions (Khazaei *et al*., 2014). Indeed, it has been considered as disruptive as genomics research (McGregor, 2013), being able to drive healthcare efficiency and dramatically enhance patient outcome as well as reduce their length of stay (Blount *et al*., 2010).

2.5. Challenges in implementing Big Data projects

Many promises for Big Data were already identified; yet, many challenges are also effective barriers to its successful implementation. Particularly in healthcare industry, data privacy and security are extremely important (Feldman *et al.*, 2012). Moreover, cultural and organizational inertia (McAfee & Brynjolfsson, 2012), as well as skilled labor constraints (Chen *et al.*, 2012) are also considered potential barriers. Finally, the initial high investment combined with a still undiscovered business value may be an added difficulty.

2.5.1 Health data Privacy and Security

In this industry, data privacy reaches various dimensions, from political and legal to individual and cultural. Therefore, Big Data raises many concerns in this matter.

Firstly, the government provides an intense net of regulations regarding data privacy, which must be complied with. These include the EU Directive 95/46/CE in Europe and the HIPAA privacy rule in the U.S. In this context, there is also a traditional, cultural and legal agreement of doctor-patient confidentiality (Feldman *et al.*, 2012).

Moreover, there are personal concerns regarding data disclosure to third parties, such as insurers and employers, which may have a conflict interest in acquiring such data, for supporting decisions like insurance pricing and recruitment (Feldman *et al*., 2012).

On the other hand, Big Data is often associated with more data collection and pooling and even cloud computing for storage and analysis. Hence, as architectures become increasingly integrated, so will the risk to data security (Mckinsey, 2011b). Today, security is a real concern, with companies fearing unintentional leak of data to non-authorized entities (Feldman *et al.*, 2012).

2.5.2 Cultural and Organizational changes

Big Data adoption implies structure and cultural changes - tremendous managerial challenges.

Firstly, in this new era, decision rights and information should be placed in the same dimension (McAfee & Brynjolfsson, 2012) with business objectives and technology capabilities being understood together. Besides, company culture should be shifted from the traditional decision-making based only on experience and intuition (McAfee & Brynjolfsson,

2012). In an ICU environment, this would translate into a learning system where a multidisciplinary team of caregivers and engineers would collaborate (Celi *et al.*, 2013).

Therefore, effectively manage this change will be critical in order to fully appropriate Big Data's benefits (McAfee & Brynjolfsson, 2012). In fact, changing the mindset of the employees in order to embrace and learn the new system has already been considered the biggest challenge in a Big Data project (Dutta & Rose, 2015).

2.5.3 Skilled labour shortage

"Data Scientists" is the new term for those who work with Big Data, demonstrating capabilities not only in data analysis but also associative thinking and creative IT (Davenport *et al.*, 2012). Nevertheless, there is a real shortage of such experts. In fact, by 2012, no university offered a Data Science program (Davenport & Patil, 2012) and Mckinsey (2011a) estimates that demand exceeds supply by 50-60% in the U.S. alone.

Therefore, as the market for their services becomes more competitive, "Data Scientists" become ever more difficult and expensive to hire, as well as to retain (Davenport & Patil, 2012).

2.5.4 High initial investment and unclear benefits

Due to its underlying characteristics, such as heterogeneity and volume, Big Data raises numerous challenges throughout the Big Data Analysis Pipeline (Agrawal *et al.*, 2012) (Annex A for the description), requiring new methods and technologies, such as, the development of storage systems capable of housing extremely large datasets (Kambatla *et al.*, 2014). Additionally, as mentioned above, it involves a company-wide integration and transformation. Hence, a high investment to implement such projects is required.

However, at the same time, as concluded in Section 2.2.1, Big Data is an extremely new concept, making business cases and quantitative evidence still absent. Moreover, for a company, it is tremendously hard to understand in advance the value of the data, as it is necessary to, beforehand, gather it and explore its potential opportunities (Zillner *et al.*, 2014).

3. Methodology

3.1. Research Focus

Big Data has many possible applications, included in the five different value pathways (Mckinsey, 2013), which are expected to face diverse barriers in implementation. This requires the thesis to focus on a single application: Big Data in early detection of adverse events. Moreover, due to relevance-seeking and distance-constraints, Portugal was chosen as a geographical focus.

3.2. Hypotheses

Based on the Literature Review, Hypotheses will be formulated for the specific case of Big Data for early detection of adverse events in Portuguese hospitals.

3.2.1 Hypothesis 1 – Data Security and Privacy

In Big Data projects for early detection of adverse events, it is believed that anonymizing streaming physiological data is essential, for security and privacy reasons (McGregor, 2013). Indeed, security and privacy are true concerns among Portuguese hospitals' CIOs, which are seeking to implement measures to improve information protection (Gomes & Lapão, 2008).

Nevertheless, most Portuguese firms have reported security breaches (IDC, 2015). Specifically for the NHS, Araújo *et al*. (2007) points out several security vulnerabilities in the IS management, including, among others, the absence of appropriate security policies and procedures.

Moreover, European Union Data Protection Directive, related with personal data protection, has been facing a major reform (European Commission, 2012). Although this is expected to adapt regulations to technological advancements (European Commission, 2012), it will impact information storage and governance expenses (IDC, 2015) and increase compliance concerns.

All things considered, security and privacy are critical issues in these specific projects and a general concern among Portuguese hospital CIOs; nevertheless, security fragilities have been found in the Portuguese realm. This, allied with complex regulation, will create higher

compliance concerns and costs. Therefore, security and privacy are expected to be effective barriers in the implementation.

H1: Security and Privacy are a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

3.2.2 Hypothesis 2 - Shortage of Skilled Labor

In Big Data projects for early detection of complications, a multidisciplinary team is required, with "Data Scientists" being a critical element. For example, in Artemis project, the team included the Hospital, University of Ontario Institute of Technology and IBM Canada (McGregor, 2013).

Nevertheless, lack of qualified Human Resources is considered one of Portuguese hospitals IS' main issue (Lapão, 2007), with roughly 50% of the workers not holding a degree (Lapão, 2005). In fact, illustrating this issue, in a study analysing the adoption of IT and IS in two Portuguese hospitals, it was concluded that the level of skilled personnel was insufficient to implement IT (Martinho *et al*., 2014).

All in all, this kind of projects requires multidisciplinary teams, with "Data Scientists" being an important element. However, Portuguese hospitals already struggle with unqualified IT personnel. Therefore, lack of skilled labour is expected to be an effective barrier.

H2: Lack of IT skilled labour is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

3.2.3 Hypothesis 3 – Cultural and Organizational inertia

In Portugal, IT is still perceived as a merely efficiency-driver, with only 30% of IT budget being dedicated to innovation rather than operations (IDC, 2015). Concurrently, in hospitals this trend is yet to be reversed, with organizational rigidity being one of the main deterrent factors in hospital innovation (Martinho *et al*., 2014).

Still, this inflexibility is also realized at the caregiver's level. In fact, one of the key obstacles in implementing telemedicine in Portugal was caregivers' lack of adoption and resistance to change (Alvares *et al.*, 2004). Nevertheless, younger generation is demonstrating further adoption towards EHR, which may imply a short-termed rigidity (Tomé *et al*., 2008).

Overall, Big Data requires a data-driven mindset, which is still embryonic in Portugal. At the same time, profound changes at the organizational level are necessary and this country has reported rigid managers and caregivers. Therefore, cultural resistance is expected to be a challenge.

H3: Cultural and organizational rigidity is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

3.2.4 Hypothesis 4 – High investment and uncertain value

As concluded in the Literature Review, the investment required in implementing Big Data projects is high and, aiming for a balanced public deficit, the Portuguese Government continues its efforts in cutting healthcare costs.

At the same time, early detection of adverse events is no exception to the generalized lack of relevant evidence in Big Data projects' benefits. In fact, case studies already mentioned, like the Artemis, are still too embryonic to report costs savings and quantitative benefits (IBM, 2013).

Therefore, budget constraints combined with uncertain quantitative benefits, may hamper the management approval and, hence, the adoption.

H4: Budget constraints and undiscovered business value is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

3.2.5 Hypothesis 5 – Absence of Knowledge on Big Data

Although it is not mentioned in the Section 2.5, from the conclusions taken in the Sections "Big Data" and "Big Data in Portugal" it is possible to infer a 5th Hypothesis.

Indeed, Big Data is still a new and ill-defined concept. In fact, in 2008, Wamba *et al.* (2015) review could only identify 1 article related to Big Data, although this trend rose to 44 in 2012. At the same time, it was concluded that Portugal is still giving its first steps in this subject, leading to the possibility that Portuguese hospital's managers and caregivers are unaware of such an innovation.

Therefore, uninformed managers and caregivers in Portuguese hospitals may be an impediment.

H5: The absence of knowledge regarding Big Data and its potential benefits is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.

3.3 Data Collection

In order to test the Hypotheses, primary data was collected through surveys and person-to-person semi-structured interviews.

3.3.1 Online Surveys

Primary quantitative data was collected through surveys.

Firstly, several studies analysing the barriers to the adoption of new technologies, such as EHR (Gans *et al*., 2005; Shields *et al*., 2007) and the Internet (Walczuch *et al.*, 2000), as well as those analysing the reasons for companies not being more data-driven (LaValle *et al*., 2010) have collected data through surveys.

Moreover, advantages inherent to online surveys, such as speed and the possibility to easily acquire higher quantities of data and download it (Evans & Mathur, 2005), were taken into account. Besides, for caregivers, anonymity revealed to be important.

Nevertheless, surveys also present disadvantages and efforts were made to mitigate them. Particularly, uninformed on the concept of Big Data, respondents, without the opportunity to clarify questions, could have difficulties in answering and understanding what kind of decisions they would make. This risk was mitigated in two ways: (1) describe a possible definition of Big Data, with an example and a commented scheme, in a section of the survey and (2) to analyse the survey with two caregivers, incorporating their feedback into the text.

Therefore, surveys were distributed both to caregivers – doctors and nurses - and managers of Portuguese hospitals. These were spread in an online format, through email, personal contact with the hospital and social networks.

The caregivers' survey had an estimated completion time of 5 minutes, being mainly multiple-choice with only an open-ended question. The managers' survey followed a similar

framework, taking an estimated time of 7 minutes. All questions were formulated in Portuguese (refer to Annex B for both surveys).

The surveys were opened from 7th April until 23rd May 2015. During this period 89 answers from caregivers (refer to Annex C for demographics) and 7 answers from managers were obtained.

3.3.2 Interviews

As a qualitative complement, interviews were undertaken.

Interviews have already been utilized to enrich the assessment of Big Data's challenges (Feldman *et al*., 2012). Besides, this method provides more complete answers and depth of the information (Harrell & Bradley, 2009). Finally, it excludes the surveys' main problem, as the interviewer may explain the concept of Big Data and clarify misunderstandings.

However, this tool may have a courtesy bias associated, which might lead the interviewee to answer what is socially acceptable, for example in which concerns data privacy. Time requirements may also be considered a disadvantage (Opdenakker, 2006).

Hence, the interviews followed a semi-structured format, being all person-to-person. Managers, caregivers and Project Managers of Big Data projects in Portuguese hospitals were the main target. These were conducted from 21st April until 13th May 2015.

Overall, 4 interviews were made to managers, 3 from different hospitals – 2 private and 1 public. A Project Manager of a Big Data project from one of these hospitals was also interviewed. Besides, doctors from still 2 different hospitals – 1 private and 1 public - were interviewed, being one of them part of the Infectious Commission. Hence, the interviews' sample included 7 individuals from 5 different hospitals.

Summing both surveys and interviews the sample included 91 caregivers – 0,08% of the population – and 11 managers – 4,4% of the population (INE, 2013).

3.3.3 Applying to Hypotheses

3.3.3.1. Hypothesis 1

Security and privacy concerns were tested through managers' surveys and interviews. The surveys included a multiple-choice question, allowing respondents to check the ones they considered barriers, including a text field.

On interviews, the emphasis was on how important data security and privacy was to the hospital and if the manager considered these projects to jeopardize this priority.

Caregivers' survey also included an option related to security and privacy in the reasons not to adopt such technology in their daily practice.

3.3.3.2. Hypothesis 2

The shortage of "Data Scientists" as a barrier was tested through managers' surveys and interviews. Firstly, it was important to assess whether the hospital had accessed to such skilled personnel and if hiring was a priority using the Net Promoter Score from Qualtrics scale, 0-10. In an interview, further discussion regarding this shortage as a barrier was performed.

3.3.3.3 Hypothesis 3

Cultural and organizational rigidity was extensively verified through both managers' and caregivers' surveys and interviews.

From the caregiver point of view three main aspects were covered.

The first was directly related with the Big Data application. The objective was to understand the perception of (1) whether they would consider the tool useful – using 0-10 scale - and (2) whether they would adopt it in their daily practice. Main reasons for resisting the adoption were verified in a multiple choice question, including a text entry.

The second dimension was related with historical behavior, complementing the previous prediction. Hence, it was crucial to understand if the caregivers have been adopting and consider useful the technology implemented by the hospital – 0-10 scale. The adoption of the last technology employed was also added in order to understand the type of tools caregivers were working with.

Finally, on an organizational perspective, it was critical to ask whether the caregivers believed in a paradigm change towards data-driven medical practice – 0-10 scale.

The interviews covered all the aspects in an open-ended manner.

From the managers' point of view the discourse was divided into organizational challenges and physicians' resistance.

Firstly, it was essential to comprehend if the institution had the required organizational aspects. That is, whether or not it had an IS department and whether this was merely operational or central to the institutions' strategy – scale 0-10. On the other hand, it was necessary to understand if the manager believed in a paradigm shift with organizational change – 0-10 scale. In an interview environment, barriers to this organizational transformation were also discussed.

A cultural inertia from the caregivers was also analysed based on the managers' experience and perception of their adoption rate.

3.3.3.4 Hypothesis 4

This evaluation was based on managers' surveys and interviews. Firstly, it was important to understand whether a cost-benefit analysis was made – 0-10 scale – and how relevant it was in the decision process – 0-10 scale. Finally, a series of options as barriers to the implementation regarding this issue were given in a multiple choice.

In an interview environment, examples of this year's projects and whether the absence of a Return on Investment (ROI) could hinder a project's approval were discussed.

3.3.3.5 Hypothesis 5

This was evaluated both at the management and caregiver level. The survey questions were identical, with an initial approach on the level of knowledge in "Big Data" and its advantages – scale 0-10. If the answer was not 0, two questions would evaluate the respondent's true knowledge (1) a true or false, multiple choice question and (2) a question on the comprehension of the application of predicting adverse events – scale 0-10.

In the specific case of interviews, it was often asked if the hospital was implementing any Big Data project.

Finally, in an overview, managers and all interviewees were asked about the three main barriers to the implementation. In the surveys, this was in the form of a 3-restricted multiple-choice with several options and a text parameter.

3.3.4 Analysis

The Hypotheses' testing was divided in a quantitative and qualitative part.

3.3.4.1 Quantitative analysis

The quantitative part was based on the surveys' results, analysed in four ways in Excel 2010.

The first was to calculate the percentage of responses from one or other option, for example, the percentage of those who would adopt the application.

The second, was related with all the rating (0-10) questions, in which a One Sample t-test was performed, comparing with the test value 6. If the rating was statistically significant higher than 6, it would be considered a high rating, and the same reasoning when significantly lower.

A third analysis was executed when comparing the results of subgroups, using a Two Sample t-test with different variances, for example, to compare the mean ratings of those who would and would not adopt the application.

Finally, a statistical descriptive summary was made, with Standard-Deviation (SD), Upper and Lower Limits, using a confidence level of 95%.

3.4.1 Qualitative analysis

This analysis was solely based on the interviews. To preserve anonymity, the hospitals were named by letters – A, B, C, D and E (refer to Annex D for a brief description of each hospital). In this context, in Hospital A two managers were interviewed as well as the Project Manager of a Big Data project. In Hospital B and C a manager was interviewed and in D and E a caregiver was interviewed.

4. Results

4.1. Overview

Table 1: Hypotheses Overview

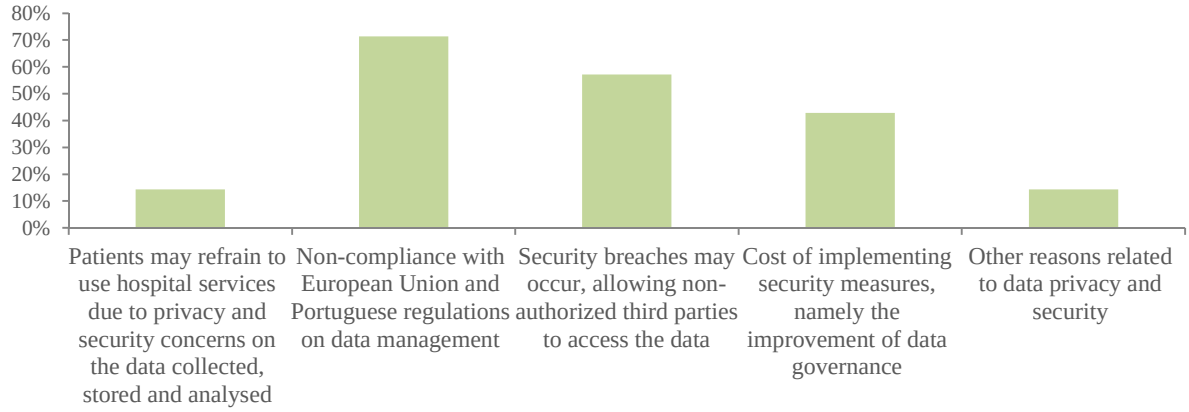
	Hypothesis	Valid?
H1	Security and Privacy are a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.	**Not valid**
H2	Lack of IT skilled labour is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.	**Valid**
H3	Cultural and organizational rigidity is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.	**Not Valid**
H4	Budget constraints and undiscovered business value is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.	**Valid**
H5	The absence of knowledge regarding Big Data and its potential benefits is a barrier to the implementation of Big Data projects for early detection of adverse events in Portuguese hospitals.	**Valid**

4.2. Hypotheses testing

4.2.1 Hypothesis 1- Data Privacy and Security

Data security and privacy were reflected in the managers' surveys as the lowest barrier, emerging only once as a top 3 barrier. Nevertheless, the non-compliance with data management regulations and eventual security breaches were considered the main concerns in this field.

Graphic 1: Main Security and Privacy Concerns from managers' surveys



On the caregivers' perspective, data privacy and security were named in the surveys as the main driver to resist acceptance (39%).

On a qualitative outlook, in the interviews with managers, data security and privacy was not considered a relevant barrier, but rather a requirement of the tool and technology used. However, from the Project Manager point of view, it could become a true issue if the data was shared between hospitals.

Firstly, Hospital C's manager considered it was not something particularly significant. In fact, he mentioned a trade-off between patients' privacy and security, meaning that it was more important that a physician had accessed to all the patients' data, than to misdiagnose due to restrained access to information.

Moreover, in Hospital A, the manager did not see data privacy and security as a barrier to the implementation, but more as a requirement in the tool – it had to guarantee confidentiality. Indeed, in what concerns its Big Data project, this was not a barrier. Still, several steps were taken in order to guarantee privacy and security, such as anonymizing flows and ensure that the access-rules were exactly the same. However, it was believed that data privacy and security issues could rise if (1) the data was collected and shared through different hospitals or (2) the application was sold.

All in all, data privacy and security are not perceived as a true impediment to managers, although it could raise some issues at a larger scale. Therefore **Hypothesis 1, from a providers' point of view is not valid**. However, it is necessary to refer that from a regulatory institution perspective, this may be a barrier – this would require further research.

4.2.2 Hypothesis 2 - Shortage of Skilled Labor

Shortage of "Data Scientists" has been qualified by the managers' surveys as the main barrier, with 86% of the respondents including it in the top 3.

Table 2: Hospital provided with skilled labour and priority to hire. One Sample t-test, Test Value=6

	N	Mean	SD	UL	LL
Provided with Data Scientists?	7	3,71*	2,81	6,31	1,11
Priority to hire?	7	3,14***	1,68	4,69	1,59

****P*<0.01, ***P*<0.05, **P*<0.10

At the same time, it was concluded that hospitals were not provided with "Data Scientists" nor was their priority to hire them – means 3,71 and 3,14 respectively, significantly different from 6. Hence, from the surveys, it is possible to conclude a shortage of this high-qualified workforce in Portuguese hospitals.

On a more qualitative overview, in manager's interviews, the results were consistent.

Hospital A's manager considered this to be the main barrier to the implementation. In fact, the hospital had only one "Data Scientist", without whom the Big Data project would have been impossible. Hence, the manager saw the shortage of skilled employees as a deterrent in the projects' extension to other areas.

On the other hand, Hospital B considered a generation gap as the third main barrier to implementation. In this manager's point of view, it was not the shortage of people with know-how, but rather the shortage of people which combined experience and know-how that was an issue.

Quantitative and qualitative conclusions point to a high scarcity of "Data Scientists" in Portuguese hospitals, which is perceived as an obstacle to implement this kind of projects. Therefore, **Hypothesis 2 is valid**.

4.2.3 Hypothesis 3 – Cultural and Organizational inertia

Cultural and organizational rigidity have been considered in managers' surveys as the third most relevant barrier, with 57% including it in the top 3 barriers.

From an organizational point of view, all the surveyed hospitals had an IS department, which was believed to be crucial to the strategy and growth – mean of 7. Nevertheless, combining this information with the shortage of "Data Scientists", it is possible to conclude that these departments are not consistent with the organizational structure mentioned in 2.5.2. Therefore, it is not possible to determine that Portuguese hospitals have the organizational structure normally associated with Big Data projects.

On the perspective of organizational change, managers and caregivers were keen to believe that one had to occur – mean 7,29 and 7,25 respectively, significantly different from 6. This indicates a positive mindset towards an organization where "Data Scientists" and caregivers work together.

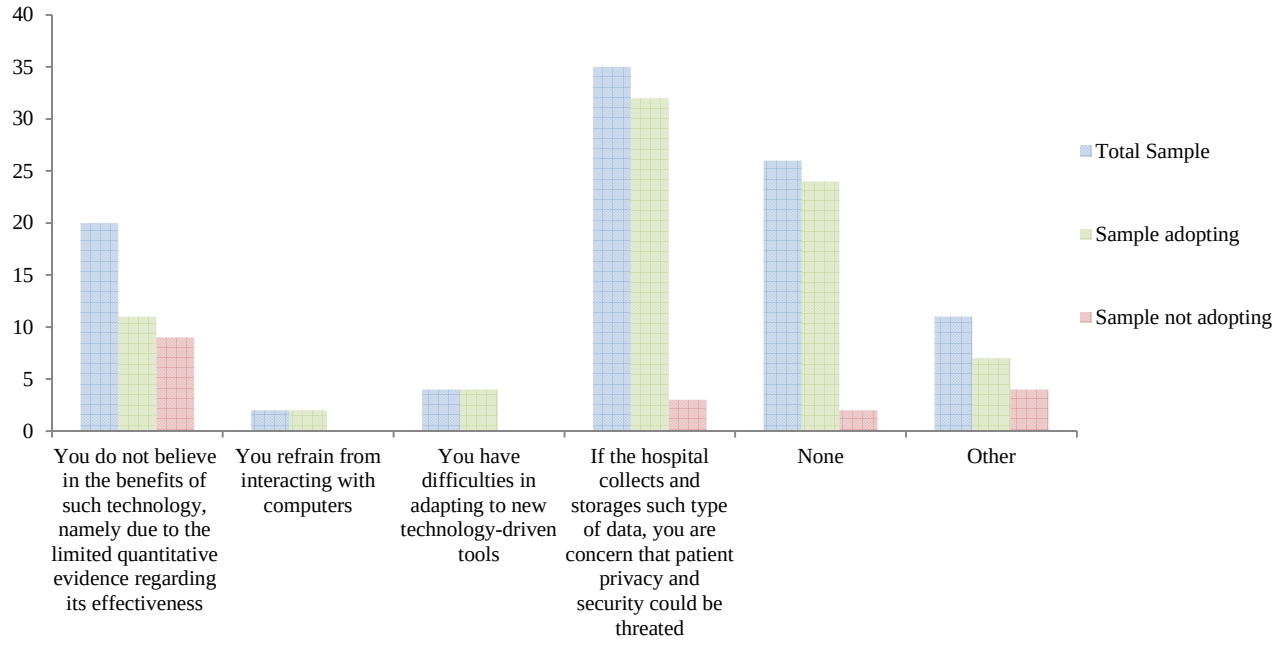
Table 3: Caregivers' and Managers' believe in data-driven medicine and organizational change. One Sample t-test, Test Value=6

	N	Mean	SD	UL	LL
Managers	7	7,29***	0,76	7,98	6,59
Caregivers	89	7,25***	1,99	7,67	6,83

****P*<0.01, ***P*<0.05, **P*<0.10

At the same time, all managers agreed that caregivers would adopt this tool. Moreover, when asked directly to caregivers, they considered the application as useful – 6,55 significantly different from 6. Furthermore, only 20% would not adopt such tool in their daily practice.

Graphic 2: Reasons to resist the application adoption



The central reason for caregivers to refrain from adopting would be patients' privacy and security concerns. However, to those who would not adopt, the cause was related with the lack of evidence regarding the technology's benefits. Moreover, a deficiency of resources, high costs and incompatible information systems were mentioned in the "other" field.

Table 4: Caregivers' believe that application is helpful and in data-driven medicine and organizational change. One Sample t-test, Test Value=6

	Whole Sample					Sample Adopting					Sample Not Adopting				
	N	Mean	SD	UL	LL	*N*	Mean	SD	UL	LL	*N*	Mean	SD	UL	LL
Application as helpful	89	6,55**	2,37	7,05	6,05	71	7,28***	1,798	7,71	6,86	18	3,67***	2,196	4,76	2,57
Believe in data-driven medicine	89	7,25***	1,99	7,67	6,83	71	7,55***	1,81	7,98	7,12	18	6,06	2,26	7,18	4,93

****P*<0.01, ***P*<0.05, **P*<0.10

In fact, those who would not adopt, considered it statistically significant (Two Sample t-test with different variances was performed with a P=1,38^{-06}) less helpful than those who would, and even not useful – mean 3,67, significantly different from 6. Besides, this group also believed significantly less (P=0,02) in a paradigm change regarding medicine and organizational structure.

Finally, on an historical point of view, caregivers have been adopting and consider advantageous the technology implemented by the hospital, with a mean of 7,18, significantly different from 6.

Table 5: Caregivers have been adopting and consider advantageous the technology implemented by the hospital. One Sample t-test, Test Value=6

	Whole Sample					Sample Adopting					Sample Not Adopting				
	N	Mean	SD	UL	LL	*N*	Mean	SD	UL	LL	*N*	Mean	SD	UL	LL
Have been adopting the technology implemented	89	7,18***	2,43	7,69	6,67	71	7,28***	2,19	7,80	6,76	18	6,78	3,26	8,40	5,15

****P*<0.01, ***P*<0.05, **P*<0.10

Furthermore, when comparing those who would and would not adopt, although the difference was not significant (P=0,54), the first group has been accepting more easily. This proximity may be explained by the fact that most of the implemented technologies show complete evidence of the benefits, namely, electronic or online clinical registration and prescription.

On a qualitative perspective, the conclusions were diverse. On the one hand, organizational structures were extremely different, although all managers agreed with a paradigm and organizational change, with no cultural barriers at the management level. On the other hand, the opinions diverged in caregivers' resistance to adoption.

Firstly, Hospital C did not hold an official IS department, being an organizational change combined with a nonexistence experience in implementing an IS area, considered a top 3 barrier. Indeed, this kind of projects did not fit with the hospital strategy of adopting only golden standards – "we are not and do not wish to be an investigation center". On the other side, Hospital A, which aims to be the most advanced hospital at the country level, and even continent, had two distinctive departments: the Informatics, which was maintenance-related, and the Development, which was multidisciplinary and responsible for the Big Data project.

Still, both managers believed in a paradigm and organizational shift. In fact, a manager from Hospital A assumed that, in the future, the hospital would invest in a team of "Data Scientists" collaborating with caregivers and managers. Moreover, another manager from this hospital confirmed that there would be no resistance to change at the management level, as long as it brought more efficiency and effectiveness. Indeed, the Project Manager explained that there was a cultural transformation inside the hospital in order to implement the project, namely regarding communication between managers and the different care units.

Concurrently, the manager from Hospital C agreed that there would be no cultural barriers at the management level, although it would be extremely difficult to have a unanimous acceptance of the tool, expecting resistance by the caregivers – "most doctors are resistant to change and accommodated to a style, there is much inertia".

In fact, a doctor from Hospital D agreed with this latter vision. In his point of view, doctors tend to accommodate to their work model. Moreover, he believed doctors see themselves as liberal workers, which obstructs teamwork, flexibility and feedback exchange. In fact, care unit directors are in charge for long periods of time, thus creating vice. Besides, a manager from Hospital A agreed that a possible relevant barrier could be the resistance from caregivers, who could, initially, distrust the technology results.

However, Hospital B was much in line with the managers' survey results, disagreeing with caregivers' resistance. In her opinion, doctors are scientists, hence educated based on the scientific method and with difficulties in everything that is subjective – "90% of their

decisions is based on data". Hence, as Big Data provides more numbers and mathematical reasoning, it will be welcomed.

Going a step further, the doctor from Hospital D related the resistance to adoption with the knowledge-acquisition process, which differs among generations. This was reemphasized by the doctor from Hospital E, which considered that resistance was only visible in older physicians. Analysing the surveys' results, there was a difference between the mean age of those that would and would not adopt – 39 against 43 – although this difference was not significant (P=0,12).

All things considered, most of the Portuguese hospitals do not have an organizational structure required to apply this kind of projects. However, at the management level it is not expected a cultural resistance to change. Moreover, although opinions diverge, balancing the results, it is possible to conclude that caregivers would not resist to such a change and to adopt this innovation. Therefore, **Hypothesis 3 is not valid.**

4.2.4 Hypothesis 4 – High investment and uncertain value

High initial investment allied with an undiscovered business value was considered in the managers' surveys as the second most relevant barrier, with 71% including it in their top 3.

Table 6: Costs-Benefit analysis execution and relevance in project decision. One Sample t-test, Test Value=6

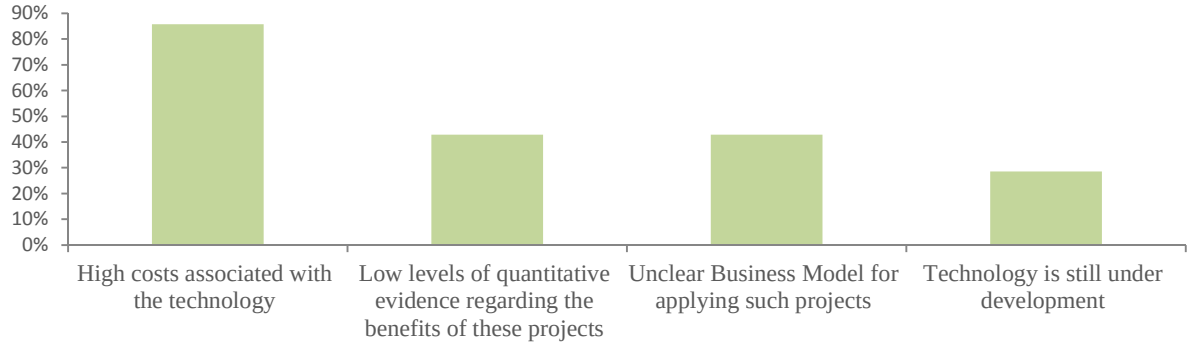
	N	Mean	SD	UL	LL
Cost-Benefit analysis	7	7,57*	1,90	9,33	5,81
Revelance	7	7,86***	1,07	8,85	6,87

****P*<0.01, ***P*<0.05, **P*<0.10

This is a reflection of (1) an often performance of cost-benefit analysis – 7,57 and (2) its relevance in the decision-making process – 7,86. In other words, if a cost-benefit analysis is extremely important, then it is expected that high investments allied with uncertain benefits become a barrier.

In a more detailed analysis, high costs associated with the technology are the main component of this barrier. This emphasizes the mentioned Portuguese budget constraints.

Graphic 3: Main barriers with high investment and lower evidence from managers' surveys



In a qualitative outlook, this is perceived as an issue, although opinions split the definition: for ones absence of evidence is more important and for others are the costs.

Firstly, Hospital C's main barrier was the absence of evidence – this is not a golden standard in medicine and it is not the hospital's objective to invest in high-risk projects, "Start-ups". In fact, cost was not considered an issue, since a relevant percentage of the budget is dedicated towards new technology, as the hospital continuously renovates its infrastructure.

On the other hand, manager from Hospital B considered the cost as a true impediment, being this project a medium-long term investment. In other words, this spending would leverage the business, as it improves quality, generates satisfaction and loyalty. Hence, cost was the only barrier.

In the same line, in Hospital A, the absence of a ROI was not an issue, as a return was evident, namely through improved care, cost savings and even the sale of the project. Moreover, it is necessary to mention that the project was funded by both the hospital and European Union – not many Portuguese hospitals would have this kind of financial power.

Quantitative and qualitative evidence is consistent with budget constraints combined with undiscovered benefits being a barrier to implementation. Hence, **Hypothesis 4 is valid**.

4.2.5 Hypothesis 5 – Absence of Knowledge on Big Data and its applications

In the managers' surveys this was positioned as the 4th most relevant barrier, with 43% placing it among the top 3.

In what concerns the managers' knowledge on the concept, the mean was of 3,86, although it was not significantly different from 6, explained by a high SD.

Table 7: Managers' knowledge on the concept of Big Data. One Sample t-test, Test Value=6

	N	Mean	SD	UL	LL
Overall Knowledge	7	3,86	3,72	7,29	0,42
Overall Knowledge from those different from 0	4	6,75	1,26	8,75	4,75
Knowledge from those different from 0 regarding the specific application	4	5,50	2,89	10	1

****P*<0.01, ***P*<0.05, **P*<0.10

In this context, 57% reported some knowledge, having a mean understanding of 6,75 – also not significantly different from 6. Nevertheless, only 1 manager answered the true or false correctly, reflecting that the previous results are self-reported and the knowledge may be, in fact, lower. Hence, in the overall sample, only 14% actually had full comprehension on Big Data.

Moreover, on the specific application focused in this thesis, the mean knowledge was of 5,50, which is slightly lower from the knowledge regarding Big Data (P=0,47).

On the caregivers' side, the overall understanding of Big Data was extremely small - 1,04, significantly lower than 6. Moreover, only 28% rated some level of knowledge, although it was significantly low - 3,72. Besides, merely 4 caregivers responded correctly the true or false, which corresponds to 4% of the caregivers having profound knowledge on Big Data.

Table 8: Caregivers' knowledge on the concept of Big Data. One Sample t-test, Test Value=6

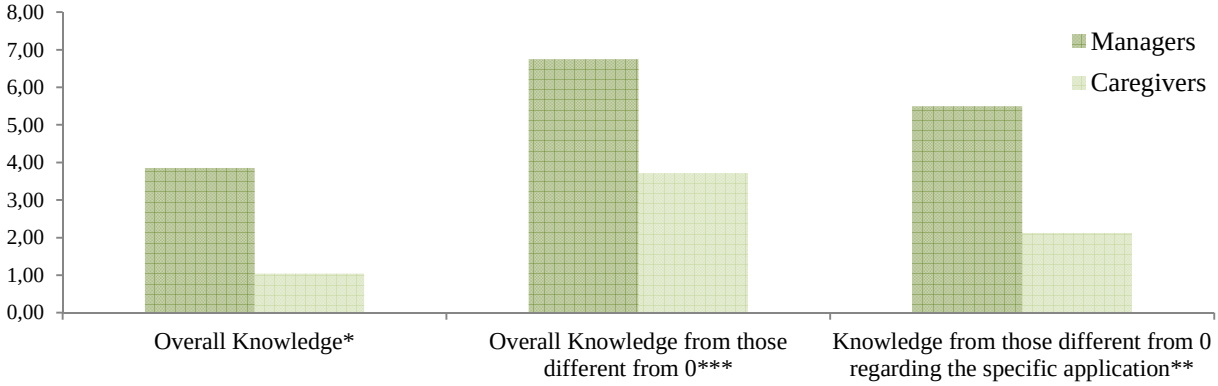
	N	Mean	SD	UL	LL
Overall Knowledge	89	1,04***	2,11	1,49	0,6
Overall Knowledge from those different from 0	25	3,72***	2,44	4,73	2,71
Knowledge from those different from 0 regarding the specific application	25	2,12***	2,297	3,07	1,17

****P*<0.01, ***P*<0.05, **P*<0.10

On the specific application of adverse events' detection, the comprehension was also reported as being low - 2,12 - even statistically significant smaller than Big Data as a whole (P=0,02).

Finally, it is possible to determine a significant knowledge gap between managers and caregivers. As illustrated in Graphic 4, managers demonstrate significantly more understanding on the concept of Big Data and in its specific application for prevision of adverse events.

Graphic 4: Managers vs. Caregivers knowledge on Big Data. Two Sample t-test with different variances ***P<0.01, **P<0.05, *P<0.10



The qualitative research was in line with these results. On the one side, Hospital C manager had no idea regarding Big Data, while the other two Hospitals were not only fully aware of the concept, but also already implementing Big Data. More specifically, Hospital B was applying Big Data in Marketing while Hospital A was using data for earlier and proactive intervention in infections, antibiotics consumption and in health deterioration of admitted patients.

All in all, the ratings regarding knowledge were significantly low for caregivers and medium, although extremely variant, for managers, existing an important gap between the two classes. Therefore, **Hypothesis 5 is valid**.

5. Discussion and Recommendations

5.1. Discussing Results

Knowledge regarding Big Data and its applications is significantly low at the caregiver level and medium at the management level, existing a gap between the two. This condition may have many implications. Firstly, if managers and caregivers are completely uninformed they would not even consider implementing such projects. Secondly, with little knowledge, managers would present difficulties in understanding how Big Data can be applied and improve the various areas of the hospital. In fact, this "lack of understating of how to use analytics to improve the business" was considered by LaValle *et al*. (2010) the main barrier for firms to become more data-driven. Finally, this information gap may hinder caregivers' adoption of the tool. Meaning that if the project is implemented but caregivers do not understand its benefits and how it works, then, distrusting the results, they might refrain from using it.

At the same time, Portuguese hospitals do not hold an organizational structure prone to Big Data projects. As discussed, although managers consider significantly relevant the hospital's IS departments in its strategy, the absence of "Data Scientists" and low priority to hire them, lead to believe that the organization is not ready for a multidisciplinary environment, where caregivers and "Data Scientists" work together. Therefore, an organizational change would be required.

Intrinsic to this transformation, the results demonstrated that managers' resistance would not be a barrier. Nevertheless, this finding is not consistent with the literature, revised in Section 2.5.2. This discrepancy may be explained by the fact that the hospitals managers are not the "HiPPO - the highest-paid person's opinion" (McAfee & Brynjolfsson, 2012), but rather the caregivers. In other words, this particular tool would not transform the way managers decide, but the caregivers'. Hence, resistance would be expected at the caregiver level.

However, whether caregivers would resist was rather controversial, although a final conclusion pointed for a general adoption. This may have several explanations. Firstly, this application would enhance the caregivers' information with quantitative data, which is intrinsic to their education on the scientific method. Secondly, the tool is intended to be personalized to the caregivers needs, improving adoption rate. Moreover, by answering the survey, the caregivers are already providing their input into the project, thus enhancing their

openness. Besides, those responding to an online survey may be more technology-friendly and the mean age was fairly low – 39.

On another point, high investment allied with undiscovered business value was considered a valid barrier. Nevertheless, there was an interesting split within this Hypothesis, with some hospitals concerned with costs while others with the lack of proved benefits. Combining this information with a high SD on the managers' knowledge on Big Data, there might be a relation between the hospital's strategy and the specific barriers it endures. For example, hospitals who are innovators or early adopters, who wish to become a top investigation center, would have high levels of knowledge, low preoccupation on undiscovered business value, but might present budget constraints. Meanwhile, late majority or laggard hospitals would not adopt the application because it is not a golden standard, having little knowledge but no major cost concerns. This relation with the Innovation Curve would require further research.

Finally, in which concerns data privacy and security, it was not considered a true impediment, but rather something to make sure is complied. This result is not in line with the literature, although 71% of managers were concerned with Portuguese and EU data management law compliance. This mismatch may be explained by managers considering that this application would not require data sharing with other hospitals to appropriate the benefits. This requirement, on the other hand, would raise important issues, according to Hospital A's Project Manager.

5.2. Recommendations

5.2.1 Hospitals

Communication. When implementing this kind of projects a Top-down approach is necessary, with a leading team of managers promoting the project (IBM, 2013). This team should establish a direct communication between the management and care units, promoting a culture of information sharing. This could substantially reduce resistance to adoption and the Big Data knowledge gap. For example, were the positive results of certain indicators communicated, caregivers' trust on the tool could improve. In fact, a cultural transformation, creating direct channels between these two levels, was an important success factor in Hospital A's project.

Multidisciplinary team. It is also recommended the formation of a team, joining clinicians with IT representatives, which would work as a decision body in defining priorities and data

needs (McKinsey, 2013). In this context, Hospital A created a multidisciplinary team, with managers, a "Data Scientist", caregivers, medicine teachers and so forth. This allowed the hospital to directly understand what the caregivers' needs truly were, thus programing the parameters accordingly and, as caregivers viewed their input in the tool, minimize resistance. Moreover, this team would share information and debate, hence, reducing any knowledge gap.

Keeping track. In the Project Manager's point of view, it is imperative to keep track of the tool's usage rate, understanding who abandoned and why. This would allow them to react immediately, thus enhancing the adoption rate. Furthermore, having quantitative evidence of the results could reduce the H4 barrier, promoting the projects' expansion.

Programs to develop skills. Modules could be developed across the hospital to enhance data and analytics skills (IBM, 2013). This would not only reduce the knowledge gap, thus enhancing adoption rate, but could also create hybrid employees, with IT and clinical skills, which could look at the data in a creative and different way – "Data Scientists".

Hiring and retaining "Data Scientists". Davenport & Patil (2012) suggest that "Data Scientists" are lured by interesting challenges, enjoying the autonomy to experiment and explore new approaches. Besides, their relationship with the rest of the company is extremely important, being the wage a symbol of their role's value inside the hospital. Therefore, these are aspects to take into account when surpassing the H2 barrier.

5.2.2 Government

Educational program for skills development. Given the skilled labour shortage, it would be crucial to establish centers of excellence to develop students' skills in Big Data analytics (Pentland *et al.*, 2013). Moreover, collaboration between these academic institutions and hospitals could be critical in exploring new Big Data applications.

Appropriate data management regulation. Regarding data privacy and security, one of the biggest barriers was compliance with regulations. Hence, it would be of utmost importance to adjust the regulation to this new reality, promoting data collection, analysis and sharing, while always guaranteeing security and privacy.

NHS common strategic plan. Hospital A's manager suggested the establishment of a cross-hospital plan towards innovation and implementation of this kind of projects – a centralized

plan. These actions could minimize the issue raised by H4, namely by reducing costs, and increase managers' awareness on Big Data. Going a step further, this could promote data pooling between NHS hospitals, which would maximize the applications' benefits, thus removing not only the absence of quantitative evidence but also the siloed data constraint.

5.2.3 Third Party Suppliers/Partners

Prove benefits. Demonstrating case studies, showing performance indices of pilot hospitals may be essential to reduce H4 barrier. In fact, this should be done both at the management and caregiver level, guaranteeing full collaboration of the latter.

Keep it flexible. It is crucial that parameters can be programmed according to the specific needs of caregivers. Indeed, one of Hospital A's requirements in choosing a partner was its capability to build a solution fully personalized from scratch.

Spread information. Forums, lectures and conferences may be promoted inside hospitals for both caregivers and managers, overcoming the knowledge barrier and gap.

Solid "Data Scientists" team. The partner should be prepared to extend the multidisciplinary team created by the hospital, with experts that can teach, develop the tool, and support the hospital's organizational change.

Guarantee Security and Privacy. Data security and privacy was not considered a barrier but rather a requirement from the tool. This means that if the technology does not fulfil this prerequisite it would not be considered. Hence, it is advisable that the privacy and security policy of the hospital is complied, with anonymized data flows and necessary access restrictions.

6. Conclusions

6.1 Conclusions overview

Overall, Big Data can effectively address healthcare systems' current challenges, but there are barriers to implement such projects in a hospital. Focusing in the application of Big Data for predicting adverse events, such as nosocomial infections, in Portuguese hospitals, three main barriers were identified.

Firstly, there is a generalized lack of knowledge regarding this phenomenon and its potential benefits. On the managers' side, only 14% had profound understanding, with a substantial disparity between levels of knowledge. Barriers such as not even consider incorporating these projects in the yearly budget or difficulties in understanding how Big Data can be applied in the various areas of the hospital may, hence, emerge. On the other hand, the gap between managers' and caregivers' knowledge may create distrust regarding the results, increasing caregivers' resistance to adoption.

Secondly, a shortage of "Data Scientists" in Portuguese hospitals was reported. Indeed, Big Data projects require individuals capable of creatively look at the data and understand how it may generate value, with IT skills allied with clinical comprehension. Hence, not having access to, or not being a priority to hire such skills, is a barrier to implementation.

Finally, a high initial investment allied with an undiscovered business value is a true obstacle. A cost-benefit analysis was considered to be often performed and relevant in the decision-making process. Therefore, a project with those characteristics is not expected to be approved by the management. However, there was an interesting split in this Hypothesis, with some hospitals more concerned with costs while others with the lack of proved benefits. Indeed, allied with the disparity in managers' knowledge, this may be connected with hospitals' positioning inside the Innovation Curve.

On the other side, two phenomenon initially identified as barriers, were discarded as being true impediments.

Big Data implementation requires a specific organizational structure where decision-power involves a multidisciplinary team formed by "Data Scientists", caregivers and managers – a patient will not have a doctor: a patient will have a team. Taking into account the collected data, Portuguese hospitals do not reveal such an organizational structure, hence the need for

an organizational transformation. The question stood on whether managers and caregivers would resist. In which concerned managers, it was concluded that they would not resist as they significantly believe that a paradigm and organizational shift is imminent and all the interviewed managers denied any cultural obstacle. This behaviour may be explained by the fact that they are not the "HiPPO" in a hospital environment. However, caregivers' resistance raised much discussion. On the one hand, caregivers considered advantageous and have been adopting the technology implemented by the hospital. Moreover, they rated the usefulness of the application as high and 80% confirmed that they would adopt the tool. Besides, caregivers significantly believed in a paradigm shift towards data-driven medicine. On the other hand, qualitative interviews revealed some concern with caregivers' accommodation, perception of being a liberal profession, resistance to new procedures and so forth. Balancing the two positions, it was concluded that caregivers', overall, would not resist to the application. Such behaviour could be explained by their education in the scientific-method, the personalization of the tool, the input given in the survey and finally a possible technology-friendliness bias on those answering to online surveys.

Furthermore, Big Data projects in a healthcare context are expected to generate much concern around data privacy and security, from political and legal to individual and cultural. However, according to the data collected, these concerns are not a barrier but rather a requirement of the technology. Still, compliance with EU and Portuguese regulation on the matter was the main concern, opening the question on whether this may be a barrier not to managers but rather to regulatory institutions.

All in all, when implementing Big Data projects on prediction of adverse events, Portuguese Hospitals will face three main validated barriers: shortage of "Data Scientists", high investment with undiscovered business value and reduced knowledge regarding Big Data and its potential benefits. Besides, an organizational change will be required but neither managers nor caregivers are expected to resist this transformation and adoption. Additionally, security and privacy are requirements to the tool, rather than barriers to the hospital's implementation.

6.2 Limitations

This thesis was developed not without limitations.

Firstly, articles on the field of Big Data are recent. Hence, there is a shortage of peer-reviewed articles in the matter, thus limiting to some extent the Literature Review, especially in which concerns Portugal's state-of-the-art.

Secondly, knowledge on Big Data among caregivers and managers is limited. Therefore, it was assumed that the explanation provided in the surveys was sufficient for them to formulate an informed answer.

Further, although the thesis aims to explore the barriers in Portuguese hospitals, most caregivers' answers were from Lisbon and Porto, which are not truly representative of all Portuguese hospitals. This was mainly due to time and space constraints.

Finally, the survey was electronic, which implies that those answering to it are more technology-friendly, thus possibly creating a bias in the answers.

6.3 Future research

There is much research to undertake in the field of Big Data and this thesis open yet more questions.

During interviews, several new barriers emerged and all require further research.

> *Absence of incentives to implement preventive tools.* According to the doctor from Hospital D, in most Portuguese hospitals there are no incentives to invest in prevention.
>
> Firstly, there is no accountability and the financial system distorts incentives. For example, as manager from Hospital B mentioned, whether a doctor washes his hands or not (essential to prevent infections) the salary is the same. In another example, according to doctor from Hospital D, most of the financing in "Centro Hospitalares" is based on a severity index calculated when the patient is released – the higher the index the more the hospital receives. Hence, there is no incentive to prevent, but rather to "produce" more. Moreover, emphasising this, the doctor explained that the hospital receives the funding with a 3-year delay.
>
> Besides, according to the same doctor, in opposite to the U.S., insurance companies are not constantly examining possible situations of malpractice, being infections normally considered as a "risk associated to the practice".

Siloed Data. One of the crucial phases in Hospital A's Big Data project was the creation of a common warehouse capable of quickly treating and extracting information without affecting the system. However, there were some barriers in this formation.

Indeed, there were several different applications with different databases inside the hospital and the data marts did not communicate with each other. Moreover, even inside the institution, some units were unwilling to share the information and suppliers hindered information pooling, as this access could jeopardize their power.

Off-the-shelf solutions are limited. According to those involved in the Big Data project, the market has produced extremely inadequate solutions - they are rigid with few parameters and are extremely expensive. Hence, the market is still underdeveloped.

Absence of promotion from Government. A barrier raised by Hospital A was the absence of a consistent role from the Health Ministry in such a paradigm evolution - there is no cross-hospital strategy from the NHS in modernizing and preparing hospitals to this shift. On the contrary, much bureaucracy is faced with several authorizations required, since it is not considered a priority.

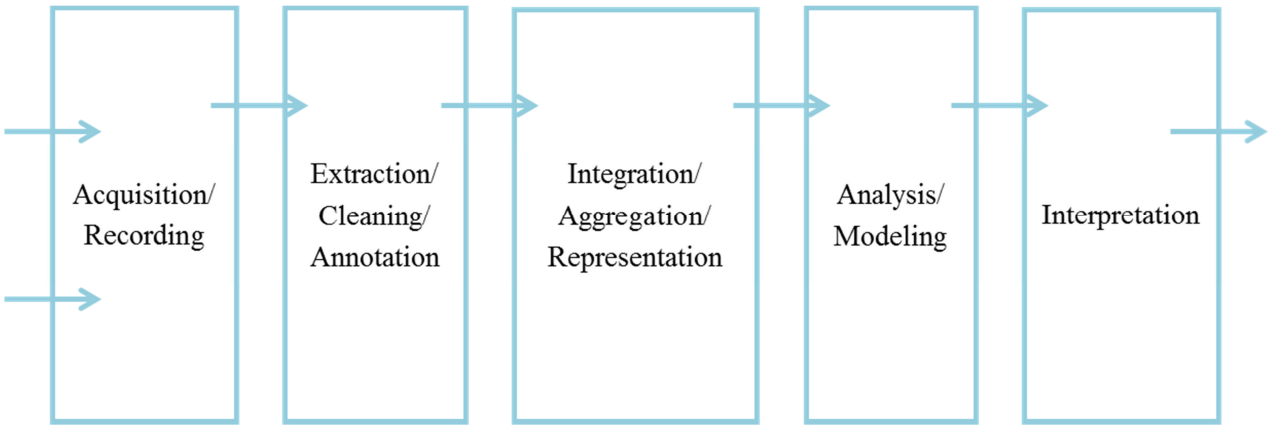
Besides these, further research on the relationship between barriers and hospitals' positioning in the Innovation Curve would be required. Are the barriers the same for early adopters and laggards?

Furthermore, this thesis could be validated with data from other parts of Portugal, such as the interior of the country, being also interesting to perform a geographical comparison.

Finally, it would be important to analyse the barriers from the perspective of other industry players, since they may be different. For example, the data privacy and security is not a concern for providers but could be for regulatory institutions.

Annex A – Big Data Analysis Pipeline

Big Data Analysis Pipeline



Adapted from Agrawal *et al.* (2012)

Annex B – Surveys for managers and caregivers

1. Survey for managers

Este questionário foi realizado no âmbito de uma tese de Mestrado da Católica-Lisbon, desenvolvida na área de "Big Data" na saúde, para a aplicação particular de previsão de intercorrências, tais como as infecções adquiridas nos hospitais.
Agradeço desde já a disponibilidade em contribuir decisivamente para esta investigação.
(tempo estimado de 7 minutos)

Qual o seu nível de conhecimento relativamente ao fenómeno "Big Data" e o seu potencial impacto no sector da saúde?

Nenhum Total

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Pode ser considerado "Big Data":
Seleccione aquelas que considera verdadeiras

- ☐ Análise de dados demográficos dos pacientes do hospital
- ☐ Análise de dados de diversos hospitais para prever qual será o resultado dos diferentes tratamentos num paciente específico
- ☐ Utilizar informação das redes sociais para identificar a evolução de surtos de gripe
- ☐ Digitalização de registos médicos

Tem conhecimento relativamente à utilização de "Big Data" para previsão de intercorrências, tais como as infecções adquiridas no hospital?

Nenhum Total

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

"Big Data" em acção, um exemplo

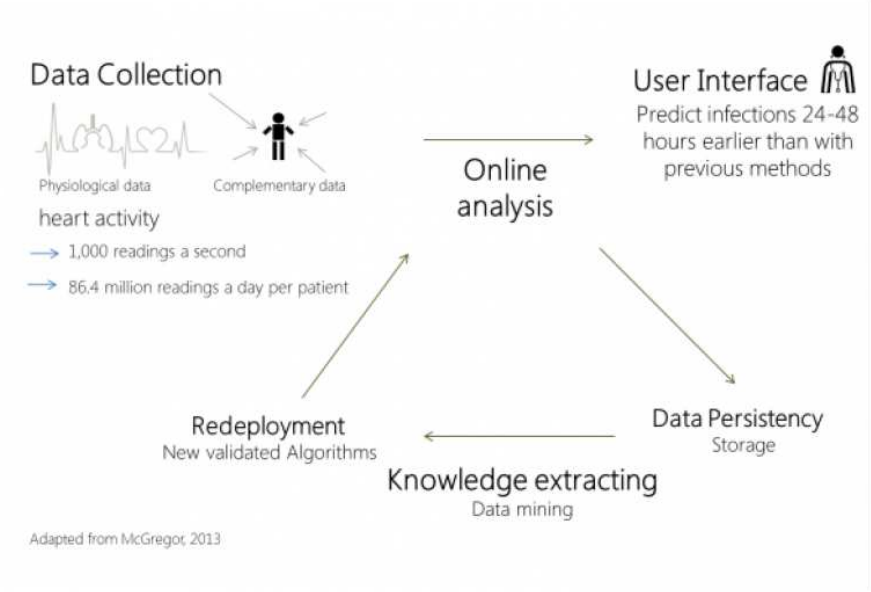
SickKids em Toronto aplica "Big Data analytics" à combinação de "Electronic Health Records" com os sinais vitais dos pacientes, recolhidos cerca de 1,000 vezes por segundo, de forma a detectar sinais de uma possível infecção, aproximadamente 24 horas mais cedo do que métodos anteriores (IBM, 2013).

Por favor considere a aplicação de "Big Data" para previsão de intercorrências, tais como as infecções adquiridas nos hospitais, na resposta às questões seguintes.

- ○ Quero saber mais

"Big Data" é um termo utilizado para descrever uma nova geração de arquitecturas e tecnologias que permitem captar valor, quase ou mesmo tempo real, de grandes quantidades de dados, extremamente variados, que estão constantemente a ser gerados em diferentes origens (IDC,2011). Deste modo, "Big Data" tem vindo a ser definido por vários investigadores através da noção de "V", incorporando características como "volume", "variety" e "velocity" (Gandomi & Haider, 2015; McAfee & Brynjolfsson, 2012), permitindo criar "value" económico (Gantz & Reinsel, 2012).

No caso especifico do exemplo descrito na página anterior, o seguinte esquema permite compreender qual o processo seguido para captar este valor.



No Hospital existe um departamento dedicado à Gestão de Sistemas/Tecnologias de Informação?

- Sim
- Não

Este departamento é considerado meramente operacional ou é determinante para a estratégia e crescimento do hospital?

É um departamento de suporte — É crucial para o crescimento

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Se não, quais as razões?

- [ ] Recursos escassos para alocar a esta área
- [ ] Sistemas/Tecnologias de Informação não são necessários para a actividade hospitalar
- [ ] Criação deste departamento não foi considerada
- [ ] Outra

Acredita que as decisões médicas serão cada vez mais apoiadas em informação baseada em dados, tornando-se o departamento de Sistemas de Informação crucial na actividade hospitalar?

Nada — Totalmente

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Acredita que os(as) médicos(as) e enfermeiros(as) adoptariam este tipo de ferramenta na sua prática diária?

- Sim
- Não

Das seguintes opções, o que contribuiria para resistir a esta adopção?

- [ ] Não acreditam nos benefícios desta inovação, particularmente uma vez que, face à escassez de evidência quantitativa, ainda existe alguma incerteza relativamente à sua eficácia
- [ ] Evitam a utilização de computadores
- [ ] Demonstram algumas dificuldades na adaptação a novas tecnologias
- [ ] Revelam alguma apreensão relativamente à privacidade e segurança da informação que o hospital passará a adquirir, armazenar e analisar
- [ ] Outra

Poderia este comportamento ser considerado uma barreira para a implementação deste tipo de projectos?

- ○ Sim
- ○ Não

Na adopção deste tipo de projectos, relativamente a segurança e privacidade, considera como barreiras à sua implementação as seguintes:

- ☐ Pacientes poderão demonstrar preocupação relativamente à segurança e privacidade da sua informação, evitando o hospital em causa
- ☐ Dificuldade em estar em total conformidade com a lei prevista para gestão de informação da União Europeia e Portugal
- ☐ Possibilidade de quebras de segurança, nas quais os dados poderão ser acedidos por terceiros não autorizados
- ☐ Custo na implementação de medidas de segurança, nomeadamente na melhoria da eficácia da gestão de dados
- ☐ Outras razões relacionadas com privacidade e segurança dos dados dos pacientes

O hospital detém colaboradores qualificados na área de "data analytics"?
"Data scientists" é o termo agora empregue para os que trabalham com "Big Data", que demonstram não só competência na área de análise de dados mas também lógica associativa e criatividade em TI (Davenport et al., 2012).

Não — Muitos

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

A contratação de "Data scientists" é considerada uma prioridade para o hospital?

Não — É a nossa prioridade

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

O orçamento deste ano contempla projectos de novas tecnologias?

- ○ Sim
- ○ Não

No processo de aprovação de projectos, é realizada uma análise de custo-benefício?

Nunca — Sempre

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Quão importante é esta análise na aceitação ou rejeição dos projectos a implementar?

Nada importante — Principal factor de decisão

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Na adopção deste tipo de projectos, considera como barreiras à sua implementação as seguintes:

- ☐ Elevado custo associado às tecnologias necessárias para implementação
  Nomeadamente para adquirir, extrair, integrar, analisar e interpretar os dados
- ☐ Escassez de evidência quantitativa relativa aos benefícios destes projectos
  Por exemplo a nível de poupança e melhoria de produtividade
- ☐ Incerteza relativamente ao Modelo de Negócio a adoptar
- ☐ Tecnologia ainda está numa fase de desenvolvimento inicial

Determine quais as três principais barreiras na implementação deste tipo de projectos:

- [ ] Privacidade e Segurança
- [ ] Rigidez cultural e organizacional
- [ ] Carência de pessoal qualificado na área de "Big Data" (i.e. Data Scientists)
- [ ] Restrições orçamentais e escassez de evidência quantitativa dos benefícios
- [ ] Sector da saúde ter os dados organizados em silos
  Dados estão divididos em silos com acesso condicionado, reduzindo a quantidade e variedade da informação à qual o hospital tem acesso e, desta forma, também os benefícios potenciais destes projectos
- [ ] Iliteracia relativamente a esta inovação
- [ ] Outra

Concluído!

Mais uma vez agradeço a disponibilidade demonstrada. Se desejar informações adicionais sobre este tema, não hesite em contactar-me através do endereço de e-mail ana_bianchi_@hotmail.com

2. Survey for caregivers

Este questionário foi realizado no âmbito de uma tese de Mestrado da Católica-Lisbon, desenvolvida na área de "Big Data" na saúde, para a aplicação particular de previsão de infecções, tais como as infecções adquiridas no hospital.
Agradeço desde já a disponibilidade em contribuir decisivamente para esta investigação.
(tempo estimado de 5 minutos)

Qual o seu nível de conhecimento relativamente ao fenómeno "Big Data" e o seu potencial impacto no sector da saúde?

Nenhum Total

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Pode ser considerado "Big Data":
Seleccione aquelas que considera verdadeiras

- ☐ Análise de dados demográficos dos pacientes do hospital
- ☐ Análise de dados de diversos hospitais para prever qual será o resultado dos diferentes tratamentos num paciente específico
- ☐ Utilizar informação das redes sociais para identificar a evolução de surtos de gripe
- ☐ Digitalização de registos médicos

Tem conhecimento relativamente à utilização de "Big Data" para previsão de infecções, tais como as infecções adquiridas no hospital?

Nenhum Total

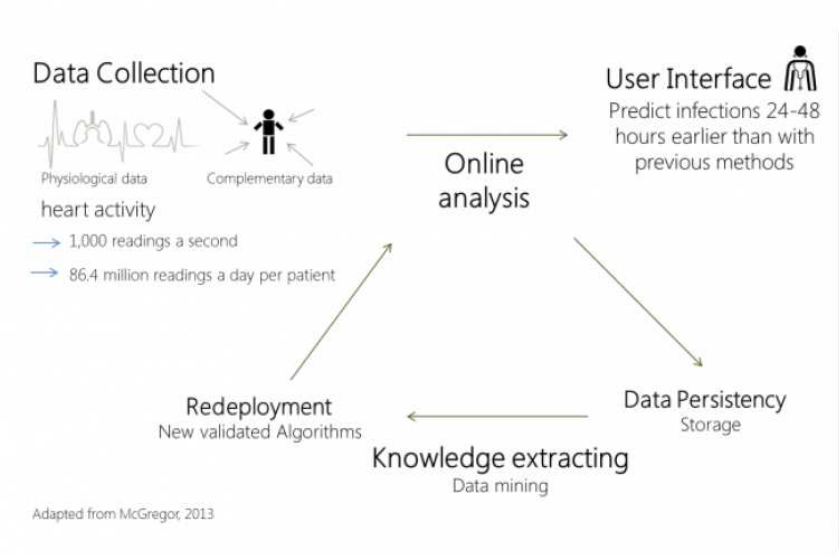
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

"Big Data" em acção, um exemplo

"Big Data" é um termo utilizado para descrever uma nova geração de arquitecturas e tecnologias que permitem captar valor, quase ou mesmo tempo real, de grandes quantidades de dados, extremamente variados, que estão constantemente a ser gerados em diferentes origens (IDC,2011). Deste modo, "Big Data" tem vindo a ser definido por vários investigadores através da noção de "V", incorporando características como "volume", "variety" e "velocity" (Gandomi & Haider, 2015; McAfee & Brynjolfsson, 2012), permitindo criar "value" económico (Gantz & Reinsel, 2012).

SickKids em Toronto aplica "Big Data analytics" à combinação de "Electronic Health Records" com os sinais vitais dos pacientes, recolhidos cerca de 1,000 vezes por segundo, de forma a detectar sinais de uma possível infecção, aproximadamente 24 horas mais cedo do que métodos anteriores (IBM, 2013).

O seguinte esquema permite compreender qual o processo seguido para captar este valor.



O processo constante do esquema é iniciado por um elemento de aquisição de dados, que continuamente recolhe quer fluxos de múltiplos dados fisiológicos, quer informação complementar, como "Electronic Health Records". Por sua vez, esta informação é incorporada numa componente de "Online Analysis" que a processa e analisa em tempo real, sendo, desta forma, capaz de detectar sinais de início de infecção de forma precoce.

Além disso, o sistema é ainda composto por uma segunda fase na qual os dados, quer resultantes da simples recolha, quer processados no elemento de "Online Analysis", são armazenados e analisados através de "Data Mining" e "Machine Learning". Esta fase permite encontrar novas correlações e "clinical rules", ou seja, algoritmos novos e validados, que serão posteriormente reintegrados na componente de "Online Analysis" que assim é constantemente alimentada, gerando um ciclo de contínuo aperfeiçoamento.

Considera vantajosa a tecnologia de informação que tem vindo a ser adoptada pelo hospital?
Por exemplo a prescrição e registo clínico electrónico e telemedicina.

Nenhuma — Alguma — Toda

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Acredita que "Big Data" para previsão de infecções - tais como infecções adquiridas no hospital - poderia ser útil no dia-a-dia da sua actividade?

Nada — Totalmente

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Adoptaria este tipo de método na sua prática diária?

- ○ Sim
- ○ Não

Das seguintes opções, o que contribuiria para resistir a esta adopção?

- ☐ Não acredita nos benefícios desta inovação, particularmente uma vez que, face à escassez de evidência quantitativa, ainda existe alguma incerteza relativamente à sua eficácia
- ☐ Evita a utilização de computadores
- ☐ Demonstra algumas dificuldades na adaptação a novas tecnologias
- ☐ Revela alguma apreensão relativamente à privacidade e segurança da informação que o hospital passará a adquirir, armazenar e analisar
- ☐ Nenhuma
- ☐ Outra

Acredita que as decisões médicas serão cada vez mais apoiadas em informação baseada em dados?

Nada — Totalmente

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Incorporou na sua prática diária a última tecnologia de informação adoptada pelo hospital?
Por exemplo a prescrição e registo clínico electrónico e telemedicina

- ○ Sim, porque é obrigatória
- ○ Sim, porque considero vantajosa
- ○ Não

Se sim, qual?

Género:

- ○ Feminino
- ○ Masculino

Idade

Hospital

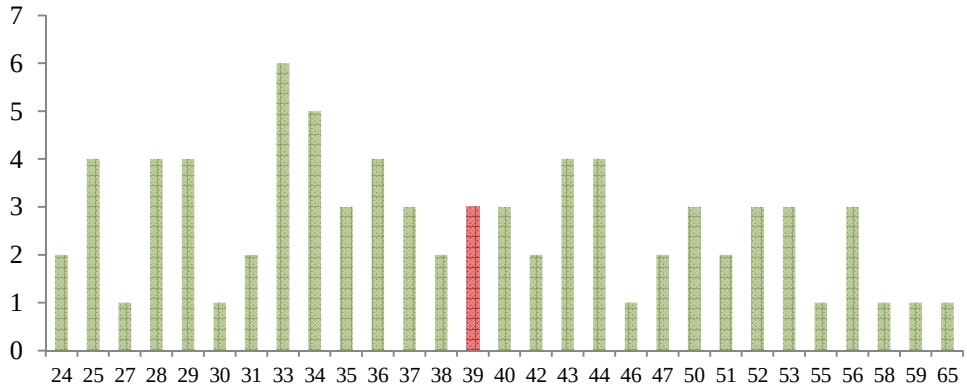
- ○ Privado
- ○ Público

Concluído!

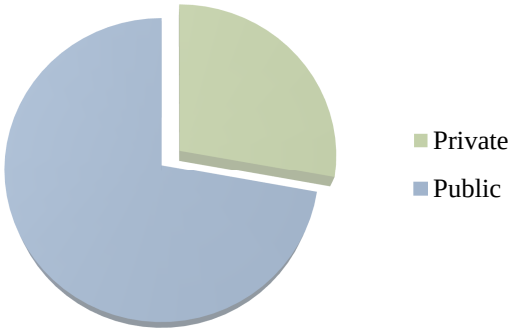
Mais uma vez agradeço a disponibilidade demonstrada. Se desejar informações adicionais sobre este tema, não hesite em contactar-me através do endereço de e-mail ana_bianchi_@hotmail.com

Annex C – Caregivers respondents' demographics

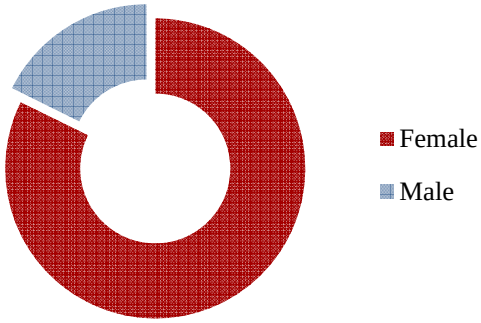
Respondents' age



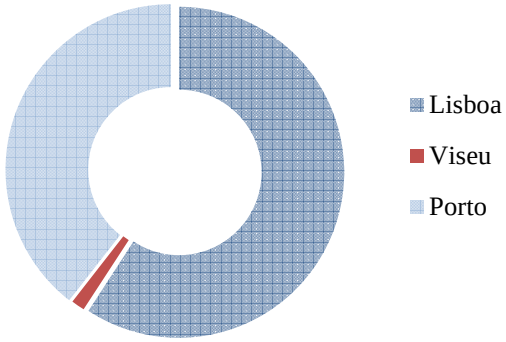
Respondents' practice



Responders' gender



Respondents' locations



Annex D – Hospitals' description

Hospital A: This public hospital is based in Porto, being one of the biggest and most advanced in the country. With more than 5000 employees, it served more than 150,000 patients in the emergency room alone. Currently, it is implementing a Big Data project, aiming to use data for earlier and proactive intervention in infections, antibiotics consumption and in health deterioration of admitted patients.

Hospital B: This hospital is the biggest private hospital in the North and belongs to one of the main private groups in Portugal. Having an initial investment of €90 million, it incorporates 35 specialities. At the moment, a cross-group Big Data project on marketing is being implemented, working on the prevention perspective.

Hospital C: This hospital is located in Porto and is a private, connected with the Catholic Church institution. Although it comprises many specialities, it is a fairly small hospital with less than 60 rooms in the hospitalization division. It is not implementing any Big Data project and does not hold a formal IS department.

Hospital D: This hospital belongs to a larger public group -"Centro Hospitalar"- and is based in Gaia. It is not currently implementing any Big Data project.

Hospital E: This private hospital belongs to a greater private group and is also not implementing any Big Data project.

References

ABEGUNDE, D. O.; MATHERS, C. D.; ADAM, T.; ORTEGON, M. & STRONG, K. (2007). The burden and costs of chronic diseases in low-income and middle-income countries. *The Lancet*. [Online] 370(9603). p.1929-1938. Available from: http://www.sciencedirect.com/science/article/pii/S0140673607616961 [Accessed 11/03/2015]

AGRAWAL, D. *et al*. (2012). Challenges and opportunities with big data. *A community white paper developed by leading researches across the United States.* [Online] Available from: http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf [Accessed 20/03/2015]

ALVARES, S. *et al.* (2004). Telemedicina: situação em Portugal. *Nascer e Crescer*. [Online] 13(2). p.89-97. Available from: http://repositorio.chporto.pt/bitstream/10400.16/547/1/Telemedicina%20-%20situa%C3%A7%C3%A3o%20em%20Portugal.pdf [Accessed 20/03/2015]

ARAÚJO, S.; FARIA, J. P. & CRUZ, J. M. (2007) *Propostas de Melhoria da Segurança dos Sistemas de Informação Clínica em Portugal*. [Online] Available from: http://quatic2007.dsi.uminho.pt/QUATIC2007_AraujoFariaCruz_IndustrialPaper.pdf [Accessed 10/03/2015]

ASCH, S. M.; KERR, E. A.; KEESEY, J.; ADAMS, J. L.; SETODJI, C. M.; MALIK, S. & MCGLYNN, E. A. (2006). Who Is at Greatest Risk for Receiving Poor-Quality Health Care? *The New England Journal of Medicine*. [Online] 354. p.1147-1156. Available from: http://www.nejm.org/doi/full/10.1056/NEJMsa044464 [Accessed 18/03/2015]

BARROS, P. P.; MACHADO, S. R. & SIMÕES, J. A. (2011). Portugal Health system review. *Health Systems in Transition.* [Online] 13(4). Available from: http://www.euro.who.int/__data/assets/pdf_file/0019/150463/e95712.pdf?ua=1 [Accessed 16/03/2015]

BATES, D. W.; SARIA, S.; OHNO-MACHADO, L.; SHAH, A. & ESCOBAR, G. (2014). Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. *Health Affairs.* [Online] 33(7). p.1123-1131. Available from: http://www.trinity.edu/eschumac/HCAD5320/HA%20--%20Big%20Data%20In%20Health%20Care%20-%20Using%20Analytics%20To%20Identify%20And%20Manage%20High-Risk%20And%20High%20Cost%20Patients.pdf [Accessed 13/03/2015]

BLOUNT, M.; EBLING, M. R.; EKLUND, J. M.; JAMES, A. G.; MCGREGOR, C.; PERCIVAL, N.; SMITH, K. P. & SOW, D. (2010). Real-Time Analysis for Intensive Care: Development and Deployment of the Artemis Analytic System. *Engineering in Medicine and Biology Magazine, IEEE*. [Online] 29(2). p.110-118. Available from: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5431948 [Accessed 16/04/2015]

BODENHEIMER, T. (2005). High and Rising Health Care Costs - Part 1: Seeking an Explanation. *Annals of Internal Medicine*. [Online] 142(10). p.847-854. Available from: http://annals.org/article.aspx?articleid=718406 [Accessed 18/03/2015]

BURAU, V. & BLANK, R. H. (2006). Comparing health policy: An assessment of typologies of health systems. *Journal of Comparative Policy Analysis: Research and Practice*. [Online] 8 (1). p.63-76. Available from: http://www.tandfonline.com/doi/pdf/10.1080/13876980500513558 [Accessed 10/03/2015]

CATLEY, C.; STRATTI, H. & MCGREGOR, C. (2008). Multi-dimensional temporal abstraction and data mining of medical time series data: Trends and challenges. *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. [Online] p.4322 – 4325. Available from: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4650166&url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D4650166 [Accessed 01/05/2015]

CELI, L. A.; MARK, R. G.; STONE, D. J. & MONTGOMERY, R. A. (2013). "Big Data" in the Intensive Care Unit. Closing the Data Loop. *American Journal of Respiratory and Critical Care Medicine*. [Online] 187(11). p.1157-1160. Available from: http://www.atsjournals.org/doi/pdf/10.1164/rccm.201212-2311ED [Accessed 17/04/2015]

CENTERS FOR DISEASE CONTROL AND PREVENTION. (2003) *Trends in aging - United States and worldwide*. Georgia: Morbidity and Mortality Weekly Report. [Online] Available from: http://www.ncbi.nlm.nih.gov/pubmed/12645839 [Accessed 10/03/2015]

CENTERS FOR DISEASE CONTROL AND PREVENTION. (2013) *CDC Health Disparities and Inequalities Report — United States, 2013*. Georgia: Morbidity and Mortality Weekly Report. [Online] Available from: http://www.cdc.gov/mmwr/pdf/other/su6203.pdf [Accessed 18/03/2015]

CENTRO HOSPITALAR SÃO JOÃO E.P.E. (2015). [Online] Available from: http://portal-chsj.min-saude.pt/pages/375 [Accessed 10/05/2015]

CHEN, H.; CHIANG, R. & STOREY, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*. [Online] 36(4). p.1165-1188. Available from: http://hmchen.shidler.hawaii.edu/Chen_big_data_MISQ_2012.pdf [Accessed 05/03/2015]

CHERNEW, M. E.; HIRTH, R. A. & CUTLER, D. M. (2003). Increased Spending On Health Care: How Much Can The United States Afford? *Health Affairs*. [Online] 12(4). p.15-25. Available from: http://scholar.harvard.edu/files/cutler/files/increased_spending_on_health_care.pdf [Accessed 18/03/2015]

CROWN, W. (2015). Potential Application of Machine Learning in Health Outcomes Research and Some Statistical Cautions. *Value in Health*. [Online] 18(2). p.137-140. Available from: http://www.sciencedirect.com/science/article/pii/S1098301514047913. [Accessed 06/03/2015]

DAVENPORT, T. H.; BARTH, P. & BEAN, R. (2012). How 'Big Data' Is Different. *MIT Sloan Management Review*. [Online] Available from: http://sloanreview.mit.edu/article/how-big-data-is-different/ [Accessed 20/03/2015]

DAVENPORT, T. H. & PATIL, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*. [Online] Available from: https://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=d8448160-e2c7-478f-b7c8-a201ea8554b2%40sessionmgr112&hid=106 [Accessed 20/03/2015]

DAVIS, K.; STREMIKIS, K.; SQUIRES, D. & SCHOEN, C. (2014) *MIRROR, MIRROR ON THE WALL - How the Performance of the U.S. Health Care System Compares Internationally.* New York: The Commonweath Fund. [Online] Available from: http://www.commonwealthfund.org/~/media/files/publications/fund-report/2014/jun/1755_davis_mirror_mirror_2014.pdf [Accessed 18/03/2015]

DELOITTE. (2014) *2014 Global health care outlook: shared challenges, shared opportunities.* [Online] Available from: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/dttl-lshc-2014-global-health-care-sector-report.pdf [Accessed 18/01/2015]

DUTTA, D. & ROSE, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. *International Journal of Production Economics* [Online] Available from: http://www.sciencedirect.com/science/article/pii/S0925527314004265 [Accessed 05/03/2015]

ESPANHA, R. (2010) *Sistemas de Informação em Saúde e Saúde Online.* [Online] Available from: http://pns.dgs.log.pt/files/2010/07/TIC-A31.pdf [Accessed 15/04/2015]

European Commission Newsroom. (2012) Commission proposes a comprehensive reform of the data protection rules. [Online] 25th January. Available from: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm [Accessed 15/04/2015]

EVANS, J. R. & MATHUR, A. (2005). The value of online surveys. *Internet Research*. [Online] 15(2). p.195-219. Available from: http://www.emeraldinsight.com/doi/pdfplus/10.1108/10662240510590360 [Accessed 08/05/2015]

FEEDZAI (2015). [Online] Available from: https://www.feedzai.com/home/risk-fraud-management/ [Accessed 10/05/2015]

FELDMAN, B.; MARTIN, E. & SKOTNES, T. (2012) *Big Data in Healthcare Hype and Hope*. [Online] Available from: http://thegrcbluebook.com/wp-content/uploads/2013/02/107279699-Big-Data-in-Healthcare-Hype-and-Hope1.pdf [Accessed 03/03/2015]

GANDOMI, A. & HAIDER, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*. [Online] 35(2). p.137-144. Available from: http://www.sciencedirect.com/science/article/pii/S0268401214001066 [Accessed 16/03/2015]

GANS, D.; KRALEWSKI, J.; HAMMONS, T. & DOWD, B. (2005). Medical Groups' Adoption Of Electronic Health Records And Information Systems. *Health Affairs*. [Online] 24(5). p.1323-1333. Available from: http://content.healthaffairs.org/content/24/5/1323.full [Accessed 07/05/2015]

GANTZ, J. & REINSEL, D. (2012). The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. *IDC IVIEW*. [Online] 74(7). p.2561-2573. Available from: http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf [Accessed 16/03/2015]

GOLDBERG, R. N. (1981). Sustained arterial blood pressure elevation associated with pneumothoraces: Early detection via continuous monitoring. *Pediatrics*. [Online] 68(6). p.775-777. Available from: http://pediatrics.aappublications.org/content/68/6/775.short [Accessed 10/05/2015]

GOMES, R. & LAPÃO, L. V. (2008). The adoption of IT security standards in a healthcare environment. *Studies in health technology and informatics*. [Online] 136. p.765-770. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18487824 [Accessed 05/03/2015]

GWATKIN, D. R.; BHUIYA, A. & VICTORA, C. G. (2004). Making health systems more equitable. *The Lancet.* [Online] 364(9441). p.1273–1280. Available from: http://www.sciencedirect.com/science/article/pii/S0140673604171456 [Accessed 18/03/2015]

HARREL, M. C. & BRADLEY, M. A. (2009). Data Collection Methods: Semi-Structured Interviews and Focus Groups. *National Defense Research Institute*. [Online] Available from: http://www.dtic.mil/dtic/tr/fulltext/u2/a512853.pdf [Accessed 09/05/2015]

HART, J. T. (1971). The inverse care law. *The Lancet.* [Online] 297(7696). p.405–412. Available from: http://www.sciencedirect.com/science/article/pii/S014067367192410X [Accessed 18/03/2015]

HOFMARCHER, M. M.; OXLEY, O. & RUSTICELLI, E. (2007). Improved Health System Performance through better Care Coordination. *OECD Health Working Papers.* [Online] 30. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/improved-health-system-performance-through-better-care-coordination_246446201766;jsessionid=ft060m4mgx6o.x-oecd-live-02 [Accessed 11/03/2015]

HORTON, R. (2012). GBD 2010: understanding disease, injury, and risk. *The Lancet*. 380. p.2053-2054.

HUGHES, G. (2011) *How big is 'big data' in healthcare?*. [Online] October 21st 2011. Available from: http://blogs.sas.com/content/hls/2011/10/21/how-big-is-big-data-in-healthcare/ [Accessed: 18/03/2015]

IBM. (2013) *Data-driven healthcare organizations use big data analytics for big gains*. U.S.: IBM Corporation. [Online] Available from: http://www-03.ibm.com/industries/ca/en/healthcare/documents/Data_driven_healthcare_organizations_use_big_data_analytics_for_big_gains.pdf [Accessed 20/03/2015]

IDC. (2015) *Portugal no Ponto de Viragem da Transformação Digital.*

INE. (2013) *Dados Estatísticos.* [Online] Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_base_dados&bdpagenumber=2&bdind_por_pagina=15&bdtemas=1117&bdfreetext=Palavra(s)%20a%20encontrar&bdnivelgeo=1&contexto=bd&atributoordenar=null&atributoordem=null&bdsubtemas=111711 [Accessed 25/05/2015]

JOURNAL OF INNOVATION MANAGEMENT (2015). *Call for Papers - Thematic Issue: Boosting Innovation with Big Data*. [Online] Available from: https://web.fe.up.pt/~ijmai/CFP_BigData.pdf [Accessed on 10/05/2015]

KAMBATLA, K.; KOLLIAS, G.; KUMAR, V. & GRAMA, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*. [Online] Available from: http://www.sciencedirect.com/science/article/pii/S0743731514000057 [Accessed 12/03/2015]

KHAZAEI, H. *et al.* (2014). Toward a Big Data Healthcare Analytics System: A Mathematical Modeling Perspective. *2014 IEEE World Congress on Services (SERVICES)*. [Online] p.208-215. Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6903267 [Accessed 16/04/2015]

KOHN, M. S.; SUN, J.; KNOOP, S.; SHABO, A.; CARMELI, B.; SOW, D.; SYED-MAHMOOD, T. & RAPP, W. (2014). IBM's Health Analytics and Clinical Decision Support. *Yearbook of Medical Informatics.* [Online] 9(1). p.154–162. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287097/ [Accessed 20/04/2015]

LAPÃO, L. V. (2005). A complexidade da saúde obriga à existência de uma arquitectura de sistemas e de profissionais altamente qualificados: O problema da saúde – inexistência de informação impossibilita a gestão. *Polytechnical Studies Review.* [Online] 2(4). p.15-27. Available from: http://www.scielo.oces.mctes.pt/pdf/tek/n4/v2n4a02.pdf [Accessed 20/03/2015]

LAPÃO, L. V. (2010). Papel crítico dos Sistemas de Informação na saúde - Considerações no âmbito do Plano Nacional de Saúde 2011-2016. *Instituto de Higiene e Medicina Tropical – Universidade Nova de Lisboa.* [Online] Available from: http://pns.dgs.log.pt/files/2010/09/lvl.pdf [Accessed 20/03/2015]

LAVALLE, S.; LESSER, E.; SHOCKLEY, R.; HOPKINS, M. S. & KRUSCHWITZ, N. (2010). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review.* [Online] Available from: http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/ [Accessed 07/05/2015]

LEE, S.; CHUN, C.; LEE, Y. & SEO, K. (2008). The National Health Insurance system as one type of new typology: The case of South Korea and Taiwan. *Health Policy*. [Online] 85 (1). p.105-113. Available from: http://www.sciencedirect.com/science/article/pii/S0168851007001546 [Accessed 10/03/2015]

LEHMAN, L. H.; ADAMS, R. P.; MAYAUD, L.; MOODY, G. B.; MALHOTRA, A.; MARK, R. G. & NEMATI, S. (2015). A Physiological Time Series Dynamics-Based Approach to Patient Monitoring and Outcome Prediction. *IEEE Journal of Biomedical and Health Informatics.* [Online] 19(3) p.1068 – 1076. Available from: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6846269&url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D6846269 [Accessed 10/05/2015]

LISBON MACHINE LEARNING SCHOOL (2015). [Online] Available from: http://lxmls.it.pt/2015/ [Accessed 10/05/2015]

LIU, J. L. Y.; MANIADAKIS, N.; GRAY, A. & RAYNER, M. (2002). The economic burden of coronary heart disease in the UK. *Heart*. [Online] 88(6). p.597–603. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1767465/ [Accessed 12/03/2015]

MARTINHO, R.; VARAJÃO, J.; CUNHA, M. & BALLONI, A. J. (2014). Tecnologias e Sistemas de Informação em entidades hospitalares – Dois casos de hospitais portugueses. *GEstão de SIstemas e Tecnologias da Informação em Hospitais*. [Online] Available from: http://repositorium.sdum.uminho.pt/bitstream/1822/31098/1/GESITI-2014.pdf [Accessed 15/03/2015]

MCAFEE, A. & BRYNJOLFSSON, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*. [Online] Available from: https://hbr.org/2012/10/big-data-the-management-revolution/ar [Accessed 18/03/2015]

MCGREGOR, C. (2013). Big Data in Neonatal Intensive Care. *Computer*. [Online] 46(6). p.54-59. Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6513228 [Accessed 01/04/2015]

MCGREGOR, C.; CATLEY, C. & JAMES, A. (2012). Variability analysis with analytics applied to physiological data streams from the neonatal intensive care unit. *2012 25th International Symposium on Computer-Based Medical Systems (CBMS).* [Online] Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6266385 [Accessed 10/05/2015]

MCKINSEY. (2008) *Cutting costs: The great health-care challenge of the century*. [Online] Available from: http://mckinseyonsociety.com/whats-working-in-health-reform/ [Accessed 17/03/2015]

MCKINSEY. (2011a). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. [Online] Available from: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation [Accessed 10/03/2015]

MCKINSEY. (2011b). Are you ready for the era of 'big data'? *McKinsey Quarterly*. [Online] Available from: http://www.mckinsey.com/insights/strategy/are_you_ready_for_the_era_of_big_data [Accessed 15/03/2015]

MCKINSEY. (2013) *The 'big data' revolution in healthcare – Accelerating value and innovation*.

MENDES, V. (2012) *Qualidade no Serviço Nacional de Saúde: Evolução recente e perspectivas futures.* A Thesis Submitted in partial fulfilment of the Requirements of Universidade Nova de Lisboa for the Degree of Master of Healthcare Management. Lisbon: Universidade Nova de Lisboa.

MORENO-SERRA, R. & SMITH, P.C. (2012). Does progress towards universal health coverage improve population health? *The Lancet.* 380. p.917–923.

MORGAN, D. & ASTOLFI, R. (2013). Health Spending Growth at Zero: Which Countries, Which Sectors Are Most Affected? *OECD Health Working Papers.* [Online] 60. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/health-spending-growth-at-zero_5k4dd1st95xv-en;jsessionid=cinbli15cigq6.x-oecd-live-02 [Accessed 25/04/2015]

MY HEALTH LONDON (2015) *Today's NHS - our current challenges.* [Online] Available from: http://www.myhealth.london.nhs.uk/your-health/call-action/nhs-today [Accessed 15/04/2015]

NOVA INFORMATION MANAGEMENT SCHOOL. (2015). *Perceção sobre a qualidade dos serviços prestados no SNS.*

OECD. (1987) *Financing and delivering healthcare: a comparative analysis of OECD Countries.*

OECD. (2014) *OECD Factbook 2014 – Economic, Environmental and Social Statistics.* [Online] Available from: http://www.oecd-ilibrary.org/economics/oecd-factbook-2014_factbook-2014-en [Accessed 20/03/2015]

OECD. (2015) *OECD.StatExtracts.* [Online] Available from: http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT [Accessed 20/03/2015]

OPDENAKKER, R. (2006). Advantages and Disadvantages of Four Interview Techniques in Qualitative Research. *FORUM: QUALITATIVE SOCIAL RESEARCH.* [Online] 7(4). Available from: http://www.qualitative-research.net/index.php/fqs/article/view/175/392 [Accessed 07/05/2015]

PAULY, M. V. (1993). U.S. health care costs: the untold true story. *Health Affairs.* [Online] 12(3). p.152-159. Available from: http://content.healthaffairs.org/content/12/3/152.full.pdf [Accessed 18/03/2015]

PEIXOTO, H. (2013) Equipa da U.Porto vence competição internacional de Data Science. *Notícias Universidade do Porto.* [Online] 5th September. Available from: http://noticias.up.pt/equipa-da-u-porto-vence-competicao-internacional-de-data-science/ [Accessed 01/05/2015]

PENTLAND, A.; REID, T. G. & HEIBECK, T. (2013). Big Dara and Health - Revolutionizing medicine and Public Health. *WISH Big Data and Health Report 2013.*

PORTUGAL. Alto Comissariado da Saúde. (2010) *Plano Nacional de Saúde 2011-2016: Política do medicamento, dispositivos médicos e avaliações de tecnologias em saúde.* [Online] Available from: pns.dgs.pt/files/2010/11/PM1.pdf [Accessed 17/03/2015]

PROPPER, C. (2000). The demand for private health care in the UK. *Journal of Health Economics*. 19. p.855–876.

RAGHUPATHI, W. & RAGHUPATHI, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems.* [Online] 2(3). Available from: http://www.biomedcentral.com/content/pdf/2047-2501-2-3.pdf [Accessed 18/03/2015]

RUSSOM, P. (2011). Big Data Analytics. *TDWI Best Practices Report*. [Online] Available from: http://www.sas.com/content/dam/SAS/en_us/doc/research2/big-data-analytics-105425.pdf [Accessed 17/03/2015]

SAKELLARIDES, C.; REIS, V.; ESCOVAL, A.; CONCEIÇÃO, C. & BARBOSA, P. (2005) *O futuro do Sistema de Saúde Português: "Saúde 2015"*. [Online] Available from: https://infoeuropa.eurocid.pt/registo/000036272/ [Accessed 20/04/2015]

SAPO (2015). [Online] Available from: http://maquinadotempo.sapo.pt/ [Accessed 10/05/2015]

SHIELDS, A. E.; SHIN, P.; LEU, M. G.; LEVY, D. E.; BETANCOURT, R. M.; HAWKINS, D. & PROSER, M. (2007). Adoption Of Health Information Technology In Community Health Centers: Results Of A National Survey. *Health Affairs*. [Online] 26(5). p.1373-1383. Available from: http://content.healthaffairs.org/content/26/5/1373.full [Accessed 05/05/2015]

SOW, D. M.; SUN, J.; BIEM, A.; BLOUNT, M. L. & EBADOLLAHI, S. (2012). Real-time analysis for short-term prognosis in intensive care. *IBM Journal of Research and Development.* [Online] 56(5). Available from: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6261584 [Accessed 20/04/2015]

STACEY, M. & MCGREGOR, C. (2007). Temporal abstraction in intelligent clinical data analysis: A survey. *Artificial Intelligence in Medicine.* [Online] 39(1). p.1-24. Available from: http://www.sciencedirect.com/science/article/pii/S0933365706001345 [Accessed 21/04/2015]

TOMÉ, A.; BROEIRO, P. & FARIA-VAZ, A. (2008). Os sistemas de prescrição electrónica. *Revista Portuguesa de Medicina Geral e Familiar*. [Online] 24(5). p.632-640. Available from: http://www.rpmgf.pt/ojs/index.php?journal=rpmgf&page=article&op=view&path%5B%5D=10554 [Accessed 15/03/2015]

UN. (2012) *UNdata – A world of information.* [Online] Available from: http://data.un.org/Data.aspx?d=PopDiv&f=variableID%3a22 [Accessed 20/03/2015]

VELTHUIS, E.; MALKA, E. & RICHARDS, M. (2013). 'Big Data' in Health Care - What Does It Mean and Will It Make a Difference? *Value in Health*. [Online] 16(7). Available from: http://www.sciencedirect.com/science/article/pii/S1098301513028118# [Accessed 20/01/2015]

VITALIDI (2015). [Online] Available from: http://www.vitalidi.com/index.php/what-is-it [Accessed 10/05/2015]

WALCZUCH, R.; VAN BRAVEN, G. & LUNDGREN, H. (2000). Internet adoption barriers for small firms in The Netherlands. *European Management Journal*. [Online] 18(5). p.561-572. Available from: http://www.sciencedirect.com/science/article/pii/S0263237300000451 [Accessed 06/05/2015]

WAMBA, S. F.; AKTER, S.; EDWARDS, A.; CHOPIN, G. & GNANZOU, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*. [Online] Available from: http://www.sciencedirect.com/science/article/pii/S0925527314004253 [Accessed 04/03/2015]

WHITE, M. (2012). Digital workplaces Vision and reality. *Business Information Review.* [Online] 29(4). p.205-214. Available from: http://bir.sagepub.com/content/29/4/205.short [Accessed 20/03/2015]

WHO. (2000) *The World Health Report 2000 – Health Systems: Improving Performance.*

WHO. (2010) *Health Systems Financing - The path to universal coverage.* Geneva: WHO Press. [Online] Available from: http://whqlibdoc.who.int/whr/2010/9789241564021_eng.pdf?ua=1 [Accessed 18/03/2015]

WHO. (2012) *Global Health Observatory Data Repository- Health expenditure ratios, Data by country*. [Online] Available from: http://gamapserver.who.int/gho/interactive_charts/health_financing/atlas.html [Accessed 20/03/2015]

WHO. (2013) *Universal Health Coverage: Supporting Country Needs*. [Online] Available from: http://www.who.int/contracting/UHC_Country_Support.pdf [Accessed 14/03/2015]

WHO. (2014a) *Global status report on noncommunicable diseases 2014.* Geneva: WHO Press. [Online] Available from: http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf?ua=1 [Accessed 12/03/2015]

WHO. (2014b) *WHO Global Health Expenditure Atlas.* Geneva: WHO Press. [Online] Available from: http://www.who.int/health-accounts/atlas2014.pdf [Accessed 16/03/2015]

ZILLNER, S.; RUSITSCHKA, S. & SKUBACZ, M. (2014). Big Data Story - Demystifying Big Data with Special Focus on and Examples from Industrial Sectors. *EU BIG – European Big Data Public Private Forum.* [Online] Available from: http://big-project.eu/publications [Accessed 11/05/2015]