Repository Istituzionale dei Prodotti della Ricerca del Politecnico di Bari

Business intelligence in the healthcare industry: The utilization of a data-driven approach to support clinical decision making

This is a pre-print of the following article

Original Citation:
Business intelligence in the healthcare industry: The utilization of a data-driven approach to support clinical decision making / Basile, Luigi Jesus; Carbonara, Nunzia; Pellegrino, Roberta; Panniello, Umberto. - In: TECHNOVATION. - ISSN 0166-4972. - STAMPA. - 120:(2023). [10.1016/j.technovation.2022.102482]

Availability:
This version is available at http://hdl.handle.net/11589/236101 since: 2023-05-15

Published version
DOI:10.1016/j.technovation.2022.102482

Publisher:

Terms of use:

(Article begins on next page)

31 December 2024

This is an Accepted Manuscript of an article published by Elsevier in Technovation in 2023, available online:
https://doi.org/10.1016/j.technovation.2022.102482
© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Business intelligence in the healthcare industry: the utilization of a data-driven approach to support clinical decision making.

Authors' names: Luigi Jesus Basile, Nunzia Carbonara, Roberta Pellegrino, Umberto Panniello.

Authors' affiliation: Polytechnic University of Bari

Corresponding author: Luigi Jesus Basile

Authors' email: luigijesus.basile@poliba.it; nunzia.carbonara@poliba.it; roberta.pellegrino@poliba.it; umberto.panniello@poliba.it.

Abstract: The pandemic has accelerated the digitization of many businesses, forcing people to use digital technologies. Using these technologies generates a huge amount of data. Business Intelligence (BI) is concerned with the extraction, analysis, and presentation of data to make decisions and improve the management of firms. This becomes particularly relevant in the healthcare sector where decisions are traditionally made on the physicians' experience. Much work has been done on applying BI in the healthcare sector. Most of these studies were focused only on IT or medical aspects, while the usage of BI for improving the management of healthcare processes is neglected. This research aims at filling this gap by investigating whether a decision support system (DSS) based on the exploitation of data through BI can outperform traditional experience-driven practices for managing processes in the healthcare domain. To achieve this objective, we develop a DSS for reducing the overall costs of a healthcare process in the oncology field. The DSS was developed in two versions: the first is experience-driven while the second is data-driven. The results of our study show that the usage of BI for managing the healthcare processes proved to improve traditional processes based only on the physicians' experience, from a business (i.e., costs reduction) and managerial (i.e., effectiveness improving) point of view.

Keywords: decision support system, decision making, business intelligence, healthcare, oncology, data driven, big data.

Declarations of interest: none

Introduction

The Covid-19 pandemic has generated deep transformations in several industries around the world. While from a human and social point of view the changes are dramatic, many new opportunities have emerged in business and education [1] [2]. The need to maintain the social distance caused by the pandemic and keep working has forced companies, employees, students, and different professionals to accelerate digital transformation. McKinsey professionals have estimated that because of COVID-19, digital technology adoption in Europe has jumped from 81% to 95%, the gap between European countries has been considerably reduced. This change would have only been achieved in 2-3 years at pre-pandemic growth rates [1]. One of the sectors more impacted by digitalization is the healthcare sector. In the U.S., telemedicine usage has grown from 0.1% of users in February 2020 to 43.5% in April 2020 [3]. Applications that leverage digital technologies are multiplying day by day. Very interesting examples come from the development of new wearable technologies which make it possible to monitor and analyse clinical data in real time [4]. All-new forms of digitization are based on the massive use of data for knowledge extraction. The business process that deals with this is the Business Intelligence (BI), defined as a combination of processes, policies, culture, and technologies for gathering, manipulating, storing, and analysing huge collections of data (the so-called "big data") coming from internal and external sources, to communicate information, create knowledge, and inform decision making. BI helps report business performance, uncover new business opportunities, and make better business decisions regarding competitors, suppliers, customers, financial issues, strategic issues, products, and services [5]. Therefore, the recent massive use of digital technologies opens many opportunities for BI and generally for the exploitation of big data for different purposes, but with the common goal of making better-informed decisions. After the pandemic, the application of BI in the healthcare sector is expected to experience a real renaissance, as witnessed by the increasing number of studies in the field and applications [6].

In the healthcare sector, BI can be considered as a boost to improve traditional decisions made by physicians (i.e. medical doctors) [7]. However, even if, on one side, there are plenty of applications based on the use of data that support physicians in selecting and monitoring prognosis and diagnosis or to improve the ICT architectures and data management systems, on the other side, from the management point of view the use of data for improving healthcare processes seems to be still limited [8] [9]. Despite the limited attention, this topic seems to be however very promising. Decision-making in healthcare is challenging since many decisions are still made based on experience and procedures rather than on rigorous approaches integrating BI into the decision-making process. The management of decision-making processes in the health sector is very complex due to the high complexity of choices. Optimization of the decision-making process may make it possible to avoid wasting resources. The objective of this paper is to investigate whether a decision support system (DSS) based on the exploitation of data through BI can outperform traditional experience-driven practices for managing processes in the healthcare domain. Ultimately, the research question we aim to answer is: "Can a data-driven DSS improve the healthcare process management better than DSS based solely on experience and literature?". The paper is organized as follows. In the second section, we provide a literature review of previous works focused on the usage of business intelligence in the healthcare sector, in the third section we describe the methodology used in this work while we discuss the main results and conclusions in the remaining two sections.

1. Business intelligence for decision making in healthcare

Business Intelligence is the process of obtaining information and then knowledge for decision-makers by collecting data from different sources, analysing the data through data mining techniques, and finally creating reports that allow easy visualization [10].

The BI process is based on the collection of data of different types, stored in data warehouses or data marts where multidimensional cube analyses are carried out to filter the information that is most useful for the purposes. Then there is a data exploration phase using statistical and visualization techniques, and the application of learning models with two different purposes: prediction or knowledge. Finally, concerning the results obtained through the learning models, optimization criteria are used to allow decision-makers to easily identify the optimal choices [11] [12].

Much work has been done in the domain of BI applied to the healthcare sector. These works can be grouped into three main categories, depending on their focus [13]. The first group includes studies focused on applying BI to refine prognoses and diagnoses and select the best treatments, by using medical informatics, data mining, and machine learning algorithms. An application of these algorithms made it possible to define working models for the early diagnosis of heart disease, using attributes such as age, sex and type of chest pain, or even cholesterol levels and blood pressure [14].

The second group of studies is on improving data management and communication performance through the usage of ICT infrastructure to ensure basic health services [15]. For instance, at the London Health Sciences Centre, BI technology and solution have been adopted for improving infection rates. They changed the hospital's old systems from online transactional processing (OLTP) BI solution to online analytical processing (OLAP) BI solution [16].

The third group of studies focuses on monitoring and improving healthcare processes. Most of these studies use data to monitor a process, by collecting data and producing–KPI to address problems related to surgical processes, to control the rate of infections linked to operations, to improve the use of time and reduce costs, and develop indicators relating to the quality of clinical services and expected life [17] [18] [19]. However, few studies–use BI to decide on healthcare process management. The use of BI to actually make a decision, in fact, still requires data manipulation after the data has been delivered because another series of steps are needed to arrive at a decision after the data has been viewed [20] [21]. Although the fundamental goal of BI is to enable informed decision-making that results in improved organizational performance [21], it has been argued that, perhaps due to the lack of integration of BI into the decision-making process, more than 50% of BI implementations fail to influence the decision-making process in any meaningful way [22]. In the healthcare sector, where decision-making is both a crucial and challenging task, many decisions are still made based on experience and procedures rather than on rigorous approaches integrating BI into the decision-making process. The motivation for this research stems from the observation that, despite a great deal of work that has been done on the application of BI technologies in the healthcare sector, these systems often fail to influence managerial decision-making [23]. In particular, to the best of our knowledge, no work has demonstrated whether data-driven decisions outperform experience-based decision-making processes in healthcare. This paper aims at contributing to the third category of studies on BI applications to support the healthcare decision-making for improving the healthcare processes management by investigating whether the data-driven decision version improves the healthcare process management better than the version based solely on experience and literature.

2. Methodology

In order to investigate whether a decision support system (DSS) based on the exploitation of data through BI can outperform traditional experience-driven practices for managing processes in the healthcare domain, we develop a three steps methodology which compares an approach integrating BI into the healthcare decision-making process with a traditional approach based on experience. To this aim, we refer to the decision process of treatment strategy selection in BRCA1/2[1] Mutation Carriers with breast cancer.

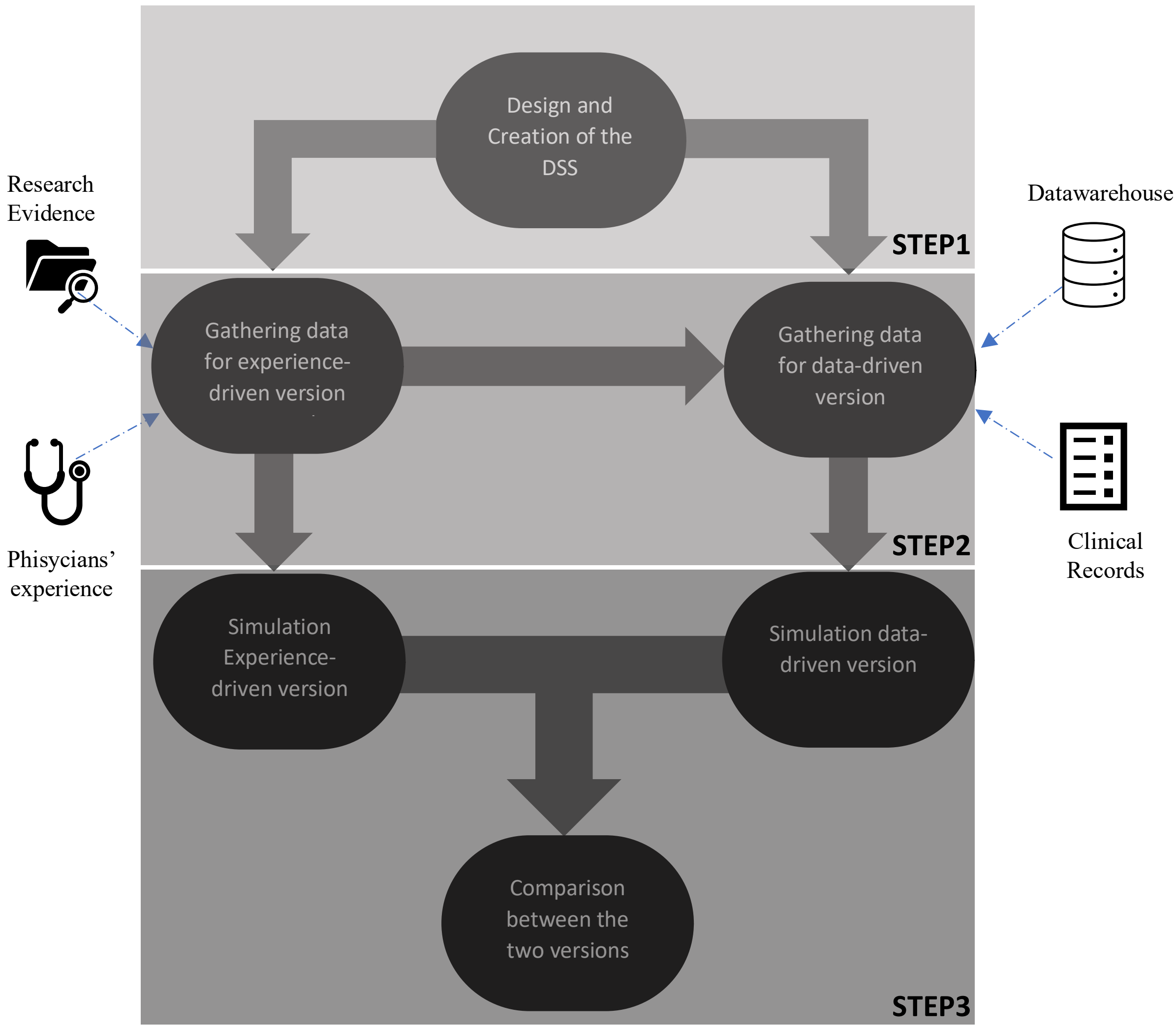

Figure 1 Steps of the methodology.

The literature on clinical management of patients at increased risk for breast cancer shows that there exist multifarious possibilities to treat high-risk BRCA mutated patients with breast cancer and reduce their risk of new tumors [24] . Although they are all equally feasible, none of the clinical guidelines

[1] BRCA1 (BReast CAncer gene 1) and BRCA2 (BReast CAncer gene 2) are genes that produce proteins that help repair damaged DNA. A woman's lifetime risk of developing breast and/or ovarian cancer is markedly increased if she inherits a harmful variant in BRCA1 or BRCA2, [29].

report treatment pathway algorithms specific to the treatment of patients with BRCA breast cancer (BC) [25]. Yet, they consume different resources, drugs, radiotherapy, surgery, diagnostics, etc., thus burdening the health-care system cost differently [26]. Therefore, improving the decision-making process supporting the selection of treatment strategies for high-risk women already diagnosed with BC may potentially produce advantages for the healthcare systems, in terms of cost and effectiveness of the processes.

The three main steps of the developed methodology, depicted in Figure 1, are described in the following.

Step1: Design and creation of the decision support system (DSS).

In this paper, we first build a cost decision-making model (i.e., DSS), which compares the healthcare costs for diverse treatment strategies for BRCA mutated women with breast cancer and calculates the cancer treatment costs that could potentially be prevented if the treatment strategy with minimum cost is chosen for treating high-risk women with breast cancer. Then we develop two versions of the DSS. The first one is the experience-driven one, which uses data provided by literature and historical data. The second is the data-driven one, it is built by modifying the input variable of the probability of BRCA mutation-positive in affected individuals extracted from the created database. To build the DSS and define possible clinical pathways for BRCA mutated patients affected with BC, identify the risk events of different clinical pathways, and associate costs, we did interviews with a multidisciplinary team of doctors of the "Giovanni Paolo II" Cancer Institute, located in Bari (Italy), a reference center on genetic pathologies oncological.

The DSS assesses and computes the cost of each possible treatment strategy throughout the lifetime of the BRCA affected patient, and thus defines the most cost-effective therapeutic pathway, given the assumption of the same clinical efficacy among the different clinical pathways. The DSS works under conditions of uncertainty, taking into account the risks and complications that may arise throughout the patient's life and therefore the costs associated with the management of such events, values obtained by the historical data. The study examines the diagnostic and therapeutic care pathway of BRCA mutated patients receiving the first diagnosis at 40 years[2] of age, e.g. two clinical pathways are: women opting for intensive radiological follow-up and those of women opting for prophylactic mastectomy and subsequent ultrasound follow-up, both of the options have been considered over 35 years, in accordance with the first eligible age for the testing program from 40 to 75 years. DSS allows to simulate the different clinical pathways under uncertainty and obtain the associated costs, thus identifying the clinical pathway that minimizes costs, called "optimal therapeutic path". Then, based on the actual practice ("as-is" scenario), the therapeutic pathway that the patient would follow without the optimization model, defined on the basis of historical data, is considered and the associated cost is calculated. The difference between the two costs of the two therapeutic paths (optimal versus as is) represents the net unit savings per affected patient, which is the output of the model.

The logic of the decision support system may be summarized in the following steps:

1. Calculation of the costs associated with the therapeutic pathways.

2. Comparison of costs of alternative therapeutic pathways and choice of the one with the lowest cost ("optimal therapeutic path").

[2] It is the first age eligible for BRCA genetic testing program: genetic counselling and genetic test.

3. Calculation of the cost of the therapeutic pathways in the “as is” scenario - the therapeutic pathway, that the patient would follow without the optimization model, defined on the basis of historical data.

4. Comparison of the cost of the “optimal therapeutic pathway” with the cost of the “as is” therapeutic pathway.

5. Calculation of the unit “saving” per affected patient: the cost savings that would be obtained by choosing the optimal therapeutic path, throughout the patient’s entire residual life. To this end, it is calculated by considering all the net potential saving (or costs) generated by the optimal path in each year, until the end of the life of the patient, discounted with a predefined discount rate, identified from the literature. Specifically, the Net Present Value (NPV) has been used to calculate the present value (actual unit “saving” per affected patient) of a series of future payments (with a discount rate of 3%) [27]

6. Identification of the most cost-effective therapeutic pathway.

Step 2: Gathering data.

The input variables of the DSS are of two types, probabilistic and deterministic, estimated differently in the two versions of the DSS. In particular, the substantial difference between the two versions of the DSS is related to the estimation of one input variable, the probability of BRCA mutation-positive in affected individuals. Several scientific articles report a relationship between the incidence of genetic mutations and ethnicity and territory, and they found that the incidence of BRCA gene mutation varies between different ethnicities, varying from 9.4% for the Middle East to 15.6% for the African ethnic group [28]. In another study, the main mutations affecting certain populations were also specifically defined and then divided by country [29]. In light of this evidence, for the data-driven version, the probability of BRCA mutation-positive in affected individuals has been estimated by using historical data available for the Puglia region and the ones available for the provinces of the region.

The input values for experience-driven DSS were extracted from the scientific literature on the topic and from historical data presented by the doctors during the interviews (Table 1 and Table 2). The costs are summarised in Appendix A, and the main source is the National Health Service (NHS).

Table 1 Model probabilistic input variables

Variables	Distribution	Values	Sources
Starting age (Affected)	Normal	Mean = 40 Std. Dev. = 2.5	[30], [31], [32] [33]
The probability of BRCA mutation positive in affected individuals	Uniform	Min = 10% Max = 20%	[30], [31], [32] [33]
Annual risk of new incidence of breast cancer if BRCA positive	20-29	0.005	[33]
	30-39	0.015	
	40-49	0.03	
	50-59	0.026	
	60-69	0.012	
	70-79	0.012	
The annual risk of contralateral breast cancer if BRCA positive	20-29	0	[33]

	30-39	0.05	
	40-49	0.04	
	50-59	0.03	
	60-69	0.03	
	70-79	0.03	
The probability that the patient is treated with radiotherapy after mastectomy		40%	Historical data
The probability that the patient is treated with radiotherapy after quadrantectomy		95%	Historical data
The probability to undergo genetic counseling	Bernoulli	45%	Historical data
The probability to undergo BRCA genetic testing	Bernoulli	45%	Historical data
Probability of detecting suspected local recurrence (skin or lymph node recurrences)		5%	Historical data
Risk of surgery complications	Uniform	Min = 10% Max = 20%	Historical data
Positive biopsy rate	Bernoulli	60%	Historical data

Table 2 Model deterministic input parameters

Variables	Values
Discount rate	0,3
% affected patients undergoing surgery after receiving BRCA test results	15%
% affected patients undergoing mastectomy before receiving BRCA test results	26%
% affected patients, BRCA-positive, choosing contralateral mastectomy (RRM) and ultrasound follow-up after mastectomy	30%
% affected patients undergoing quadrantectomy before receiving BRCA test results	70%
% affected patients, BRCA-positive, choosing intensive breast screening (intensive follow up) after quadrantectomy (Chance 1a)	20%
% affected patients, BRCA-positive, choosing bilateral mastectomy (RRM) and ultrasound follow-up after quadrantectomy (Chance 1b)	80%
% affected patients undergoing monolateral mastectomy after receiving BRCA test results, if BRCA positive	70%
% affected patients undergoing bilateral mastectomy after receiving BRCA test results, if BRCA positive	30%

The dataset used for estimating the probability of BRCA mutation-positive in affected individuals in the data-driven DSS contains information on female patients with cancer who underwent genetic testing to detect mutations in the BRCA1 and BRCA2 genes during the period 2004-2019 in the Puglia region. The dataset was obtained by extraction from Datawarehouse and Paper Clinical Records. All data were provided either by laboratories performing the genetic analysis on-site or by pathology clinicians (oncologists, gynecologists) who requested the genetic analysis from laboratories outside the region. In particular, data were collected from four institutions, IRCCS Cancer Institute "Giovanni Paolo II" in Bari, Policlinico of Bari, Ospedale Riuniti in Foggia, and PO

Vito Fazzi Hospital in Lecce. Ultimately, the dataset initially considered contains information on 2,256 patients from the Puglia region in Italy. In Table 3 the schematization of the attributes and the typology of data are reported. In detail, the attributes we worked on are the province of birth and BRCA test outcome. The province of birth allowed us to characterize the results of the data-driven to show that with the BI we could reach more precise outcomes. The BRCA test outcome has three possible outcomes: Carriers or "C", Non-Carriers or "NC" and Variant of Uncertain Significance or "VUS".

Attribute	Type	Values And Meaning
Patient condition	Binomial categorical	Identify whether the patient is healthy or sick.
Sex	Binomial categorical	F=female; M=male
Date of birth	Range numeric	day/month/year
Place of birth	Nominal categorical	Municipalities of Puglia or other Italian regions
Residence	Nominal categorical	Municipalities of Puglia or other Italian regions
Age at diagnosis	Numeric ratio	Age at which a tumor was contracted
Post-test year	Numeric ratio	Year in which the patient received the result of the test
Histotype	Nominal categorical	Result of histological examination related to the location of the neoplasm
Neoplasm place	Nominal categorical	Where the tumor is located
Outcome Test BRCA	Nominal categorical	C(*Carrier*)=carrier of a pathogenic mutation in one of the two genes; *VUS* (a *variant of uncertain significance*) = carrier of a mutation of uncertain meaning in one of the two genes; NC (*Non-Carrier*) = non-carrier
BRCA1	Nominal categorical	Alphanumeric mutation identification code in the BRCA1 gene
BRCA2	Nominal categorical	Alphanumeric mutation identification code in the BRCA2 gene

Table 3 Attributes of the database

We selected 1,873 of the 2,255 individuals available according to the province of birth, the 382 excluded individuals were born either in other Italian regions or in other countries. Of these 1,873, for this study, it was necessary to calculate the input variable as the ratio between the number of individuals found to be carriers at the genetic test, in the table indicated with "C", and the total number of individuals affected by an oncological disease, all divided by province. In this paper, in order to calculate the "Probability of BRCA mutation in affected patients" we focused our attention on the "C" outcome of the test, and we calculate the probability as the frequency of occurrence. The results obtained are summarised in the table 4.

Table 4 Data-driven DSS: Probability of BRCA mutation in affected patients.

		BRCA Test Outcome			**Total**	**Probability of BRCA mutation in affected patients**
		C	**NC**	**VUS**		
Provinces	**Bari**	151	532	56	739	20.43%
	BAT	74	94	4	172	43.02%
	Brindisi	35	80	6	121	28.93%

	Foggia	16	18	5	39	41.03%
	Lecce	137	378	40	555	24.68%
	Taranto	66	173	8	247	26.72%
Region	**Puglia**	479	1275	119	1873	25.57%

Step 3: Making a comparison between experience-driven version and data-driven version.

As the final step, we run the two versions of the DSS and made a comparison. To this aim, we develop a plan of experiments designing different scenarios by modifying two parameters: the percentage of patients who have undergone genetic counselling and the percentage of patients who undergo BRCA testing. Genetic counselling means that women who have reached the age of 40 can undergo an examination in which geneticists and oncologists try to reconstruct the family tree of hereditary genetic diseases and identify the risk associated with BRCA gene mutations. The BRCA test detects the BRCA1 and BRCA2 gene mutations and identifies carriers. The mutations are of different types, and in the case of VUS, the relationship between these mutations and BC is still unclear [34]. The comparison was made in three scenarios:

- First Scenario (Baseline): BC BRCA patients are 45% likely to have genetic counselling and 45% likely to have BRCA testing, as it is in the current practice.
- Second Scenario: All patients have genetic counselling, while the genetic testing has been performed as in the current practice (the probability of BRCA remains at 45%).
- Third Scenario: Finally, the third scenario was constructed on the assumption that all patients perform genetic counselling and BRCA genetic testing.

In order to take into account, the uncertainties that characterize the input data, the Monte Carlo simulation has been used. It is a numerical method that can consider multiple sources of uncertainty in the estimation and decision problems, as they are in the actual environment [35]. The simulation was done in the @Risk for Excel environment, with 1000 sample iterations of the identified variables.

To make the comparison between the two versions of the DSS and among the results of the data-driven version to understand if the differentiation carried out by provinces of the Puglia region led to benefits in the decision, we compared the results looking for a statistically significant difference in the different scenarios. The methodology applied to analyse the difference among the results in statistical terms uses the definition of "confidence interval" [36]. We calculated the confidence interval associated with a confidence level of 95% for the data-driven version (for the region of Puglia and the provinces of the region of Puglia) and the experience-driven version.

3. Results

The simulation of the first two scenarios did not lead to a statistically significant difference between the two versions. Therefore, it is possible to state that there is no difference between the use of the two versions. The results obtained will be presented in Appendix B.

The results in the third scenario were much more pronounced. The data-driven version achieved a mean net saving per patient equal to € 7,783.12, up to a maximum value of € 45,411.96, and a probability of getting saving 77.1%. For the experience-driven model we obtained a mean net saving per patient of € 6,360.68, up to a maximum value of € 39,000.9, and the probability of getting savings of 75.7%, as summarised in the table 5.

Table 5 Results of the third scenario.

Version	Place	Variable	Mean	Max	Std. Dev.	Prob. Of Savings
Data-driven	**Puglia region**	25.57%	€ 7,783.12	€ 45,411.96	€ 8,003.45	77.1%
	Province of Bari	20.43%	€ 6,922.94	€ 45,168.89	€ 6,922.94	77.8%
	Pr. of BAT	43.02%	€ 9,549.42	€ 45,641.79	€ 8,893.74	81.5%
	Pr. of Brindisi	28.93%	€ 7,872.31	€ 48,641.00	€ 7,834.62	77.3%
	Pr. of Foggia	41.03%	€ 9,114.14	€ 38,115.22	€ 8,866.50	77.7%
	Pr. of Lecce	24.68%	€ 7,412.29	€ 45,911.38	€ 7,563.05	76.9%
	Pr. of Taranto	26.72%	€ 7,919.19	€ 40,299.14	€ 7,770.73	78.3 %
Experience-driven	**Not specified**	Uniform Distribution (10%-20%)	€ 6,360.68	€ 39,000.97	€ 6,458.55	75.7%

The statistical difference analysis in the table, in this case, provides a positive result, confirming the statistical difference between the two versions.

Table 6 Analysis of the statistical difference for the third scenario.

Version	Place	Variable	Mean	Std dev	Confidence	Lower bound	Upper bound
Data-driven	**Puglia region**	25.57%	€ 7,783.12	€ 8,003.45	€ 496.05	€ 7,287.07	€ 8,279.17
	Province of Bari	20.43%	€ 6,922.94	€ 6,922.94	€ 429.08	€ 6,493.86	€ 7,352.02
	Pr. of BAT	43.02%	€ 9,549.42	€ 8,893.74	€ 551.23	€ 8,998.19	€ 10,100.65
	Pr. of Brindisi	28.93%	€ 7,872.31	€ 7,834.62	€ 485.59	€ 7,386.72	€ 8,357.90
	Pr. of Foggia	41.03%	€ 9,114.14	€ 8,866.50	€ 549.54	€ 8,564.60	€ 9,663.68
	Pr. of Lecce	24.68%	€ 7,412.29	€ 7,563.05	€ 468.75	€ 6,943.54	€ 7,881.04
	Pr. of Taranto	26.72%	€ 7,919.19	€ 7,770.73	€ 481.63	€ 7,437.56	€ 8,400.82
Experience-driven	**Not specified**	Uniform Distribution (10%-20%)	€ 6,360.68	€ 6,458.55	€ 400.30	€ 5,960.38	€ 6,760.98

4. Conclusion

Over the last few years and especially during the pandemic, people all over the world have been forced to use technology platforms and tools to overcome the problems caused by social distance and to carry out their daily activities: work, study, medical visits, and shopping. In this general context, people are unknowingly generating a large amount of data, which can be used in different sectors and for different purposes. The use of data to make decisions is revolutionizing and generating a renaissance in many sectors. In the health sector, the use of data is increasing interest in different areas of scientific research: computer science, management, and medicine. The literature review shows that this interest has been mainly reserved for the application of Business intelligence to improve the ICT architectures and data management systems or to support prognoses and diagnoses, while the usage of BI for improving the management of healthcare processes is still overlooked. There seem to be no studies investigating whether a decision support system (DSS) based on the exploitation of data through BI can outperform traditional experience-driven practices for managing processes in the healthcare domain. So, the research question of this paper was: "Can data-driven DSS improve the healthcare process management better than DSS based solely on experience and literature?". In order to answer this research question, we develop a DSS for reducing the overall costs of a specific healthcare process in the oncology field. The DSS was developed in two versions: the first is experience-driven (i.e., based only on scientific and historical literature data), while the second is data-driven (i.e., based on additional information coming from hospital data). In order to make the comparison, we also develop a plan of experiments designing different scenarios by modifying two parameters: the percentage of patients who have undergone genetic counselling and the percentage of patients who undergo BRCA testing. The results obtained for the first two scenarios have shown that there is no difference among the two versions of the DSS, where it is not possible to say that the data-driven version obtained distinct results than the experience-driven one. We hypothesized that the reason for these results is related to how we constructed the DSS. The input variables related to genetic counselling and BRCA genetic testing are a fundamental resource for the DSS. The scarce presence of data related to genetic counselling and testing leads the model not to perform at its best. This may have led to a non-statistically significant difference between the two versions.

While in the third scenario it was possible to state that there was a statistically significant difference between the results (95% confidence level). This result allowed the research question to be answered, as the difference in effectiveness in terms of results obtained is based on the use of data extracted for the single case as provided by BI and not on the generalization of global cases, which is of fundamental importance given the link between BRCA genetic mutations and ethnicity of origin. In fact, in the case considered, it emerges that in the hypothesis of extending genetic counselling and testing to all patients, the results obtained by a data-driven DSS are better than those obtained by an experience-driven because they allow making diversified decisions, and therefore more effective, depending on the characteristics of the starting population. In other words, by using a data-driven approach it is possible to understand if the decision to extend the test and counselling to all patients is efficient, thus allowing the decision-maker to make informed decisions and avoid wasting money. The results obtained highlight the importance of using specific real data, which can be obtained through the application of BI techniques. The idea of substituting only one of the input variables allowed us to understand how the model varies in the two versions, and especially the sensitivity to variation. In general, the fundamental importance of the input data from genetic counselling and genetic testing emerged, in fact, the way the model structure was constructed without these data we cannot obtain results due to the absence of material on which the model can work.

The main academic implications are related to the use of decision support systems in conjunction with Business Intelligence. Computer science techniques and managerial techniques for strategic decision support can be seen as a single entity, especially after the exponential generation of data. The approach suggested through this paper sees data-driven decision support in conjunction with managerial implications. This research aims to fill the gap in the healthcare industry about the use of BI to improve the decision making process in the oncology domain, by investigating whether a decision support system (DSS) based on the exploitation of data through BI can outperform traditional experience-driven practices for managing processes in the healthcare domain. The results of our study demonstrate that the usage of BI for managing the healthcare processes proved to improve traditional activities and processes mostly based only on the physicians' experience.

Moreover, the implications are two folds. From a managerial point of view, we demonstrated that BI could improve the management of the decision-making process by providing doctors with a mapping of all possible pathways thus helping them in making the best decision. From a business point of view, we demonstrated that BI can also improve the effectiveness of the decision-making process thus leading to financial savings. Clinicians in this case can quickly consult the DSS to identify the decision that will save money on the treatment pathway with the same clinical effectiveness. The savings made by using the DSS can be managed to expand the clinical offerings of hospitals and cut unnecessary waste of money. Furthermore, for this model in detail, the results were very satisfactory when the outcome of genetic counselling and BRCA genetic testing was known, so it is suggested for its users to have a database and to suggest patients to perform these clinical tests to have data to work on.

Nowadays, several people are at the same time fascinated and scared about the exploitation of new technologies based on data and artificial intelligence (AI) [37]. The question is that most people make confusion among big data algorithm and artificial intelligence. The aims of these two powerful tools are completely different, i.e., the big data algorithms are used to improve the cognitive and calculation capability of humans, on the other hand, AI tries to mimic the capability of humans [38]. People are scared about the "ethic-choices" that an AI technology can take and if it is reliable. In this paper, we built the model and used BI as a "decision making support" for the physicians to reach an improved comprehension of all the possible outcomes, and not as a substitution of the decision-makers. So, from our point of view, the doctors keep full control of the decision making process by achieving a full comprehension of the problem.

The limitations of this work are strictly related to the specificity of the application of the model and to the fact that the clinical efficacy of individual treatment pathways has not been investigated in detail as it is a subject of discussion among physicians in the field, and we focused only on the business-management point of view. Besides, the model could be further improved by considering not only the views of experts but also those of patients, to manage the whole healthcare decision-making processes. For future research developments, it is suggested to extract other input variables that could further improve the model. Besides, according to the very limits of the model, a study on the clinical effectiveness of the various clinical pathways could be conducted involving doctors specialized in the oncology branch and involving a sample of patients to get their point of view. Furthermore, complex learning models that can be used to achieve higher levels of performance were not used for this paper. Therefore, a future research development sees the use of such models (neural networks, support vector machines, etc.).

Appendix A

Table A.1 Model input parameters: costs.

Activity	Cost (€)	Notes	Reference
Quadrantectomy	2,354.00	Without complications	NHS: DRG code 259
	2,717.00	With complications	NHS: DRG code 260
Intensive breast screening (intensive follow up)	263.31	mammography and breast magnetic resonance imaging (MRI)	NHS: DRG codes 87371 – 88929 – 897
Biopsy	52.08	core-biopsy	NHS: DRG code 85111
Mastectomy including reconstructive surgery	8,265.00	Without complications	NHS: DRG codes 258 - 461
	8,872. 00	With complications	NHS: DRG codes 257 - 461
Bilateral mastectomy including reconstructive surgery	16,530.00	Without complications	NHS: DRG codes 258 - 461
	17,744.00	With complications	NHS: DRG codes 257 - 461
Ultrasound follow-up	56.55	Breast examination and ultrasound	NHS: DRG codes 88731 - 897
Surgery for local recurrences (skin or lymph node recurrences)	4,583.00		NHS: DRG code 19881
Plastic surgery after complications or for breast implant replacement after 15 years	4,924.00		NHS: DRG code 461
Radiotherapy	2,936.00	cost per regimen in combination with systemic therapy	NHS: DRG code 409
Genetic counseling	20.76		NHS: DRG code MANCA
BRCA testing	1,107.00		Primary data collection

Appendix B

Baseline Scenario (Probability of affected patient going to genetic counselling = 45%, Probability of affected patient getting BRCA test = 45%)

The simulation was run once for the experience-driven model and each province of the region and the region the data-driven model. A table summarising the results of the decision support system is presented below. The experience-driven model has a value of the variable "Probability of BRCA mutation-positive in affected individuals" an element extracted during the Monte Carlo simulation from a uniform distribution in which the minimum is 10% and the maximum is 20%. The average value of savings using the DSS is € 1,388.50 while the maximum value is € 26,761.02, with a probability that the savings are greater than 0 of 16.3%. The reference that should be observed is that of the data-driven model for the region of Puglia, as it summarises the results obtained from the seven simulations: average value of savings of € 1,568.76, the maximum value of € 38,371.47, and a probability of savings greater than 0 of 16.4%, summarized in the table B.1. At first sight, the results in this scenario are significantly different, but to state this with certainty we carried out a statistical analysis.

Table B.1 Results of the first scenario.

Version	Place	Variable	Mean	Max	Std. Dev.	Prob. Of Savings
Data-driven	**Puglia region**	25.57%	€ 1,568.76	€ 38,371.47	€ 4,602.99	16.4%
	Province of Bari	20.43%	€1,377.95	€ 42,363.79	€ 4,417.11	14.5%
	Pr. of BAT	43.02%	€ 1,811.95	€ 39,473.89	€ 5,057.70	17.01%.
	Pr. of Brindisi	28.93%	€ 1,555.72	€ 32,989.08	€ 4,425.92	16.5 %
	Pr. of Foggia	41.03%	€ 1,887.48	€ 41,199.38	€ 5,400.93	17.2%
	Pr. of Lecce	24.68%	€ 1,514.16	€ 38,614.40	€ 4,530.43	15.9%
	Pr. of Taranto	26.72%	€ 1,555.91	€ 40,508.81	€ 4,760.37	14.8%
Experience-driven	**Not specified**	Uniform Distribution (10%-20%)	€ 1,388.50	€ 26,761.02	€ 3,836.37	16.3%

The statistical significance analysis between the two models was carried out following the model identified in a scientific paper [36]. The results in the table show that it cannot be said that there is a significant statistical difference given a significance α=0.05 and the number of observations equal to the iterations of the simulation, i.e. 1000. There is no statistical significance because the comparison of the two confidence intervals for the data-driven and experience-driven model predicts a strong overlap, showed in the table B.2.

Table B.2 Analysis of the statistical difference for the first scenario.

Version	Place	Variable	Mean	Std dev	Confidence	Lower bound	Upper bound
Data-driven	**Puglia region**	25.57%	€ 1,568.76	€ 4,602.99	€ 285.29	€ 1,283.47	€ 1,854.05
	Province of Bari	20.43%	€ 1,377.95	€ 4.417.11	€ 273.77	€ 1,104.18	€ 1,651.72
	Pr. of BAT	43.02%	€ 1,811.95	€ 5.057.70	€ 313.47	€ 1,498.48	€ 2,125.42
	Pr. of Brindisi	28.93%	€ 1,555.72	€ 4.425.92	€ 274.32	€ 1,281.40	€ 1,830.04
	Pr. of Foggia	41.03%	€ 1,887.48	€ 5.400.93	€ 334.75	€ 1,552.73	€ 2,222.23
	Pr. of Lecce	24.68%	€ 1,514.16	€ 4.530.43	€ 280.79	€ 1,233.37	€ 1,794.95

	Pr. of Taranto	26.72%	€ 1,555.91	€ 4.760.37	€ 295.05	€ 1,260.86	€ 1,850.96
Experience-driven	Not specified	Uniform distribution (10%-20%)	€ 1,388.50	€ 3.836.37	€ 237.78	€ 1,150.72	€ 1,626.28

Second Scenario (Probability of affected patient going to genetic counseling = 100%, Probability of affected patient having BRCA test = 45%)

How the simulations were carried out are similar to those of the first scenario and summarised in the table. The results also in this case seem different at first analysis. In detail, the results obtained from the experience-driven model are an average saving of € 2,982.77, a maximum value of savings of € 36,398.22, and a probability of savings greater than zero equal to 34.2%, while the data-driven model led to an average value of € 3,593.96, a maximum value of € 37,497.16 and a probability of savings greater than zero equal to 34.6%, summarised in the table 10.

Table B.3 Results of the second scenario.

Version	Place	Variable	Mean	Max	Std. Dev.	Prob. Of Savings
Data-driven	Puglia region	25.57%	€ 3,593.96	€ 37,497.16	€ 6,640.63	34.6%.
	Province of Bari	20.43%	€ 3,011.27	€ 32,731.45	€ 5,307.66	35.1%
	Pr. of BAT	43.02%	€ 4,118.40	€ 37,665.96	€ 7,368.89	36.6%,
	Pr. of Brindisi	28.93%	€ 3,306.63	€ 40,950.40	€ 6,812.43	34.2%
	Pr. of Foggia	41.03%	€ 4,109.61	€ 43,225.16	€ 7,649.24	32.2%
	Pr. of Lecce	24.68%	€ 3,883.75	€ 38,186.79	€ 5,959.52	36.6%
	Pr. of Taranto	26.72%	€ 3,872.25	€ 39,894.5	€ 6,926.86	36.1%
Experience-driven	Not specified	Uniform Distribution (10%-20%)	€ 2,982.77	€ 37,497.16	€ 5,564.84	34.6%.

The analysis of whether the difference between the two models was statistically significant was again negative in the table 11, mainly because a very high confidence level was preferred. Please refer to the methodology section for a more detailed analysis of the results.

Table B.4 Analysis of the statistical difference for the second scenario.

Version	Place	Variable	Mean	Std dev	Confidence	Lower bound	Upper bound
Data-driven	Puglia region	25.57%	€ 3,593.96	€ 6,640.63	€ 411.58	€ 3,182.38	€ 4,005.54
	Province of Bari	20.43%	€ 3,011.27	€ 5,307.66	€ 328.97	€ 2,682.30	€ 3,340.24
	Pr. of BAT	43.02%	€ 4,118.40	€ 7,368.89	€ 456.72	€ 3,661.68	€ 4,575.12
	Pr. of Brindisi	28.93%	€ 3,306.63	€ 6,812.43	€ 422.23	€ 2,884.40	€ 3,728.86
	Pr. of Foggia	41.03%	€ 4,109.61	€ 7,649.24	€ 474.10	€ 3,635.51	€ 4,583.71
	Pr. of Lecce	24.68%	€ 3,883.75	€ 5,959.52	€ 369.37	€ 3,514.38	€ 4,253.12
	Pr. of Taranto	26.72%	€ 3,872.25	€ 6,926.86	€ 429.32	€ 3,442.93	€ 4,301.57
Experience-driven	Not specified	Uniform distribution (10%-20%)	€ 2,982.77	€ 5,564.84	€ 344.91	€ 2,637.86	€ 3,327.68

References

[1] S. Fernandez, P. Jenkins e B. Vieira, «McKinsey & Co.,» 15 December 2020. [Online]. Available: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/europes-digital-migration-during-covid-19-getting-past-the-broad-trends-and-averages.

[2] M. Ienca e E. Vayena, «On the responsible use of digital data to tackle the COVID-19 pandemic,» *Nature Medicine,* vol. 26, p. 463–464, 2020.

[3] A. Bosworth, J. Ruhter, L. Samson, S. Sheingold, C. Taplin e W. Tarazi, «MEDICARE BENEFICIARY USE OF TELEHEALTH VISITS: EARLY DATA FROM THE START OF THE COVID-19 PANDEMIC,» *ASPE (Assistant Secretary for Planning and Evaluation),* 2020.

[4] G. Yilmaz, M. Rapin, D. Pessoa, B. M. Rocha, A. M. d. Sousa, R. Rusconi, P. Carvalho, J. Wacker, R. P. Paiva e Olivier Chételat, «A Wearable Stethoscope for Long-Term Ambulatory Respiratory Health Monitoring,» *Sensors,* 2020.

[5] É. Foley e M. Guillemette, «What is Business Intelligence?,» *IJBIR,* 2010.

[6] G. M. Sechi, M. Migliori, G. Dassi, A. Pagliosa, R. Bonora, A. Oradini-Alacreu, A. Odone, C. Signorelli e A. Zoli, «Business Intelligence applied to Emergency Medical Services in the Lombardy region during SARS-CoV-2 epidemic,» *Acta Biomed,* pp. 39-44, 2020.

[7] E. R. Safwan, R. Meredith e F. Burstein, «Business Intelligence (BI) system evolution: a case in a healthcare institution,» *Journal of Decision Systems,* pp. 25:sup1, 463-475, 2016.

[8] B. Patil, R. Joshi e D. Toshniwal, «Association Rule for Classification of Type-2 Diabetic Patients,» in *Second International Conference on Machine Learning and Computing.*, 2010.

[9] K. Liu e C. Lu., «BBN-Based Decision Support for Health Risk Analysis,» in *Fifth International Joint Conference on INC, IMS and IDC*, 2009.

[10] M. R. Llave, «A Review of Business Intelligence and Analytics in Small and Medium-Sized Enterprises,» *International Journal of Business Intelligence Research,* vol. 10, n. 1, pp. 19-41, 2019.

[11] D. Larsona e V. Chang, «A review and future direction of agile, business intelligence, analytics,» *International Journal of Information Management,* vol. 36, pp. 700-710, 2016.

[12] A. Uriarte, E. Zúñiga, M. Moris e N. Amos, «How can decision makers be supported in the improvement of an emergency department? A simulation, optimization and data mining approach,» *Operations research for health care,* 2017.

[13] M. I. Mashinchi, A. Ojo e F. J. Sullivan, «Analysis of Business Intelligence Applications in Healthcare Organizations,» *HICSS ,* 2019.

[14] A. Methaila, P. Kansal, H. Arya e P. Kumar, «EARLY HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES,» *CSE 2014.,* 2014.

[15] A. Ahmad, I. Muraina, R. Razak e M. Abdullah, «Better Healthcare: Exploring Business Intelligence for Healthier Malaysian Rural Dwellers.,» 2013.

[16] O. Ali, A. Nassif e L. Capretz, «Business intelligence solutions in healthcare a case study: Transforming OLTP system to BI solution,» in *2013 Third International Conference on Communications and Information Technology (ICCIT),*, 2013.

[17] N. Mustafee, J. Powell, S. Martin, A. Fordyce e A. Harper, « Investigating the use of real-time data in nudging patients,» *Emergency department (ED) attendance behaviour, Torbay Hospital.,* 2017.

[18] J. Almeida, «A disruptive Big data approach to leverage the efficiency in management and clinical decision support in a Hospital,» *Porto biomedical journal, 1 1,* pp. 40-42., 2016.

[19] D. Gonçalves, M. Santos e J. Cruz, «Analysis of the Quality of Life after an Endoscopic Thoracic Sympathectomy: A Business Intelligence Approach.,» in *2010 Second International Conference on Advances in Databases, Knowledge, and Data Applications,1-6*, 2010.

[20] B. Cooper, H. Watson, B. Wixom e D. Goodhue, «Data Warehousing Supports Corporate Strategy at First American Corporation,» *MIS Quarterly,* vol. 24, n. 4, pp. 547-567, 2000.

[21] V. Vinekar, J. Teng e A. Chennamaneni, «The Interaction of Business Intelligence and Knowledge Management in Organizational Decision-Making,» *Journal of International Technology and Information Management,* vol. 18, n. 2, pp. 143-159., 2009.

[22] I. S. KoSarvar e R. Abdullaev, «A Study on the Aspects of Successful Business Intelligence System Development,» *International Conference on Computational Science,* pp. 729-732, 2007.

[23] M. Karami, M. Fatehi, M. Torabi, M. Langarizadeh, A. Rahimi e R. Safdari, «Enhance hospital performance from intellectual capital to business intelligence,» *Radiology management,* 2013.

[24] A. Mehrgou e M. Akouchekian, «The importance of BRCA1 and BRCA2 genes mutations in breast cancer development,» 15 May 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4972064/#. [Consultato il giorno 20 December 2020].

[25] F. Carol, F. Debra, D. K. Shelley e Q. Ruben, «A systematic review of international guidelines and recommendations for the genetic screening, diagnosis, genetic counseling, and treatment of BRCA-mutated breast cance,» *Cancer Management and Research,* vol. 11, pp. 2321-2337, 2019.

[26] P. B. Van Der Nat, D. v. Veghel, E. Daeter, H. J. Crijns, J. Koolen, S. Houterman, M. A. Soliman e B. A. d. M. &. M. B. S. Group, «Insights on value-based healthcare implementation from Dutch heart care,» *International Journal of Healthcare Management,* vol. 13, n. 3, pp. 189-192, 2020.

[27] C. Gamble, L. Havrilesky, M. E.R., J. Chino, S. Hollenbeck, J. Plichta, P. Marcom, S. Hwang, N. Kauff e R. Greenup, «Cost Effectiveness of Risk-Reducing Mastectomy versus Surveillance in BRCA Mutation Carriers with a History of Ovarian Cancer,» *Ann Surg Oncol,* 2017.

[28] M. J. Hall, J. E. Reid, L. A. Burbidge, D. Pruss, A. M. Deffenbaugh, C. Frye, R. J. Wenstrup, B. E. Ward, T. A. Scholl e W. W. Noll, «BRCA1 and BRCA2 mutations in women of different ethnicities undergoing testing for hereditary breast-ovarian cancer.,» 2009. [Online]. Available: https://doi.org/10.1002/cncr.24200. [Consulted 1 October 2020].

[29] «BRCA Gene Mutations: Cancer Risk and Genetic Testing,» National Cancer Institute, 19 November 2020. [Online]. Available: https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet. [Consulted 20 December 2020].

[30] K. Malone, J. Daling, D. Doody, L. Hsu, L. Bernstein, R. Coates, P. Marchbanks, M. Simon, J. McDonald, S. Norman, B. Strom, R. Burkman, G. Ursin, D. Deapen, L. Weiss, S. Folger, J. Madeoy, D. Friedrichsen, N. Suter, M. Humphrey, R. Spirtas e E. Ostrander, «Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years,» *Cancer Res.,* vol. 66, n. 16, pp. 297-308, 2006.

[31] M. Palma, E. Ristori, E. Ricevuto, G. Giannini e A. Gulino, «BRCA1 and BRCA2: the genetic testing and the current management options for mutation carriers,» *Crit Rev Oncol Hematol.,* vol. 57, n. 1, pp. 1-23, 2006.

[32] F. Fostira, Saloustros, A. A. P, V. D, K. D, M. D, T. V, G. D, Y. G, F. I e Konstantopoulou, «Germline deleterious mutations in genes other than BRCA2 are infrequent in male breast cancer,» *Breast Cancer Res Treat,* vol. 169, n. 1, pp. 105-113, 2018.

[33] H. Tuffaha, A. Mitchell, R. Ward, L. Connelly, J. Butler, S. Norris e P. Scuffham, «Cost-effectiveness analysis of germ-line BRCA testing in women with breast cancer and cascade testing in family members of mutation carriers,» *Genet Med,* vol. 20, n. 9, pp. 985-994, 2018.

[34] F. Karami e P. Mehdipour, «A comprehensive focus on global spectrum of BRCA1 and BRCA2 mutations in breast cancer. BioMed research international,» 2013. [Online]. Available: https://doi.org/10.1155/2013/928562. [Consulted 1 October 2020].

[35] J. Mun, «Modeling risk: Applying Monte Carlo simulation, real options analysis, forecasting, and optimization technique,» *John Wiley & Sons.,* p. Vol 347, 2006.

[36] N. Carbonara e R. Pellegrino, «Revenue guarantee in public–private partnerships:,» *Construction Management and Economics,* 2018.

[37] R. McDougall, «Computer knows best? The need for value-flexibility in medical AI.,» *Journal of Medical Ethics,* vol. 3, n. 45, pp. 156-160, 2019.

[38] E. Di Nucci, «Should we be afraid of medical AI?,» *Journal of Medical Ethics,* pp. 1-3, 2019.