Meta analysis of the use of Bayesian networks in breast cancer diagnosis

Metanálise do uso de redes bayesianas no diagnóstico de câncer de mama

Metaanálisis del uso de los modelos bayesianos en el diagnóstico de cáncer de mama

Priscyla Waleska Simões [1]
Geraldo Doneda da Silva [1]
Gustavo Pasquali Moretti [1]
Carla Sasso Simon [1]
Erik Paul Winnikow [1]
Silvia Modesto Nassar [1]
Lidia Rosi Medeiros [1]
Maria Inês Rosa [1]

[1] Universidade do Extremo Sul Catarinense, Criciúma, Brasil.

Correspondence
M. I. Rosa
Laboratório de Epidemiologia, Universidade do Extremo Sul Catarinense.
Av. Universitária 1105, Criciúma, SC 88806-000, Brasil.
mir@unesc.net

Abstract

The aim of this study was to determine the accuracy of Bayesian networks in supporting breast cancer diagnoses. Systematic review and meta-analysis were carried out, including articles and papers published between January 1990 and March 2013. We included prospective and retrospective cross-sectional studies of the accuracy of diagnoses of breast lesions (target conditions) made using Bayesian networks (index test). Four primary studies that included 1,223 breast lesions were analyzed, 89.52% (444/496) of the breast cancer cases and 6.33% (46/727) of the benign lesions were positive based on the Bayesian network analysis. The area under the curve (AUC) for the summary receiver operating characteristic curve (SROC) was 0.97, with a Q* value of 0.92. Using Bayesian networks to diagnose malignant lesions increased the pretest probability of a true positive from 40.03% to 90.05% and decreased the probability of a false negative to 6.44%. Therefore, our results demonstrated that Bayesian networks provide an accurate and non-invasive method to support breast cancer diagnosis.

Medical Informatics; Bayes Theorem; Breast Neoplasms

Resumo

O objetivo deste estudo foi avaliar a acurácia das redes bayesianas no apoio ao diagnóstico de câncer de mama. Foram realizadas revisão sistemática e metanálise, que incluíram artigos e relatórios publicados entre Janeiro de 1990 e Março de 2013. Foram incluídos estudos transversais prospectivos e retrospectivos que avaliaram a acurácia do diagnóstico de lesões de mama (condição alvo) usando as redes bayesianas (teste em avaliação). Quatro estudos primários que incluíram 1.223 lesões de mama foram analisados, 89,52% (444/496) dos casos de câncer de mama e 6,33% (46/727) das lesões benignas foram positivas tendo-se como base a análise das redes bayesianas. A área dentro da curva SROC (característica de operação do receptor sumária) foi 0,97, com um valor Q* de 0,92. O uso de redes bayesianas no diagnóstico de lesões malignas aumentou a probabilidade pré-teste para um verdadeiro positivo de 40,03% para 90,05% e diminuiu a probabilidade de um falso negativo para 6,44%. Portanto, nossos resultados demonstraram que as redes bayesianas oferecem um método acurado e não invasivo no apoio ao diagnóstico de câncer de mama.

Informática Médica; Teorema de Bayes; Neoplasias da Mama

Introduction

Breast cancer is the most common type of cancer in women, and it is also a common cause of cancer-related mortality, both in developing and developed countries. Approximately 1.4 million new cases occurred in 2008 worldwide, representing 23% of all cancers [1,2]. Fortunately, the early detection of breast cancer can improve the chance of successful treatment and recovery [3].

In recent decades, artificial intelligence has become widely accepted in medical applications [4]. One such application is Bayesian networks, which are becoming widely used to represent knowledge domains in the presence of uncertainty from randomness. Bayesian networks can be used as an analysis and decision aid in the interpretation of the results of a diagnostic test (e.g. mammography), signs and symptoms when uncertainty is known to be a dominant factor [5].

A Bayesian network is a graphical model that represents probabilistic relationships among variables of interest [6]. Such networks consist of a qualitative component (i.e., the structural model), which provides a visual representation of the interactions among variables, and a quantitative component (i.e., a set of local probability distributions), which permits probabilistic inference. Together, these components determine the unique joint probability distribution over the variables in a specific problem [7]. In clinical medical practice, professionals can calculate probabilities using Bayes' theorem without a computer for a specific diagnosis with limited parameters (i.e., a few conditional probabilities). If the factors that modify the probability of disease have interactions, however, the complexity of such calculations can increase exponentially, making it difficult to solve without computational support. In this case, Bayesian networks may be useful [8].

In the Bayesian networks, the nodes represent uncertain variables, and typically, there is a primary or "root" node that represents the variable of interest, other nodes impact the probability of that primary node [6,7]. For example, in medical diagnosis, estimating the probability of an event, such as the malignancy of a breast mass ("root" node), given a set of evidence (i.e., demographics, image characteristics, etc.) is a problem that can be solved with Bayesian networks [8]. Patient risk factors, signs, symptoms, and the results of diagnostic test are inputs of the system [8].

Each node necessarily contains mutually exclusive and collectively exhaustive instances [6,7]. Mutually exclusive instances refer to events that cannot occur at the same time. For example, if a coin toss is the variable of interest, heads and tails are the mutually exclusive instances of that variable because they cannot occur simultaneously [8].

When both the structure and probabilities are established, the Bayesian network can be used to determine the probability of one node based upon the available information about other conditionally dependent nodes using an inference algorithm. Inference is the reasoning process used to draw conclusions from available evidence based on the principles of probabilistic reasoning and Bayes' theorem [6,7,8].

This theory may be used to differentiate between benign and malignant breast diseases, using radiologists' descriptions of breast imaging findings, helping to define which patients should be referred for biopsy and which should not be referred for this procedure [9]. These systems can also perform more complex reasoning tasks, such as mammography-histology correlation, and detect sampling error better than radiologists [10]. This technique can also be used to help the general practitioner to determine which breast lesions identified on physical examination should be directed to mammography or for mastologist. It can assist the pathologist to do the pathological diagnosis too.

The purpose of this systematic review was to assess the accuracy of Bayesian networks in patients with breast lesions.

Methods

Search strategy

The adopted search strategy was to perform a search, between January 1990 and March 2014, in the following databases: Medical Literature Analysis and Retrieval System Online (MEDLINE) through PubMed, Cancer Literature (CancerLit), Latin American and Caribbean Health Sciences (LILACS), Excerpta Medical Database (Embase), SciVerse Scopus (Scopus), Cochrane Central Register of Controlled Studies (Cochrane), Spanish Bibliographic Index of Health Sciences (IBECS), Biological Abstracts (BIOSIS), and Web of Science.

We also searched papers in grey literature (which includes Google Scholar, published papers from conferences, government technical reports, and other materials that are not controlled by scientific publishers).

The databases were searched using the keywords included in the Medical Subject Headings (MeSH) and their synonyms, including the following terms: "breast neoplasms", "breast lesions", "breast cancer", "breast tumor", "mammary neo-

plasms", "mammary cancer", and "mammary tumor". The papers with these terms were associated with the evaluated test, "Bayesian network", and with the terms "sensitivity" and "specificity". The full search strategy will be available from the authors upon request.

The "*" symbol was used to allow the recovery of all variations of the original words with suffixes. The above terms were combined using the Boolean operators "AND", "OR", and "NOT".

The search was limited to studies in human females. There was no language restriction. The reference lists from all retrieved primary studies were checked. The references cited in meta-analyses, guidelines, and comments identified in the previously mentioned databases were also checked. We contacted the authors of the studies published with incomplete information; however, we received no responses to the e-mails sent.

Study screening and eligibility

The initial analysis of the abstracts and titles identified by the search strategy in the aforementioned databases was independently performed by three researchers (M.I.R., P.W.S., and G.D.S.); the assessment of English articles was performed by M.I.R., P.W.S., and C.S.S., and articles in other languages were assessed by different reviewers, E.P.W. and G.D.S. with translations performed as necessary. Disagreements about the inclusion or exclusion of each study were initially resolved by consensus; when consensus was not possible, differences were arbitrarily resolved by L.R.M.

Concordance between the reviewers was computed in primary studies using the Kappa Coefficient of Agreement (κ) [11,12,13]. We used the categories proposed by Altman in 1991 [14].

We included cross-sectional studies that used Bayesian networks (i.e., the evaluation test) as diagnostic test to assess breast lesions (i.e., the target conditions). The diagnostic test evaluation consisted of the results provided by the Bayesian networks (i.e., positive or negative).

We analyzed the studies involving women with breast tumors (benign or malignant) who had undergone surgery or core biopsy followed by histological analysis. The reference diagnostic test was the result of histological analysis of paraffin-embedded sections; a breast cancer result by a Bayesian network was considered to be correct if it did not differ from the histological analysis.

For inclusion in our systematic review, the final histological diagnosis of breast lesions had to characterize the lesion as benign or malignant. The diagnostic being considered was the result of the Bayesian networks and was identified as the diagnosis with the highest *a posteriori* probabil-

ity calculated from each study, ie, no matter how small the difference between the probabilities of a diagnosis of cancer against a diagnosis of benign lesion or how big the final probability of cancer; thus, a patient whose probability of having a malignancy of 49% against 51% probability of being a false positive could be classified by the Bayesian network as breast cancer. However, this method has been used successfully in the treatment of false-positives and false-negatives [8].

Borderline lesions were included, allowing the consideration of the applicability of Bayesian networks to the pattern classification boundary [6,7].

In this systematic review, we excluded studies presenting exclusively benign or breast cancer as the reference standard. Thus, the primary outcome analyzed was the accuracy of breast lesion result by Bayesian network.

Data collection and quality evaluation methodology

From the studies, we extracted the data (i.e., design, location, year, sample, average age, and prevalence), outcomes, and Bayesian network architectures, including the qualitative part (i.e., the description of the input and output nodes), the quantitative part (i.e., how to obtain the conditional probabilities), and the software used to build the Bayesian network and perform inference, in addition to the patient characteristics.

M.I.R. and P.W.S. independently abstracted data on the prevalence of benign and malignant breast lesions, study (year), study design, design and setting, number of lesions, age in years, Bayesian network architecture (description of outcomes and inputs) including its design, cut-off conditional and unconditional probabilities, and software used. Other reviewers (E.P.W. and L.R.M.) independently extracted data from the articles published in languages other than English, and translations were performed when necessary. Disagreements regarding the data extraction were resolved by consensus initially, and, when this was not possible, the differences were resolved by S.M.N.

Each reviewer also calculated the pretest probability (i.e., prevalence), sensitivity, specificity, positive and negative likelihood ratios, and posttest probabilities from the primary studies that used Bayesian networks for breast cancer diagnoses. As previously detailed, the studies that did not contain the necessary data to construct a 2x2 contingency table were excluded.

The methodological quality assessment was conducted using the *Quality Assessment of Diagnostic Accuracy Studies* (QUADAS) criteria modified by Cochrane, which consists of 11 study

characteristics that have the potential to introduce bias. These items were classified as positive (without bias), negative (potential bias), or insufficient information [12,13].

Summary of data and statistical analysis

To assess agreement between each study's eligibility and methodological quality, and the consistency between the Bayesian network results and the paraffin and histological results, we calculated the percentage observed using the κ [11,12].

A 2x2 contingency table was constructed for each selected study, and all biopsies were classified as benign or malignant. We calculated the sensitivity, specificity, likelihood ratios, and post-test probability.

For studies in which one cell of the 2x2 contingency table contained a value of zero, 0.5 was added. However, the studies in which zero occurred in two or more cells were excluded from the analysis [13,14].

We aimed to produce pooled estimates of sensitivity and specificity using a meta-analysis, which was performed using the Meta-Disc software version 1.4 (developed by the Clinical Biostatistics Unit, Hospital Ramón y Cajal, Madrid, Spain) [15] and Review Manager (RevMan) version 5.0.21 (developed at The Nordic Cochrane Centre, Copenhagen, Denmark) [16].

A bivariate analysis were used to calculate the pooled estimates of sensitivity, specificity, and likelihood ratios, along with 95% confidence intervals (95%CI) to estimate the values presented in the summarized meta-analysis [17,18,19]. All measurements that were summarized by the random effects model of DerSimonian & Laird [20], which considers the presence of heterogeneity in the studies, were calculated, and all global averages were weighted with 95%CI.

The heterogeneity of the results of sensitivity and specificity across studies was analyzed by the χ^2 distribution with n-1 degrees of freedom, and the inconsistency I^2 [21,22]. The heterogeneity of the positive and negative likelihood ratios of the different studies was examined with the Cochran Q test (establishing the weights of the studies by the inverse of the variance), the χ^2 distribution with n-1 degrees of freedom, the inconsistency (I^2), and χ^2 were used to estimate the variation between the studies [21,22].

As there was no heterogeneity, a summary receiver operating characteristic (SROC) curve was generated considering the data from all thresholds using the method of Littenberg and Moses [22]. This curve can change, depending on the threshold and the ROC curve used to define an abnormal test, thereby resulting in an expected oscillation between sensitivity and specificity.

The SROC curve can be considered an excellent summary graph; however, for the purpose of comparison, we calculated the Q* point as an additional statistic [18,22]. The Q* point is the one at which the sensitivity and specificity are equal in the SROC, and indicates when a test approximates the desired performance of 100% sensitivity and specificity, similar to the area under the curve (AUC) for an receiver operating characteristic (ROC) curve. Higher Q* values are associated with better performance in diagnostic tests [18,22]. The AUC and Q* considered in the SROC curve were estimated using the trapezoidal numerical integration method Meta-Disc [15].

Results

Identification of studies and eligibility

The process of study selection is shown in Figure 1. We identified 100 citations in the searched databases. After the initial review of titles and abstracts, 24 full papers were recovered, four of which were considered to be eligible for our systematic review.

Four primary studies, which included breast lesions from 1,223 women, met the inclusion criteria and were analyzed (as shown in Table 1) [10,23,24,25]. The overall concordance between the eligibility and methodological quality of the studies was 84% ($\kappa = 0.67$), indicating good agreement [12]. There was disagreement, which was resolved by consensus, between the reviewers on the inclusion and exclusion criteria in these four studies.
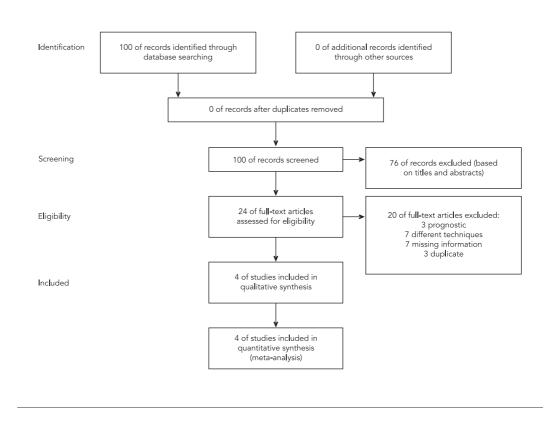
• **Study descriptions**

Table 1 provides detailed descriptions of the studies, tests, and standards used. The mean participant age was found in only one study [10]. Two studies were performed in the United States [10,24], one in Ireland [23], and another in England [25].

Breast cancer was found in 496 cases (prevalence of 40.56%, i.e., the number of positive cases among all individuals included in the studies, considering the positive and negative cases), and 727 (59.44%) patients had benign lesions. Table 2 summarizes the results for all studies included in the meta-analysis.

The Bayesian network documented in these studies had the same possible outcomes (i.e., benign and malignant); however, the nodes representing the input variables showed different characteristics relevant to breast cancer diagnosis.

Figure 1

Study selection process.

The Bayesian network in Kahn et al. [24] used information from the patients' histories, such as the physical exam and mammography results, and the conditional probabilities were obtained by reviewing the indexed medical literature, census data and statistical health reports, i.e., conditional probabilities for architectural distortion, previous biopsy at the same site were estimated as well; values for demographic variables were derived from published epidemiological data; and statistical studies published in peer-reviewed radiology journals provided most of the data for knowledge base, such as values of conditional probabilities findings and mammographic findings for breast cancer. The software used for learning conditional probabilities was BNG [26], and IDEAL was used for inference [27].

The Bayesian network documented by Hamilton et al. [23] used 10 cytological characteristics as input information obtained by fine needle aspiration cytology (FNAC), and the conditional probability matrices relating each of the diagnostic clues and their outcomes to diagnosis (benign, malignant) were defined by a cytopathologist. The software for generating these Bayes-

ian networks was developed at the University of Arizona (USA).

The Bayesian network developed by Cruz-Ramírez et al. [25] considered the patients' age and 10 cytological features obtained by FNAC; the conditional probabilities were obtained in an automated way from two databases [28,29,30]. These databases come from the field of pathology, regarding the cytodiagnosis of breast cancer using FNAC of the breast lesion; the first database, collected retrospectively by a single observer with 10 years' experience of reporting FNAC, during 1992-1993. The second database, collected prospectively by 19 observers with 5-20 years' experience of reporting FNAC, contains 322 consecutive adequate specimens, during 1996-1997. The software used for learning conditional probabilities was the Power Constructor (CBL2-Learning) [31,32], and Netica software was used for inference [33].

The Bayesian network presented by Burnside et al [10] consisted of 25 hierarchical descriptors of BI-RADS classification [34,35], and the conditional probabilities were obtained from indexed medical literature, i.e., the BI-RADS descriptors links

Table 1

Characteristics of the studies included in the systematic review.

Characteristic	Study (year)				Total
	Kahn Jr. et al. [24] (1997)	Hamilton et al. [23] (1994)	Cruz-Ramírez et al. [25] (2007)	Burnside et al. [10] (2004)	
Design and setting	Cross-sectional; USA	Cross-sectional; Northern Ireland	Cross-sectional; England	Cross-sectional; USA	
N *	77	40	1,014	92	1,223
Age in years (SD/ range)	NA	NA	NA	58.2 (±10.5) (range, 37-84 years)	
Breast cancer (prevalence %)	25 (32.47)	19 (47.50)	423 (41.72)	29 (31.52)	496 (40.56)
Architecture Qualitative (nodes)					
Outcomes	Benign/Malignant	Benign/Malignant	Benign/Malignant	Benign/Malignant	
Inputs	Five patient history items **; Two physical findings ***; 15 mammographic findings	10 cytological features	Age; 10 cytological features	25 hierarchical descriptors of BI-RADS	
Quantitative					
Conditional and unconditional probabilities	Peer-reviewed medical literature #; Census data; Health statistics reports	One cytopathologist ##	Two databases ###	Peer-reviewed medical literature #	
Software	BNG [26] (learning); IDEAL [27] (inference)	Developed in the Optical Sciences Center, University of Arizona, USA	Power constructor (CBL2-learning) [31,32]; Netica (inference) [33]	GeNIe Modeling Environment developed by the Decision Systems Laboratory of the University of Pittsburgh, USA [36]	

BI-RADS: Breast Imaging-Reporting and Data System; NA: not available.

* Number of lesions (malignant, benign and normal tissue);

** Age at menarche, age at first live birth, number of first-degree relatives with breast cancer, previous biopsy;

*** Nipple discharge, architectural distortion;

The authors of the original references used data from the medical literature to fill the parameters of conditional probabilities (dependent relationship

between the variable and the outcome) and unconditional probabilities (prevalence of breast cancer) of the architecture of Bayesian network used;

A medical specialist who informed, through interviews, conditional probabilities (dependent relationship between the variable and the outcome) and unconditional probabilities (prevalence of breast cancer) of the architecture of Bayesian network used;

These databases come from the field of pathology, regarding the cytodiagnosis of breast cancer using fine needle aspiration cytology (FNAC) of the breast lesion; the first database, collected retrospectively by a single observer with 10 years' experience of reporting FNAC, during 1992-1993. The second database, collected prospectively by 19 observers with 5-20 years' experience of reporting FNAC, contains 322 consecutive adequate specimens, during 1996-1997. Reference test: Histology.

Table 2

2x2 contingency table.

Study (year)	Bayesian network				Sensitivity	Specificity
	Biopsy positive		Biopsy negative			
	TP	FN	FP	TN		
Kahn Jr. et al. [24] (1997)	23	2	6	46	92.0 (74.0-99.0)	88.5 (76.6-95.6)
Hamilton et al. [23] (1994)	17	2	0	21	89.5 (66.9-98.7)	100.0 (83.9-100.0)
Cruz-Ramírez et al. [25] (2007)	381	42	36	555	90.1 (86.8-92.7)	93.9 (91.7-95.7)
Burnside et al. [10] (2004)	23	6	4	59	79.3 (60.3-92.0)	93.7 (84.5-98.2)
Total	**444**	**52**	**46**	**681**	**89.5 (86.5-92.1)**	**93.7 (91.7-95.3)**

FN: false negative; FP: false positive; TN: true negative; TP: true positive.

with diseases of the breast applying probabilities derived from the literature. The software used in the development of these Bayesian networks was the GeNIe Modeling Environment developed at the University of Pittsburgh (USA) [36].

We seek with this systematic review to present the use of Bayesian networks as a method of decision support and diagnosis of breast cancer; therefore, despite differences in input variables of Bayesian networks of studies included in the meta-analysis, all showed the same Bayesian networks outcome, were used as decision support and diagnosis, and used Bayes' theorem to calculate inference and a posteriori probability.

• **Methodological quality assessment**

Methodological quality assessments of the studies were performed according to a modified version of QUADAS [11,12] and are illustrated in Table 3.

The reviewers disagreed on 3 of the 11 items; this disagreement was resolved by consensus. The presence of withdrawals from the sample was not clarified in all studies; additionally, it was not possible to determine whether data interpretation was conducted in a blinded manner in the included studies. Fifty percent of the studies did not describe whether the clinical information available was that used in clinical practice, and the remainder showed bias on that item. Additionally, 25% of the studies did not provide sufficient information to enable an assessment of whether the sample was representative, that is, whether they considered the patients receiving routine tests.

Three studies performed well and received a positive rating on at least 8 of the 11 items [10,24,25]. The inter-observer agreement in the analysis of methodological quality with QUADAS was 94%

($\kappa = 0.86$), indicating a good agreement [14]. As described above, all disagreements were resolved by consensus.

• **Summary of diagnostic performance**

The overall inter-observer agreement between the Bayesian networks and the paraffin examination was 92.4% (95%CI: 90.9%-93.9%) ($\kappa = 0.84$), indicating a good agreement [14]. The overall sensitivity was 89.5% (95%CI: 86.5%-92.1%), and the specificity was 93.7% (95%CI: 91.7%-95.3%), as shown in Table 2.

The sensitivity shown in Table 2 demonstrates that there was no heterogeneity (χ^2, p = 0.41), and the inconsistency ($I^2 = 0\%$) was intermediate [37]. The specificity plot shown in Table 2 suggests that there was no heterogeneity, as assessed by χ^2 (p = 0.19), and the inconsistency ($I^2 = 36.8\%$) was intermediate [37].

The pooled positive likelihood ratio (Table 4) was 13.55 (95%CI: 10.25-17.92), meaning that a positive result from the Bayesian networks increased the odds that the patient had breast cancer (i.e., a true positive, TP) by 13.55 times. We did not observe heterogeneity in the Cochran Q test (p = 0.4154), at $\tau^2 = 0.0$, and there was no inconsistency ($I^2 = 0.0\%$).

The pooled negative likelihood ratio shown in Table 4 was 0.12 (95%CI: 0.09-0.18), a good result because it is close to zero and indicates that a negative result of Bayesian networks decreases the odds of a malignant breast lesion by a factor of 0.12. We found no heterogeneity with either the Cochran Q test (p = 0.2941) or with τ^2 (0.0337) and low inconsistency ($I^2 = 19.2\%$).

The pretest probability (i.e., prevalence) of the presence of cancer increased the probability of a positive test result from 40.56% to 90.24% (95%CI: 90.19%-90.29%) and a negative result

Table 3

Results of the risk of bias assessment for each study, according to *Quality Assessment of Diagnostic Accuracy Studies* (QUADAS).

Characteristic	Kahn Jr. et al. [24] (1997)	Hamilton et al. [23] (1994)	Cruz-Ramírez et al. [25] (2007)	Burnside et al. [10] (2004)
Representative spectrum	+		+	+
Acceptable reference standard	+	+	+	+
Acceptable delay between tests	+	+	+	+
Partial verification avoided	+	+	+	+
Differential verification avoided	+	+	+	+
Incorporation avoided	+	+	+	+
Reference standard results blinded	+	+	+	+
Index test results blinded				
Relevant clinical information	-	-		
Uninterpretable results reported	+	+	+	+
Withdrawals explained				

+: without bias; -: potential bias; blank: insufficient information.

Table 4

Summary of the likelihood ratios and post-test probabilities (random effects model) [14].

Study (year)	Likelihood ratio (95%CI)		Post-test probability * (95%CI)	
	Positive	Negative	Positive	Negative
Kahn Jr. et al.[24] (1997)	7.97 (3.72-17.07)	0.09 (0.02- 0.34)	79.37 (78.34-80.40)	4.59 (4.06-5.12)
Hamilton et al. [23] (1994)	38.50 (2.47- 599.32)	0.13 (0.04-0.41)	90.89 (89.48-92.30)	6.44 (5.24-7.64)
Cruz-Ramírez et al. [25] (2007)	14.79 (10.76-20.33)	0.11 (0.08-0.14)	91.38 (91.33-91.43)	6.68 (6.63-6.73)
Burnside et al. [10] (2004)	12.49 (4.75-32.83)	0.22 (0.11-0.45)	85.18 (84.43-85.94)	9.23 (8.62-9.85)
Total	**13.55 (10.25-17.92)**	**0.12 (0.09-0.18)**	**90.24 (90.19-90.29)**	**7.80 (7.76-7.84)**

95%CI: 95% confidence interval.

* Values of the calculated post-test probability: Pre-test probability = prevalence = 40.56%, Pre-test odds = prevalence/
(1-prevalence), Post-test odds = LH* x pre-test odds, Post-test probability = post-test odds/(1+post-test odds).

from the Bayesian network decreased the probability of a FP from 40.56% to 7.8% (95%CI: 7.76%-7.84%) (Table 4).

In the analysis of breast cancer versus benign lesions, the area under the SROC curve shown in Figure 2 was high (0.97) [38], supporting the use of Bayesian networks in breast cancer diagnosis; the Q * point was 0.92. In Figure 2 each point represents a single study, the middle line is the main curve, and other curves (first and third) represent the confidence interval.
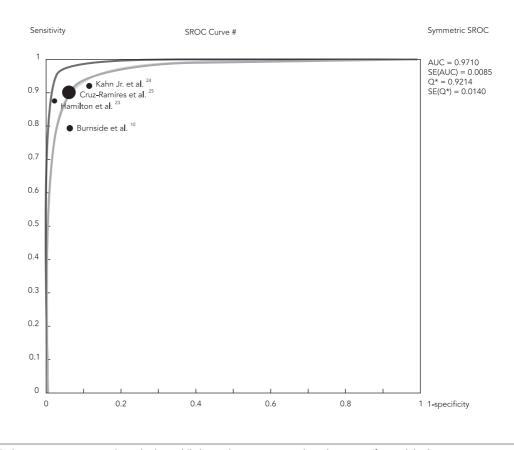
Discussion

This systematic review is the first to examine the accuracy of Bayesian networks in supporting breast cancer diagnoses. Our results demonstrated that this computational model can represent a non-invasive and accurate method that can be used to support breast cancer diagnosis.

For the development and modeling of Bayesian networks there are dozens of free tools or demo versions; however, though there are also

Figure 2

Summary Receiver Operating Characteristic curve.

Sensitivity   SROC Curve #   Symmetric SROC

AUC = 0.9710
SE(AUC) = 0.0085
Q* = 0.9214
SE(Q*) = 0.0140

Each point represents a single study, the middle line is the main curve and another curves (first and third) represent the confidence interval.

dozens of Bayesian networks applied to support the diagnosis of breast cancer presented by the scientific literature and included in our systematic review [10,23,24,25], these are made available by contacting the researchers who developed them and are used in the research centers that developed them.

By the research conducted, we observed the absence of cost-effectiveness studies on the use of this technology. Computational intelligence methods, such as Bayesian networks, have been introduced into clinical practice with the primary aims of assisting physicians in the diagnostic process by preparing therapeutic decisions and predicting various outcomes [39].

The mathematical formalism for Bayesian analysis originated in the theorem proposed by Thomas Bayes in 1763; as previously detailed, it states that conditional probabilities can be obtained through two approaches: (a) using information derived from expert knowledge or literature, and (b) the probabilities and structure obtained by Bayesian learning from large databases [7]. In the past, studies typically used the first approach [9]. Our systematic review included studies with both approaches; the specific approach used may relate to the period in which the study was performed.

Research using this theory to support breast cancer diagnosis began in the 1990s with a study in Ireland that aimed to determine the diagnosis and Bayesian network using information from fine needle aspiration [23]. That study [23], which was included in our systematic review and published in 1994, was a pioneer in using this computational model, and despite having not used an automated technique for the computation of the Bayesian network, that study reported high accuracy and favorable results that support its use in clinical practice.

Moreover, among the methods used in breast cancer diagnosis, mammography is generally

considered to be the best method available for breast cancer screening. However, some types of cancers detected in mammograms are missed by radiologists. Systems based on Bayesian networks applied to mammography seek to reduce false negatives by highlighting suspected areas for radiologists [40]. This feature was also noted in the study by Burnside [10], which is included in our systematic review.

A study performed by Laming & Warren [41] in 2000 reinforces this statement because while mammography is the primary method used for breast cancer screening, approximately 16% to 31% of cancers detected in mammograms can be missed when interpreted by a single radiologist. In this regard, a systematic review performed by Taylor & Potts in 2008 [42] revealed that the dual analysis in which two radiologists assessed the image increased the rate of cancer detection by 3-11 per 100,000 women screened. Thus, systems based on Bayesian networks can assist the analysis of suspicious areas that deserve review in those cases when it is not possible for two radiologists to analyze the mammograms [42].

Our systematic review allowed the extraction and reconstruction of diagnostic data from cross-sectional studies. The methodological quality of the included studies was high, although some QUADAS issues had negative evaluations, such as lacking information concerning whether the sample was representative, blinding in the use of Bayesian networks, lacking relevant clinical records, and no record of the subjects who were removed from the sample.

Although an extensive and detailed search strategy was employed, which enabled retrieval of all publications regardless of language, the terms used may have contributed to the failure to locate certain publications that could be relevant to our systematic review. A bivariate analysis used preserves the two-dimensional nature of the diagnostic data and considers the measurement variability within and between studies [19]. We used the most current guidelines indicated in the preparation of systematic reviews as described in the *Handbook for Systematic Reviews of Diagnostic Accuracy of Cochrane* [11,12,17,18,19,21,22].

The studies included in our systematic review presented a dichotomous outcome for breast cancer diagnoses; however, there were some differences in the composition of the input nodes. The study by Kahn Jr. et al. [24] considered the patients' histories and physical findings and some mammographic data. Hamilton et al. [23] considered several cytological characteristics. Cruz-Ramírez et al. [25] used the age of the patient and the cytological characteristics, and Burnside et al. [10] considered 25 descriptors to produce hierarchical classifications in BI-RADS. Despite these differences, all considered issues were relevant to the breast cancer diagnosis.

Regarding economics, some studies have shown that the women treated in public institutions have more advanced stages of the disease, less access to modern therapies, and a lower survival rate than the patients treated in private institutions [43,44]. The triple test (physical examination, mammography, and cytology by fine needle aspiration) has been employed as a method to accurately diagnose palpable breast lesions. When the three diagnostic methods are consistent, the elimination of biopsy as a confirmatory test is usually recommended and can result in reduced spending [43]. In this context, artificial intelligence techniques, such as Bayesian networks using above characteristics, now represent a reality and may decrease the uncertainty present in biopsies derived from suspicious nodules, enabling the reduction of public health costs [43,44].

Bayesian networks have potential to enhance the diagnostic process by instilling consistency and repeatability. The use of this system, together with pre-interpreted diagnostic information, could also provide an effective computer-based training system for breast cancer diagnostic [45]. This comparison does not imply that the Bayesian network could replace the specialist but may indicate that technology can calculate diagnostic across many variables, incorporate complex dependencies among variables, and aid, for example, the radiologists' interpretations [45].

Our meta-analysis showed that Bayesian networks have increased the probability of a breast cancer diagnosis by 49.68%, suggesting that this type of tool can be useful in evaluating suspicious lesions. Our findings also indicate that, given a negative diagnosis, Bayesian networks decreased the likelihood of false positives by 32.76% supporting their utility in evaluating lesions that are deemed to be most likely benign.

In conclusion, probabilistic computer models like Bayesian networks represents a noninvasive method that may substantially aid physicians attempting to diagnose breast cancer in a timely and accurate manner.

Resumen

El objetivo de este estudio fue evaluar la exactitud de las redes bayesianas para apoyar el diagnóstico de cáncer de mama. Se realizó una revisión sistemática y un metaanálisis, que incluyeron artículos y estudios publicados entre enero de 1990 y marzo de 2013. Se incluyeron estudios transversales prospectivos y retrospectivos, que evaluaron la exactitud del diagnóstico de lesiones mamarias (condición de destino), utilizando redes bayesianas (prueba de evaluación). Se analizaron cuatro estudios que incluyeron 1.223 lesiones de mama primarias, un 89,52% (444/496) de los casos de cáncer de mama, y un 6,33% (46/727) de las lesiones benignas se tomaron como base de análisis de las redes bayesianas. El área bajo la curva SROC (característica operativa del receptor) fue de un 0,97, con un valor de Q* de un 0,92. El uso de las redes bayesianas en el diagnóstico de las lesiones malignas aumentó la probabilidad pre test de un verdadero positivo desde un 40,03% a un 90,05%, y la disminución de la probabilidad de un falso negativo de un 6,44%. Por lo tanto, nuestros resultados demuestran que las redes bayesianas ofrecen un método preciso y no invasivo en el apoyo del diagnóstico del cáncer mamario.

Informática Médica; Teorema de Bayes; Neoplasias de la Mama

Contributors

All authors contributed to the study conception, literature search, data extraction and analysis, statistical analysis, preparation, revision of the manuscript, definition of intellectual content, and approved the final version of the manuscript.

References

1. American Cancer Society. Global cancer facts and figures. 2nd Ed. Atlanta: American Cancer Society; 2011.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011; 61:69-90.
3. World Health Organization. Breast cancer and screening. http://www.emro.who.int/ncd/publications/breastcancerscreening.pdf (accessed on Mar/2014).
4. Lisboa PJ. A review of evidence of health benefit from artificial neural networks in medical intervention. Neural Netw 2002; 15:11-39.
5. Massad E. Métodos quantitativos em medicina. Barueri: Manole; 2004.
6. Pearl J. Causality: models, reasoning and inference. London: Cambridge University Press; 2000.
7. Neapolitan RE. Learning bayesian networks. New Jersey: Pearson Prentice Hall; 2004.
8. Burnside ES. Bayesian networks: computer-assisted diagnosis support in radiology. Acad Radiol 2005; 12:422-30.
9. Burnside ES, Rubin DL, Shachter RD. Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. Stud Health Technol Inform 2004; 107:13-7.

10. Burnside ES, Rubin DL, Shachter RD, Sohlich RE, Sickles EA. A probabilistic expert system that provides automated mammographic-histologic correlation: initial experience. AJR Am J Roentgenol 2004; 182:481-8.

11. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3:25.

12. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006; 6:9.

13. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.0. The Cochrane Collaboration; 2009. http://srdta.cochrane.org.

14. Altman DG. Some common problems in medical research. In: Altman DG, editor. Practical statistics for medical research. 9th Ed. London: Chapman and Hall; 1999. p. 396-439.

15. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006; 6:31.

16. The Cochrane Collaboration. http://www.cochrane.org/ (accessed on Mar/2014).

17. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120:667-76.

18. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58:982-90.

19. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008; 149:889-97.

20. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7:177-88.

21. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. BMJ 2001; 323:157-62.

22. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results for several studies in metanalysis. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: John Wiley Professional; 2001. p. 285-312.

23. Hamilton PW, Anderson N, Bartels PH, Thompson D. Expert system support using Bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast. J Clin Pathol 1994; 47:329-36.

24. Kahn Jr. CE, Roberts LM, Shaffer KA, Haddawy P. Construction of a Bayesian network for mammographic diagnosis of breast cancer. Comput Biol Med 1997; 27:19-29.

25. Cruz-Ramírez N, Acosta-Mesa HG, Carrillo-Calvet H, Nava-Fernández LA, Barrientos-Martínez RE. Diagnosis of breast cancer using Bayesian networks: a case study. Comput Biol Med 2007; 37:1553-64.

26. Ngo L, Haddawy P, Krieger RA, Helwig J. Efficient temporal probabilistic reasoning via context-sensitive model construction. Comput Biol Med 1997; 27:453-76.

27. Srinivas S, Breese J. IDEAL: a software package for analysis of influence diagrams. Uncertain Artif Intell 1990; 1990:212-9.

28. Cross SS, Dubé AK, Johnson JS, McCulloch TA, Quincey C, Harrison RF, et al. Evaluation of a statistically derived decision tree for the cytodiagnosis of fine needle aspirates of the breast (FNAB). Cytopathology 1998; 9:178-87.

29. Cross SS, Downs J, Drezet P, Ma Z, Harrison RF. Which decision support technologies are appropriate for the cytodiagnosis of breast cancer? In: Jain A, Jain A, Jain S, Jain L, editors. Artificial intelligence techniques in breast cancer diagnosis and prognosis. Singapore: World Scientific; 2000. p. 265-95.

30. Cross SS, Stephenson TJ, Mohammed T, Harrisont RF. Validation of a decision support system for the cytodiagnosis of fine needle aspirates of the breast using a prospectively collected dataset from multiple observers in a working clinical environment. Cytopathology 2000; 11:503-12.

31. Cheng J, Greiner R. Learning Bayesian belief network classifiers: algorithms and systems. In: Stroulia E, Matwin S, editors. Advances in artificial intelligence. Ottawa: Springer; 2001. p. 141-51.

32. Cheng J, Greiner R, Kelly J, Bell D, Liu W. Learning Bayesian networks from data: an information-theory based approach. Artif Intell 2002; 137:43-90.

33. Norsys Software Corporation. Netica application, version 4.16. http://www.norsys.com (accessed on Mar/2014).

34. D'Orsi CJ, Kopans DB. Mammography interpretation: the BI-RADS method. Am Fam Physician 1997; 55:1548-52.

35. D'Orsi CJ, Newell MS. BI-RADS decoded: detailed guidance on potentially confusing issues. Radiol Clin North Am 2007; 45:751-63.

36. University of Pittsburgh Decision Systems Laboratory. GENIe. http://www.sis.pitt.edu/~dsl/ (accessed on Mar/2014).

37. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557-60.

38. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240:1285-93.

39. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004; 86:334-8.

40. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. Arch Gynecol Obstet 2009; 279:881-90.

41. Laming D, Warren R. Improving the detection of cancer in the screening of mammograms. J Med Screen 2000; 7:24-30.

42. Taylor P, Potts HW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. Eur J Cancer 2008; 44:798-807.

43. Al-Mulhim AS, Sultan M, Al-Mulhim FM, Al-Wehedy A, Ali AM, Al-Suwaigh A, et al. Accuracy of the "triple test" in the diagnosis of palpable breast masses in Saudi females. Ann Saudi Med 2003; 23:158-61.

44. Simon S, Bines J, Barrios C, Nunes J, Gomes E, Pacheco F, et al. Clinical characteristics and outcome of treatment of Brazilian women with breast cancer treated at public and private institutions. The Amazone Project of the Brazilian Breast Cancer Study Group (GBECAM). Cancer Res 2009; 69(24 Suppl 3):3082.

45. Burnside ES, Davis J, Chhatwal J, Alagoz O, Lindstrom MJ, Geller BM, et al. Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings. Radiology 2009; 251:663-72.