

A Targeted Review of Breast Cancer Studies of Concordance for an Internationally-Implemented Artificially Intelligent Clinical Decision-Support System

Yull Edwin Arriaga (✉ yull.edwin.arriaga@IBM.com)

IBM Watson Health <https://orcid.org/0000-0002-7229-2028>

Rezzan Hekmat

IBM Watson Health

Karlis Draulis

IBM Watson Health

Suwei Wang

IBM Watson Health

Anita M Preininger

IBM Watson Health

Laura C. Morgan

IBM Life Sciences

Kristina Lindsley

IBM Life Sciences

Winnie Felix

IBM Watson Health

Irene Dankwa-Mullan

IBM Watson Health

Kyu Rhee

IBM Watson Health

Gretchen P Jackson

IBM Watson Health

Research article

Keywords: Artificial intelligence, clinical decision-support system, breast cancer, treatment decision concordance, treatment choice

Posted Date: November 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-101188/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Breast cancer has the highest incidence and is the leading cause of cancer-related mortality among women worldwide. IBM Watson® for Oncology (WfO), an artificial intelligence-based clinical decision-support system, provides therapeutic options for consideration to cancer-treating physicians. We conducted a targeted review of studies evaluating concordance of therapeutic options offered by the system with treatment decisions by practicing clinicians in breast cancer.

Methods: PubMed, EMBASE, Cochrane, trial registers, conference abstracts, and an internal publication database were searched to identify studies evaluating the concordance of system-generated therapeutic options with treatment decisions by individual clinicians and multidisciplinary tumor boards for breast cancer patients reported in peer-reviewed abstracts or papers published in English between 01/01/2015 and 11/15/2019.

Results: Ten breast cancer concordance studies (4703 patients) that met the inclusion criteria were identified and analyzed; the identified studies were from China, India, and Thailand. The weighted mean concordance for all studies was 67.4% (SD 16.0%, range 55.0% - 98.0%). The weighted mean concordance of the system with multidisciplinary tumor boards was 88.2%, (SD 9.7%, range 76.5% - 98.0%), which was substantially higher than concordance between the system and individual clinicians (61.5% , SD 10.1%, range 55.0% -76.0%).

Conclusion: Concordance between system-generated therapeutic options and treatment decisions of multidisciplinary tumor boards or individual clinicians for breast cancer demonstrated overall agreement between the system and decisions of practicing cancer-treating physicians in China, India and Thailand. As multidisciplinary tumor boards may lead to higher quality clinical decision-making compared to those of individual clinicians in practice, the relatively higher concordance of the system with multidisciplinary tumor boards suggests a role for clinical decision support to inform clinicians of evidence-informed treatment options.

Background

Breast cancer is a global health problem as it the most common malignancy and one of the leading causes of cancer-related mortality in women.^{1,2} Among all invasive malignancies in the United States (US), breast cancer has the highest annual incidence and is the second leading cause of cancer-related death in women.³ The incidence of breast cancer in the US rose at an annual rate of 0.2% between 2005 and 2011.⁴ In 2020, it is estimated that 276,480 new breast cancer cases will be diagnosed in the US, with 42,170 breast cancer-related deaths in females.³ Moreover, in low- and middle-income countries, breast cancer is a leading cause of morbidity, disability, and mortality as well.^{2,5} In 2015, the number of new female breast cancer cases and breast cancer-related deaths in China was 268,600 and 69,500 respectively.³⁶

Along with a global rise in the burden of breast cancer, a shortage of cancer care services exists.⁶ The increasing demand for cancer care,⁷ without a proportional increase in services, poses challenges to cancer-care institutions, providers, and patients. Thus, there is a growing need for informatics resources and infrastructure to provide clinical decision-support systems (CDSS) that can help oncologists keep pace with medical advances and rapid practice changes for optimal care of patients with breast cancer.

Over the past decade, results of randomized clinical trials have led to substantial changes in the management of breast cancer. Examples include the avoidance of axillary lymph node dissection in patients with early stage, (i.e., T1-T2) hormone-receptor (HR) positive breast cancer with 1 to 2 positive sentinel lymph nodes.⁸ A 21-gene tumor expression assay now guides adjuvant chemotherapy decision-making in patients with HR positive, human epidermal growth factor receptor 2 (HER2) negative, axillary lymph node-negative breast cancer.⁹ A contemporary first-line treatment option for patients with metastatic HR positive, HER2 negative breast cancer is a selective inhibitor of cyclin-dependent kinases 4 and 6 combined with an aromatase inhibitor.¹⁰ As of December 2019, the US Food Drug Administration (FDA) listed 69 drugs approved for the treatment of breast cancer.¹¹ Moreover, an increasing number of novel breast cancer drugs and multimodal treatment strategies are now undergoing evaluation in clinical trials worldwide. The clinical trial registry, ClinicalTrials.gov, listed 487 female breast cancer clinical trials in December of 2019.¹² Consequently, oncologists and their patients face a complex and fast changing breast cancer management landscape. Through shared-decision making¹³, and based on evidence, best practice, and cost-effectiveness considerations, they must choose among diagnostic and staging tests, therapeutic modalities and their sequence. Unfortunately, limited health resources and variability in breast cancer practice patterns in different regions of the world pose a challenge to the ability of oncologists and their patients to make personalized treatment decisions.

Published studies reporting on the performance and implementation of artificial intelligence (AI)-based CDSS aiding oncologists in cancer treatment decision making are quite limited. Furthermore, there is dearth of therapeutic AI-CDSS implemented in routine oncology practice, that consider key patient attributes and incorporate evidence from peer-reviewed literature and cancer treatment guidelines supporting oncologists with personalized treatment plan suggestions. IBM Watson® for Oncology (WfO) is an AI-based CDSS¹⁴ that considers select patient data for a given cancer type and provides a set of evidence-informed therapeutic options for consideration by cancer-treating physicians. The options presented by the system are accompanied by published evidence in the medical literature, facilitating personalized, evidence-informed treatment options for patients. Studies have evaluated the acceptability and validity of therapeutic options suggested by this CDSS, as well as concordance between the tool's treatment suggestions and treatment decisions made by cancer-treating physicians at the point of care for a variety of cancer types.¹⁵⁻¹⁷ To summarize the performance of the CDSS for breast cancer treatment, we conducted a targeted review of peer-reviewed studies evaluating concordance between therapeutic options offered by the system and treatment decisions by cancer-treating clinicians in practice. With our targeted review we hope to advance the field of informatics applied to clinical oncology

by providing knowledge about the performance of an AI-based CDSS aiding clinicians in breast cancer treatment-decision making.

Methods

Aims and Research Questions

The aims of this study were to summarize and analyze the results of a targeted review of peer-reviewed published studies reporting the concordance of therapeutic options from the CDSS with individual clinicians and multidisciplinary tumor boards (MTB) treatment recommendations for breast cancer. Our specific research questions were:

- What are the overall concordance rates between system-generated therapeutic options and practicing oncologists' treatment decisions?
- Are there differences in concordance rates between system-generated therapeutic options and individual clinicians-treatment decisions compared to concordance rates between system-generated therapeutic options and MTB-treatment decisions?
- What are the concordance rates in different subgroups of patients with breast cancer, according to age, menopausal status, cancer stage, HR status, HER2 status, molecular subtype?
- What are the concordance rates of system-generated therapeutic options with practicing oncologists' treatment decisions by country?

Studies Eligibility Criteria

We included prospective and retrospective studies reporting the concordance of WfO's therapeutic options with practicing oncologists' treatment decisions in breast cancer. We selected studies published in English in peer-reviewed journals and peer-reviewed abstracts from oncology conferences. We excluded studies reporting the concordance of WfO therapeutic options with practicing oncologists' decisions in more than one cancer type lacking separate results for breast cancer. We did not exclude publications based on the country where the study was performed. We did not find peer-reviewed published studies reporting on the performance of AI-CDSSs aiding oncologists in breast cancer treatment decision making other than WfO. Table 1 summarizes the study inclusion and exclusion criteria.

Table 1
Study eligibility criteria for selection of WfO breast cancer concordance studies.

Inclusion Criteria	Exclusion Criteria
Study design/type: • Retrospective or prospective study • Peer-reviewed publication or abstract from an oncology conference reported in English Population: • Female breast cancer patients • Any stage according to the American Joint Committee on Cancer (AJCC) staging system: I through IV • Any breast cancer type: HR positive or negative, HER2 positive or negative, triple negative Comparisons: • WfO therapeutic options compared with breast cancer treatment decisions by individual clinicians or MTB Outcomes: • Rate of treatment decision concordance	Study design/type: • Commentary, opinion paper, review article, or editorial essay without new clinical results • Report with duplicative data (e.g., multiple publications on same study, abstract when a full text article was available) Population: • Study reported multiple cancers but did not provide separate results for breast cancer Comparisons: • Study without a comparison group (e.g., study reporting WfO usability, acceptability, learnability, integration into the workflow, or treatment decision impact only) • Study evaluating the technical performance of WfO for non-use cases (e.g., incorporated diagnostic tests not recognized as required for making treatment recommendations) • Study that reported concordance with WfO therapeutic options with other type of CDSS, or in conjunction with another CDSS Outcomes: • Study did not report concordance outcome

Abbreviations: HR, hormone receptor; HER2, human epidermal growth factor receptor 2; WfO, Watson for Oncology; MTB, multidisciplinary tumor board; CDSS, clinical decision support system.

Information Sources

A systematic review was conducted by a 2-person team of Watson Health analysts (KL and LM) with expertise in literature-based research to identify and evaluate CDSS in oncology. The current report presents a subset of studies identified from such systematic review, limited to those reporting breast cancer treatment decision concordance rates of the system with cancer-treating physicians in practice, from January 1, 2015 to November 15, 2019. Studies comparing the performance of the system with other oncology therapeutic CDSSs or with treatment decisions derived from real world evidence were excluded.¹⁸

Comprehensive searches were performed in PubMed, EMBASE, the Cochrane Library, and online clinical trial registers (ClinicalTrials.gov and the World Health Organization Clinical Trial Portal). Using the

EMBASE platform, we searched conference abstracts from the American Society of Clinical Oncology (ASCO), San Antonio Breast Cancer Symposium, European Society for Medical Oncology (ESMO), ESMO Asia, and International Gynecologic Cancer Society online. We supplemented these searches by performing manual searches of abstracts not indexed in EMBASE but available online from the Chinese Society of Clinical Oncology, with search of references cited by included studies to identify any additional pertinent studies.

Study Selection

We used Endnote¹⁸ to manage and remove duplicate references. Two experienced reviewers (KL, LM) screened titles and abstracts from each unique record for relevancy in DistillerSR (Evidence Partners), a software designed for supporting literature reviews. For all records classified as relevant based on the title and abstract, one reviewer assessed the full text report to determine final eligibility for inclusion using established criteria. For this targeted review, a second reviewer assessed all studies based on criteria presented in Table 1. In cases of uncertainty, a consensus decision was reached through discussion.

Data Collection Process

Two additional reviewers (RH and KD) confirmed and extracted data in those studies that were identified by initial reviewers. The additional reviews were conducted using a pre-defined data collection form that included data listed below.

- Citation information: study title, authors, and year of publication.
- Study characteristics: country where study was performed, name of institution.
- Study design: retrospective or prospective, number of patients, and date of study initiation and completion.
- Clinical context: mean patient age in years, and menopausal status, The system use case (MTB or individual clinician) and version utilized in the study.
- Outcomes: definition and percent of concordance between CDSS therapeutic options and practicing oncologists' treatment decision, concordance according to breast cancer stage by the AJCC staging

(edition used according to the year of study completion), HR status, HER2 status, molecular sub-type, and reported reasons for discordance.

Analysis

We analyzed treatment concordance (agreement) between the system's therapeutic options and treatment decisions made by either MTB or individual clinicians in breast cancer patients. Decisions made by MTB or individual clinicians were defined as concordant if they agreed with the CDSS treatment options labeled as "Recommended" or "For Consideration." Concordance was calculated based on the number of concordant treatment decisions divided by the total number of patients in each study. Mean concordance was calculated as a weighted average based on the number of patients in each study, assuming no patient was included in more than one study (independent samples) and data were normally distributed. We summarized concordance by patient subgroups according to the AJCC breast cancer stage edition utilized in the study (e.g. 7th), HR status, HER2 status, as well as luminal A, luminal B, and triple negative breast cancer (TNBC).

Results

Identification and Selection of Studies

Comprehensive searching yielded 1502 total unique records (Fig. 1). After title and abstract screening, we retrieved 211 full text reports. We further screened the 211 studies to identify those reporting concordance between system-generated therapeutic options with treatment decisions of practicing oncologists in breast cancer. Of the 211 studies, we excluded 201 for the following reasons: 89 did not evaluate this system, 14 did not have clinical results, 5 provided only generic information about the device, 56 had different research questions or outcomes, 18 evaluated the system's therapeutic option concordance with practicing oncologists' treatment decisions in malignancies other than breast cancer, 4 used the system outside of the approved indications or populations, 2 utilized attributes not recognized by the system,^{19,20} 7 were conference abstracts already published in full text, and 6 reported on an ongoing trial where results were not yet available. Of the 56 reports that had different research questions or outcomes, 26 did not report concordance as an outcome, 1 compared concordance of system-generated therapeutic options with real world evidence treatment decisions derived from a cohort of US breast cancer patients,²¹ and 1 compared concordance of system-generated therapeutic options with treatments recommended by a breast cancer genomic test.²² In total, our targeted review included 10 breast cancer concordance studies.

Breast Cancer Concordance Studies According to Use Case

The 10 concordance studies included in this review were retrospective, enrolling a total of 4703 patients distributed across regions of China, India, and Thailand (Table 2). These studies compared system-generated therapeutic options with treatment decisions made by MTBs (5 studies) or individual clinicians (5 studies). Across all 10 studies, the mean weighed concordance was 67.4% (standard deviation (SD)

16.0%, range 55.0% – 98.0%). The mean weighed average concordance of the system with MTBs of 88.2%, (SD 9.7%, range 76.5% – 98.0%) was higher than the mean weighed average concordance between WfO and individual clinicians of 61.5% (SD 10.1%, range 55.0% – 76.0%).

Table 2
Mean weighed average concordance by WfO use case

Study/Location	Number of patients	Concordance
Multidisciplinary Tumor Board Studies		
Zhou N, et al. 2017. China ²³	119	79.0%
Yue L, Yang L. 2017. China ²⁴	31	98.0%
Somashekhar SP, et al. 2018. India ²⁵	638	93.0%
Zhou N, et al. 2019. China ²⁶	120	82.0%
Xu J et al. 2019. China ²⁷	132	76.5%
MTB subtotals	1,040	88.2%
Individual Clinicians Studies		
Suwanvecho S, et al. 2017. Thailand ²⁸	172	71.0%
Jiang Z, et al. 2018. China ²⁹	1997	55.0%
Suwanrusme H, et al. 2018. Thailand ³⁰	92	76.0%
Pan H et al. 2019. China ³¹	1301	69.4%
Suwanvecho S et al. 2019. Thailand ³²	101	60.0%
IC subtotals	3663	61.5%

Abbreviations: MTB, multidisciplinary tumor board; IC, individual clinician.

Concordance with Multidisciplinary Tumor Boards: Specific Studies

Of the five studies that compared this CDSS with MTBs, four were conducted in China and the largest one in India. A retrospective observational study involving 638 patients in India compared treatment decision agreement between the system and a 15-member expert MTB²⁵. These patients were either naïve to treatment, had disease recurrence after systemic therapy, or had surgery at a tertiary comprehensive cancer center. A 93% concordance rate between the MTB and the CDSS was found.²⁵ Subgroup analysis showed that concordance was inversely related to age; patients 75 years of age and older had the lowest

concordance. Concordance for stages I and IV were lower than stages II and III. Concordance for metastatic HR positive breast cancer and TNBC was 75% and 85%, respectively, and 98% for HER2 positive metastatic disease. Local treatment preferences accounted for 23% of non-concordant cases.²⁵

Another study of 120 patients from China demonstrated an 82.0% concordance between a multidisciplinary tumor board and the system.²⁶ Treatment decision concordance for luminal A, luminal B, and triple negative breast cancers were 63.0%, 87.0% and 79.0% respectively, although there were relatively few of each subtype of cancer patients in this study. Concordance was 86.7% for stages I and II disease combined and 79.5% for stage III. This was one of the few studies reporting concordance based on menopausal status, with no differences found based on this patient clinical characteristic.²⁶

Concordance with Individual Clinicians: Specific Studies

There was an overall concordance of 55% between practicing individual clinicians and the CDSS in a large, cross-sectional retrospective study of 1997 patients from China.²⁹ These patients had non-metastatic breast cancer and were at high risk of relapse or metastatic disease. Among this group, concordance for stage II and stage IV disease was 66% and 50%, respectively. TNBC displayed the highest concordance between individual clinicians and the CDSS across all breast cancer types with 69% agreement.²⁹ Consistent with this result, a retrospective study in China of 1301 breast cancer patients demonstrated an overall concordance of 69.4% between individual clinicians and the system.³¹ In this study, concordance for neoadjuvant and adjuvant chemotherapy decisions were 96.7% and 65.0%, respectively, and concordance for TNBC was 89.3%. Multivariate analysis showed that the concordance of the chemotherapy choice was lower in patients ≥ 70 years of age, compared to ≤ 40 years of age (OR = 0.33, 95% CI, 0.14–0.78).³¹

Concordance by Country

The mean weighed average breast cancer treatment concordance varied by country, ranging from 60.7% for studies comparing WfO with individual clinicians in China^{29, 31} to 93.0% for a study comparing the system with a MTB at one tertiary cancer center in India²⁵ (Table 3). There were substantial differences in concordance by country and CDSS use case.

Table 3
Mean weighed average concordance by country and WfO use case

	China		India		Thailand
CDSS use case	Multidisciplinary Tumor Board (MTB)		Individual Clinician	Multidisciplinary Tumor Board (MTB)	Individual Clinician
Number of Studies (patients)	4 (404)		2 (3298)	1 (638)	2 (365)
Concordance	80.5%		60.7%	93.0%	69.2%

Abbreviations: CDSS, clinical decision support system; MTB, multidisciplinary tumor board.

Concordance by AJCC Stage

Treatment decision concordance by breast cancer stage was reported in 4 studies,^{25,26,27,29} as shown in Table 4. In 3 of these studies,^{25,27,29} concordance was higher for patients with non-metastatic breast cancer, as compared to metastatic disease. In the study reported by Someshakar et al.,²⁵ concordance was higher for stages II and III than for stage I.

Table 4
Concordance by AJCC stage

Study/Location	Non-Metastatic			Metastatic		
	Stage	Number of patients	Concordance	Stage	Number of patients	Concordance
Jiang Z et al. 2018. China ²⁹	stage II	NR	66.0%	stage IV	NR	50.0%
Somasekhar SP et al. 2018. India ²⁵	stage I	61	80.0%	stage IV	124	86.0%
	stage II	262	97.0%			
	stage III	191	95.0%			
Zhou N et al. 2019. China ²⁶	stages I & II	80	86.3%	stage IV	1*	NA
	stage III	39	79.5%			
Xu J et al, 2019. China ²⁷	non-metastatic	92	79.4%	stage IV	40	70.0%

Abbreviations: AJCC, American Joint Committee on Cancer; NR, not reported; NA, not applicable. *There was 1 patient with metastatic breast cancer in the Zhou N et al study.

A lower concordance in patients older than 70 was found in 2 studies.^{25,29} The reported reasons for treatment decision discordance between MTBs or individual clinicians with the CDSS varied among studies. Documented reasons were related to availability of treatments recommended by the system, the absence of some clinician-preferred treatments as options in the system, patient preferences, and age $\geq$ 70 years.

Discussion

Treatment Concordance by Use Case: MTB vs. Individual Clinicians.

To our knowledge, this study is one of the first to summarize the performance of an oncology AI-based CDSS, measured by concordance of its therapeutic suggestions with treatment decisions made by cancer-treating physicians in practice for female breast cancer patients in diverse, international, cancer care settings. We found substantial agreement between system-generated therapeutic options and both treatment decisions of MTBs as well as individual clinicians in a large number of patients with breast cancer. Our targeted review demonstrate that the system's suggested treatment options agreed with therapies selected by cancer-treating physicians in China, India, and Thailand, countries where we identified breast cancer treatment decision concordance studies.

The CDSS exhibited a higher treatment decision concordance with MTBs, as compared to individual clinicians. MTBs provide multidisciplinary team management that generally results in decreased mortality, improved quality of life, and reduced costs in cancer patient care.^{33,34} The higher concordance between system-generated therapeutic options and treatments agreed upon by MTB experts is consistent with the quality of therapeutic options suggested by this CDSS.³⁵ Furthermore, the lower rate of concordance with decisions made by individual physicians as compared to MTBs supports a role for an AI-based CDSS in aiding individual oncologists during the complex clinical task of breast cancer treatment decision making.

Concordance in Different Countries and Breast Cancer Subgroups

According to system use case and country, we identified a study conducted at a tertiary cancer center in India with a higher breast cancer treatment decision concordance between the CDSS and MTB,²⁵ as compared to similar use case studies in China.^{23,24,26,27} Likewise, individual clinicians in Thailand had higher concordance with the CDSS in 3 studies^{28,30,32} as compared to 2 large individual clinicians studies from China.^{29,31} Differences in breast cancer treatment decision concordance between the system and individual clinicians or MTBs in different countries are multifactorial and likely explained by differences in oncology practice patterns at the institutional and national levels.

Successful implementation of a CDSS in medical practice can be achieved by identifying and addressing barriers to CDSS clinical adoption. Successful CDSS implementation relies on factors such as quality, complexity, usability, learnability, transparency, workflow integration, and cost-effectiveness of a CDSS.

Furthermore, there is need for early involvement of end users in the development and enhancement of these systems. Consideration of regional health regulatory requirements and localization efforts to address regional differences in clinical practice are also important.³⁷

A lower concordance between system-generated therapeutic options and MTBs, as well as individual clinicians in breast cancer patients ≥ 70 years of age, was reported in 2 large studies.^{25,29} Age-related differences in patient and cancer care across China, India and the US may account for the lower concordance reported in older patients. Breast cancer stage at presentation, patient functional status, comorbidity burden, socioeconomic support, cultural values and treatment preferences may play a role in concordance, with a need for well-designed studies to promote evidence-informed management of elderly patients with breast cancer.³⁸ Prospective studies are also needed to evaluate the technical performance and clinical impact of the system in different subgroups of breast cancer patients.

Evaluation of CDSS Clinical Decision Quality and Impact

Concordance between CDSS and clinical decisions made in practice is limited as a measure of decision quality, which should be based on evidence and best practices. We selected concordance as a metric because many early adopters of CDSS have performed concordance studies to demonstrate reasonable agreement and build trust with end users. Evaluations of CDSSs often measure treatment decision adherence to guideline recommendations from the National Comprehensive Cancer Network (NCCN), Chinese Society of Clinical Oncology (CSCO) or other established guidelines. A study measured adherence of treatment decisions to guideline recommendations in 57 patients with advanced breast, colon, endometrial, esophageal, hepatic, gastric, ovarian, and rectal cancer in China which found a high adherence of WfO treatment options to NCCN and CSCO guidelines.³⁹ A study of 69 patients with colon and rectal cancers from South Korea found an adherence of WfO treatment options to NCCN guidelines and treatment decision concordance with a local MTB of 88.4% and 87.0% respectively.¹⁷

A blinded panel of cancer experts re-evaluating treatment decisions by both humans and CDSS or measuring treatment decision adherence to guideline recommendations helps reduce bias associated with the source of recommendations. A blinded study comparing WfO therapeutic options and treatment recommendations by individual physicians in breast, colon, lung and rectal cancers at a regional referral hospital in Thailand employed an expert panel of 3 oncologists to evaluate the quality of treatment decisions offered by clinicians and the CDSS.³² The expert panel, which was blinded to the source of treatment decision, compared treatments recommended by the individual clinicians and system-generated therapeutic options, rating 71% of these paired options as either identical or acceptable.³²

The concordance studies we identified in this targeted literature review reflect the intended use of the CDSS, which is to support cancer-treating physicians by providing therapeutic options that reflect best evidence and current practice. The system was designed according to a premise that humans are more likely to make optimal treatment decisions when supported by a CDSS. For institutions lacking a MTB, a CDSS may help fill this gap by providing individual clinicians with a choice of evidence-informed

therapeutic options. Consistent with this idea, use of the system as part of clinicians overall decision-making process significantly impacted treatment decisions in several studies. A large cross-sectional observational study measured the impact of the CDSS in treatment decision-making by individual clinicians in 1197 patients with breast cancer in China.²⁹ Participating physicians, initially blinded to system-generated therapeutic options, saw impact of the system on treatment recommendations in 5.0% of cases after the system's therapeutic options were disclosed to them. The adherence of breast cancer treatments to NCCN and CSCO guidelines increased from an 89.0% baseline to 97.0% in the 5.0% of cases where the physicians reevaluated treatment recommendations after viewing system-generated therapeutic options, as a part of their clinical decision-making process. Another study performed at a tertiary cancer center in India examining treatment decisions before and after exposure to the CDSS's therapeutic options in 1000 cases of breast, colon, lung, and rectal cancer cases showed a 13.6% decision impact for treatment decisions made by MTB, which were reevaluated during clinical decision making that included viewing therapeutic options offered by the CDSS. This demonstrates that even treatment decisions of expert MTBs in tertiary cancer care settings may potentially be improved by use of a CDSS.⁴⁰

Limitations

This targeted review has several limitations related to risk of bias. The reported concordance of system-generated therapeutic options with oncologists' treatment decisions from individual studies were combined, analyzed and reported according to CDSS use case. There is an inherent risk of bias introduced by the methodology utilized in the studies included in our targeted review, which were all retrospective. Moreover, the studies we reviewed had different sample sizes, proportions of patients by cancer stage, and used various versions of the CDSS. Our targeted review identified and included in the final analysis 10 eligible studies performed in 3 countries, all in the Asian continent. Therefore, the results of our review may not reflect the system's performance in the US, other Western countries or other regions of the world such as Africa or Latin America.

We did not include studies evaluating the usability, end user satisfaction, workflow integration, or the clinical impact of the system in breast cancer treatment decision making. These factors are likely to play a key role in the performance, implementation, acceptability, and clinical adoption of a CDSS. Nevertheless, assessing technical performance by measuring clinical decision concordance is an important first step in fostering end-user trust and adoption of a CDSS. Technical performance studies are necessary to address a CDSS accuracy as a potential confounder in future workflow or clinical decision impact studies. Strengths of our study are the inclusion of a large number of breast cancer patients in diverse clinical settings, inclusion of peer-reviewed publications only, and clinically relevant analysis of treatment decision concordance based on CDSS use case (MTB versus individual clinicians) and breast cancer subgroups.

Conclusion

In summary, this study is one of the first targeted reviews of breast cancer treatment decision concordance studies in women for an internationally-deployed AI-based CDSS. The concordance between the CDSS, MTBs, and individual clinicians demonstrated good system agreement with practicing oncologists in China, India and Thailand. A higher concordance was observed between the system and MTBs than the system and individual clinicians, likely reflecting a greater agreement between multidisciplinary expert consensus with evidence- and guideline-informed therapeutic recommendations of the system. This finding suggests a role of the CDSS in treatment decision support in breast cancer practice. Concordance varied across countries, reflecting regional differences in breast cancer practice. Non-concordant treatment decisions were likely related to physician preference, absence of some prescribed cancer therapies in practice as treatment options in the system, as well as differences in treatment due to factors such as patient or family preferences and availability of social support. Prospective randomized clinical trials are needed to assess the usability, workflow integration, user satisfaction, and clinical impact of oncology AI-based CDSSs on treatment decisions and relevant clinical outcomes such as progression-free survival, overall survival, quality of life, and patient reported outcomes.

Abbreviations

HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AI, artificial intelligence; WfO, Watson for Oncology; CDSS, clinical decision-support system; MTB, multidisciplinary tumor board; TNBC, triple negative breast cancer; NCCN, National Comprehensive Cancer Network; CSCO, Chinese Society of Clinical Oncology; ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology.

Declarations

Author's contributions

YA, RH, KD, SW, WF, ID, KR, and GJ were involved in the conception and study design. RH, KD, LM, and KL were involved in the acquisition of data (systematic and targeted reviews of the literature, study selection, and data collection processing). YA, RH, KD, SW, AP, ID, and GJ were involved in data analysis and interpretation (summarization of results from selected studies, discussion, and interpretation of results). YA, AP, RH, KL, ID, and GJ were involved in writing the manuscript. All authors were involved in the review and revision of the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

We acknowledge the support of IBM Watson Health for the completion of this targeted review.

Availability of data and materials

The publications of the 10 peer-reviewed studies included in the current targeted review are publicly available.

Ethics approval and consent to participate

To the best of our knowledge, all 10 peer-reviewed published studies included in the current targeted review were conducted in accordance with the principle of the International Conference of Harmonization Good Clinical Practice (IHC GCP) guidelines and the Declaration of Helsinki. The study protocol for each included study was approved by a local Ethics Committee or by an Institutional Review Board (IRB).

Consent for publication

Not applicable.

Competing Interests

At the time of completion of the final manuscript, all authors were full-time employees of IBM.

Funding

The study was funded and supported by IBM Watson Health.

References

1. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global cancer in women: burden and trends. *Cancer Epidemiol Biomarkers Prev.* 2017;26(4):444–57. Doi:10.1158/1055-9965.Epi-16-0858.
2. The Lancet Editorial. Breast cancer in developing countries. *The Lancet.* 2009;374(9701):1567.
3. Siegel RL, Miller KD, Jemal A. Cancer Statistics. 2020. *CA Cancer J Clin* 2020;70(1):7–30. Doi:10.3322/caac.21590.
4. Kohler BA, Sherman RL, Howlader N, Jemal A, Ryerson AB, Henry KA, et al. Annual report to the nation on the status of cancer, 1975–2011, featuring incidence of breast cancer subtypes by race/ethnicity, poverty, and state. *Journal of the National Cancer Institute.* 2015;107(6):djv048.
5. Tfayli A, Temraz S, Abou Mrad R, Shamseddine A. Breast cancer in low-and middle-income countries: an emerging and challenging epidemic. *Journal of Oncology*, 2010; article ID 490631, 5 pages. Doi 10.1155/2010/490631.
6. Zafar SN, Siddiqui AH, Channa R, Ahmed S, Javed AA, Bafford A. Estimating the global demand and delivery of cancer surgery. *World J Surg.* 2019. 10.1007/s00268-019-05035-6. doi.
7. American Society of Clinical Oncology. The state of cancer care in America, 2014: a report by the American Society of Clinical Oncology. *J Clin Onc Pract.* 2014;10:119–42.
8. Giuliano EA, Hunt KK, Ballman KV, Beitsch PD, Whitworth PW, Blumencranz PW, et al. Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis a randomized clinical trial. *JAMA.* 2011;305:569–75.

9. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med*. 2018;379:111–21.
10. Finn RS, Martin M, Rugo HS, Jones S, Im S-A, Gelmon K, et al. Palbociclib and letrozole in advanced breast cancer. *N Engl J Med*. 2016;375:1925–36.
11. National Institutes of Health. .
12. National Institutes of Health. U.S. National Library of Medicine. Clinicaltrials.gov.
13. Elwyn G, Frosch D, Thompson R, et al. Shared decision making: a model for clinical practice. *J Gen Intern Med*. 2012;27(10):1361–7. doi:10.1007/s11606-012-2077-6.
14. International Business. Machines (IBM) Watson™ for Oncology.
15. Kim M, Kim BH, Kim JM, et al. Concordance in postsurgical radioactive iodine therapy recommendations between Watson for oncology and clinical practice in patients with differentiated thyroid carcinoma. *Cancer*. 2019;125(16):2803–09. Doi:10.1002/cncr.32166.
16. Choi YI, Chung JW, Kim KO, et al. Concordance rate between clinicians and Watson for oncology among patients with advanced gastric cancer: early, real-world experience in Korea. *Can J Gastroenterol Hepatol*. 2019, 8072928. Doi:10.1155/2019/8072928.
17. Kim EJ, Woo HS, Cho JH, Sym SJ, Baek JH, Lee WS, et al. Early experience with Watson for oncology in Korean patients with colorectal cancer. *PloS One*. 2019;14:e0213640.
18. Endnote. <https://www.endnote.com>.
19. Kim YY, Oh SJ, Chun YS, Lee WK, Park HK. Gene expression assay and Watson for Oncology for optimization of treatment in ER-positive, HER2-negative breast cancer. *PLoS One*. 2018;13(7):e0200100. doi:10.1371/journal.pone.0200100.
20. Kim D, Kim YY, Lee JH, Chung YS, Choi S, Kang JM, et al. A comparative study of Watson for Oncology and tumor boards in breast cancer treatment. *Korean J Clin Onc*. 2019;15:3–6.
21. McNamara DM, Goldberg SL, Latts L, et al. Differential impact of cognitive computing augmented by real world evidence on novice and expert oncologists. *Cancer Med*. 2019;8(15):6578–84. doi:10.1002/cam4.2548.
22. Somashekhar SP, Yehadka R, C RK, Rajgopal AK, Rauthan A, Patil P. Triple blinded prospective study assessing the impact of genomic: Endopredict and artificial intelligence Watson for oncology (WFO) on MDT's decision of adjuvant systemic therapy for hormone receptor positive early breast carcinoma. *J Clin Oncol*. 2019;37 (suppl; abstr e18013).
23. Zhou N, Zhang CT, Lv HY, Li TJ, Zhu JJ, Jiang M, et al. Concordance study between IBM Watson for oncology (WFO) and clinical practice for breast and lung cancer patients in China. *Ann Oncol*. 2017;28:x170.
24. Yue L, Yang L. Clinical experience with IBM Watson for oncology (WFO) for multiple types of cancer patients in China. *Ann Oncol*. 2017;28:x162.
25. Somashekhar SP, Sepúlveda MJ, Puglielli S, Norden AD, Shortliffe EH, Rohit Kumar C. Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary

- tumor board. *Ann Oncol*. 2018;29:418–23.
26. Zhou N, Zhang CT, Lv HY, Hao CX, Li TJ, Zhu JJ, et al. Concordance Study Between IBM Watson for oncology and Clinical Practice for Patients with Cancer in China. *Oncologist*. 2019;24:812–19.
 27. Xu J, Sun T, Hua S. Concordance assessment of IBM Watson for Oncology with MDT in patients with breast cancer. *Cancer Research*. 2019;79:Abstract P3-14-06.
 28. Suwanvecho S, Suwanrusme H, Sangtian M, Norden AD, Urman A, Hicks A, et al. Concordance assessment of a cognitive computing system in Thailand. *J Clin Oncol*. 2017;35:6589.
 29. Jiang Z, Xu F, Sepúlveda MJ, Li J, Wang H, Liu Z, et al. Concordance, decision impact and guidelines adherence using artificial intelligence in high-risk breast cancer. *J Clin Oncol*. 2018;36:18566.
 30. Suwanrusme H, Issarachai S, Umsawasdi T, Suwanvecho S, Decha W, Danwka- Mullan I, et al. Concordance assessment of a clinical decision support software in patients with solid tumors. *J Clin Oncol*. 2018;36:18584.
 31. Pan H, Tao J, Qian M, Zhou W, Qian Y, Xie H, et al. Concordance assessment of Watson for oncology in breast cancer chemotherapy: first China experience. *Transl Cancer Res*. 2019;8:389–401.
 32. Suwanvecho S, Shortliffe EH, Suwanrusme H, Issarachai S, Jirakulaporn T, Taechakraichana N, et al. A blinded evaluation of a clinical decision support system at a regional cancer care center. *J Clin Oncol*. 2019;37:6553.
 33. Stephens MR, Lewis WG, Brewster AE, Lord I, Blackshaw GR, Hodzovic I, et al. Multidisciplinary team management is associated with improved outcomes after surgery for esophageal cancer. *Dis Esophagus*. 2006;19:164–71.
 34. Horvath LE, Yordan E, Malhotra D, Leyva I, Bortel K, Schalk D, et al. Multidisciplinary care in the oncology setting: historical perspective and data from lung and gynecology multidisciplinary clinic. *J Oncol Pract*. 2010;6:21–6.
 35. Fennell ML, Prabhu Das I, Clauser S, Petrelli N, Salner A. The organization of multidisciplinary care teams: modeling internal and external influences on cancer care quality. *J Natl Cancer Inst Monogr*. 2010;40:72–80.
 36. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115–32. Doi:10.3322/caac.21338.
 37. Kux BR, Majeed RW, Ahlbrandt J, Rohrig R. Factors influencing the implementation and distribution of clinical decision support systems (CDSS). *Stud Health Technol Inform*. 2017;243:127–31.
 38. Varghese F, Wong J. Breast cancer in the elderly. *Surg Clin North Am*. 2018;98(4):819–33.
 39. Yu Z, Wang Z, Ren X, et al. Practical exploration and research of Watson for oncology clinical decision support system in real-world and localized practice. *J Clin Oncol*. 2019;37.suppl;abstr e18304.
 40. Somashekhar SP, Sepulveda MJ, Shortliffe EH, Rohit Kumar C, Rauthan A, Patil P. A prospective blinded study of 1,000 cases analyzing the role of artificial intelligence: Watson for oncology and

change in decision making of a multidisciplinary tumor board (MDT) from a tertiary care cancer center. J Clin Oncol. 2019;37:6533.

Figures

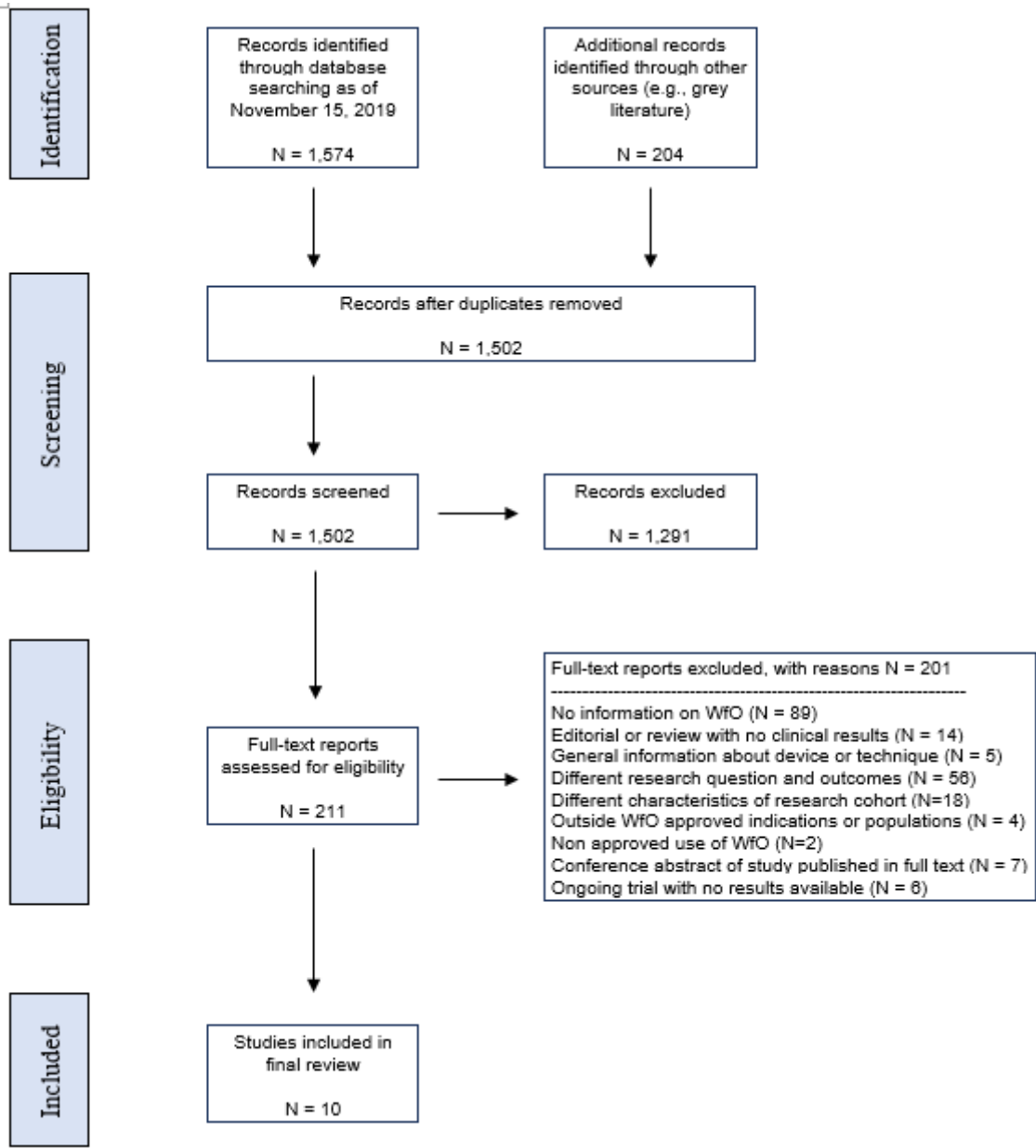


Figure 1
Study identification and selection flow diagram. Abbreviations: WfO, Watson for Oncology.