

Concordance Study Between IBM Watson for Oncology and Clinical Practice for Patients with Cancer in China

Na Zhou,^a Chuan-Tao Zhang,^a Hong-Ying Lv,^a Chen-Xing Hao,^a Tian-Jun Li,^a Jing-Juan Zhu,^a Hua Zhu,^a Man Jiang,^a Ke-Wei Liu,^a He-Lei Hou,^a Dong Liu,^a Ai-Qin Li,^b Guo-Qing Zhang,^c Zi-Bin Tian,^d Xiao-Chun Zhang^a

Departments of ^aMedical Oncology, ^bOut-patient, ^cThoracic Surgery, and ^dGastroenterology, the Affiliated Hospital of Qingdao University, Qingdao University, Qingdao, People's Republic of China

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Artificial Intelligence • Watson for Oncology • Concordance • China

Abstract _

Background. IBM Watson for Oncology (WFO), which can use natural language processing to evaluate data in structured and unstructured formats, has begun to be used in China. It provides physicians with evidence-based treatment options and ranks them in three categories for treatment decision support. This study was designed to examine the concordance between the treatment recommendation proposed by WFO and actual clinical decisions by oncologists in our cancer center, which would reflect the differences of cancer treatment between China and the U.S.

Patients and Methods. Retrospective data from 362 patients with cancer were ingested into WFO from April 2017 to October 2017. WFO recommendations were provided in three categories: recommended, for consideration, and not recommended. Concordance was analyzed by comparing the treatment decisions proposed by WFO with those of the multidisciplinary tumor board.

Concordance was achieved when the oncologists' treatment decisions were in the recommended or for consideration categories in WFO.

Results. Ovarian cancer showed the highest concordance, which was 96%. Lung cancer and breast cancer obtained a concordance of slightly above 80%. The concordance of rectal cancer was 74%, whereas colon cancer and cervical cancer showed the same concordance of 64%. In particular, the concordance of gastric cancer was very low, only 12%, and 88% of cases were under physicians choice.

Conclusion. Different cancer types showed different concordances, and only gastric cancers were significantly less likely to be concordant. Incidence and pharmaceuticals may be the major cause of discordance. To be comprehensively and rapidly applied in China, WFO needs to accelerate localization. ClinicalTrials.gov Identifier: NCT03400514. **The Oncologist** 2019;24:812–819

Implications for Practice: IBM Watson for Oncology (WFO) has begun to be used in China. In this study, concordance was examined between the treatment recommendation proposed by WFO and clinical decisions for 362 patients in our cancer center, which could reflect the differences of cancer treatment between China and the U.S. Different cancer types showed different concordances, and only gastric cancers were significantly less likely to be concordant. Incidence and pharmaceuticals may be the major causes of discordance. To be comprehensively and rapidly applied in China, WFO needs to accelerate localization. This study may have a significant effect on application of artificial intelligence systems in China.

Introduction _

Artificial Intelligence has increased its penetration in the support of our lives. In the field of medicine, computational analysis tools and decision support systems can help with clinical processes and management of medical data and knowledge. Its applications range from assistant tools

for diagnosis and investigation of disease to therapeutic procedures [1].

IBM Watson for Oncology (WFO), trained by Memorial Sloan Kettering Cancer Center (MSKCC), uses specific attributes found in a patient's case to identify potential cancer

Correspondence: Xiaochun Zhang, M.D., Ph.D., Department of Medical Oncology, the Affiliated Hospital of Qingdao University, No. 59 Haier Road, Qingdao, People's Republic of China. Telephone: 86-0532-82913271; e-mail: zhangxiaochun9670@126.com Received April 28, 2018; accepted for publication July 6, 2018; published Online First on September 4, 2018. http://dx.doi.org/10.1634/theoncologist.2018-0255

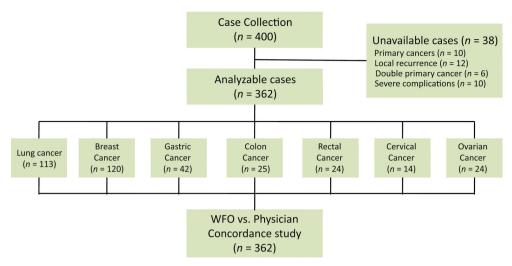


Figure 1. CONSORT diagram. Abbreviation: WFO, Watson for Oncology.

treatment options for physicians to use when making patient care decisions. The treatment options are generally consistent with the National Comprehensive Cancer Network guidelines and are supported by the MSKCC-curated literature reflecting MSKCC experience and expertise. Links are provided to the MSKCC-curated literature supporting each cancer treatment option, as well as supplemental information from the published medical literature, clinical trials, and manufacturers prescribing information for oncology drugs. At present, Watson for Oncology has been applied in 14 countries worldwide, including China, the U.S., Holland, Thailand, India, Korea, Poland, Slovakia, and Bangladesh. In a double-blind study involving 362 patients of Manipal Comprehensive Cancer Centre in India, treatment recommendations from WFO had a high degree of consistency with the Centres multidisciplinary tumor board [2].

Our group recently reported the concordance between Watson for Oncology and clinical practice in our cancer center for patients with breast and lung cancer [3]. In this study, we have expanded the sample size and cancer spectrum using the updated version of Watson for Oncology to explore the concordance of the suggested therapeutic regimen between WFO and physicians in our cancer center, which could reflect the similarities and differences between the East and West in the treatment of cancer.

MATERIALS AND METHODS

Study Design

This study was approved by the Affiliated Hospital of Qingdao University ethics committee. We randomly selected patients with cancer, including those with lung cancer, breast cancer, gastric cancer, colon cancer, rectal cancer, cervical cancer, and ovarian cancer, from our institutional database according to the criteria of Watson for Oncology (supplemental online Appendix 1). A total of 400 patients were enrolled between April 2017 and October 2017. Case data

were extracted and entered into the Watson system. WFO provided therapeutic recommendations in three categories: recommended, for consideration, and not recommended. Data were analyzed retrospectively to compare the WFO's recommendations and actual therapeutic regimen in our hospital. It should be noted that in the data analysis process, we found some actual regimen applications that were not available in WFO, which were defined as "physician's choice." Overall, physicians recommendations were defined as concordant with WFO if they corresponded to the recommended or consideration categories and were defined as nonconcordant if they corresponded to the not recommended or not available categories.

Statistical Analysis

Differences between the baseline clinicopathological characteristics of the groups were assessed using Pearson's χ^2 or Fisher's exact test and indicated accordingly. A logistic regression model was estimated to control for some determinants of concordance, with odds ratios and 95% confidence intervals reported. All analyses were performed with SPSS statistical software (version 18.0; IBM, Chicago, IL).

RESULTS

Baseline Characteristics of Patients

Of 400 eligible patients, 362 were recruited for the study. The whole cancer spectrum and reasons for nonparticipation are shown in Figure 1. Among the participants, 113 (31.2%) patients with lung cancer, 120 (33.1%) with breast cancer, 42 (11.6%) with gastric cancer, 25 (6.9%) with colon cancer, 24 (6.6%) with rectal cancer, 14 (3.9%) with cervical cancer, and 24 (6.6%) with ovarian cancer were successfully recruited. The baseline clinicopathological characteristics of the 362 patients pooled from our database are listed in Table 1.

Table 1. Characteristics of patients using Watson for Oncology

Characteristics	Lung cancer, n (%)	Breast cancer, n (%)	Gastric cancer, n (%)	Colon cancer, n (%)	Rectal cancer, n (%)	Cervical cancer, n (%)	Ovarian cancer, n (%)
Age, median (range), yr	61 (44–78)	59.5 (28–74)	60 (30–75)	55.5 (35–76)	50.5 (34–68)	61.5 (39–69)	56 (27–80)
Age distribution, yr							
<65	78 (69.03)	113 (94.17)	31 (73.81)	20 (80.00)	19 (79.17)	11 (75.57)	20 (83.33)
≥65	35 (30.97)	7 (5.83)	11 (26.19)	5 (20.00)	5 (20.83)	3 (21.43)	4 (16.67)
Sex							
Male	72 (63.72)	0 (0)	31 (73.81)	20 (80.00)	15 (62.50)	0 (0)	0 (0)
Female	41 (36.28)	120 (100)	11 (26.19)	5 (20.00)	9 (37.50)	14 (100)	24 (100)
TNM stage							
I–II	8 (7.08)	80 (66.67)	3 (7.14)	4 (16.00)	1 (4.17)	6 (42.86)	3 (12.50)
III	33 (29.20)	39 (32.5)	21 (50.00)	9 (36.00)	12 (50)	2 (14.28)	8 (33.33)
IV	72 (63.72)	1 (0.83)	18 (42.86)	12 (50.00)	11 (45.83)	6 (42.86)	13 (54.17)
Surgery							
Yes	22 (19.47)	120 (100)	30 (71.43)	23 (92.00)	21 (87.50)	8 (57.14)	20 (83.33)
No	91 (80.53)	0 (0)	12 (28.57)	2 (8.00)	3 (12.50)	6 (42.86)	4 (16.67)

Lung Cancer

By the histology of 113 patients with lung cancer, 22% had small cell lung cancer (SCLC) and 78% had non-small cell lung cancer (NSCLC). Both age and sex distributions were concurrent between histology groups and among tumor stage groups (supplemental online Table 1).

Overall, the treatment recommendations were concordant in 81.3% (93 of 113) of cases. The concordance of patients with SCLC was 92%; the concordance of patients with NSCLC was 79.99% (Fig. 2A). According to tumor stage, treatment recommendations were concordant in 87.5% of patients at stage II, 75.8% at stage III, and 84.6% at stage IV (Fig. 2B). There was no significant difference among different stages and histologies of lung cancer (Table 2).

Breast Cancer

Because it is essentially different to treat breast cancer via molecular classification, we analyzed the concordance according to the molecular classification. Luminal B type was found in the majority of breast cancer cases, which accounted for 66.67%; triple-negative breast cancer (TNBC) accounted for 26.67%, and only eight patients had luminal A type. There were no significant differences observed between menstrual states (supplemental online Table 2).

Overall, the treatment recommendations were concordant in 64.2% (77 of 120) of cases. The concordance of breast cancer patients with luminal B and TNBC was 68.8% and 59.4%, respectively. However, treatment recommendations were concordant in only 37.5% (three of eight) of patients with luminal A, approximately half of that with luminal B (Fig. 2C). According to tumor stage,

treatment recommendations were concordant in 65% of patients at stage II and 64.1% of stage III. Treatment regimen from the only patient at stage IV was under physician's choice (Fig. 2D). There was also no significant difference among different stages and histologies of breast cancer (Table 3).

Gastrointestinal Tumor

Because of the similarity of the physiological system and therapeutic regimen, we evaluated gastric cancer and colorectal cancer together in this paper. The overall treatment recommendations were concordant in only 11.9% (5 of 42) of patients with gastric cancer, and up to 83.33% of the therapeutic regimen was physician's choice (Fig. 2E). Treatment recommendation concordance was also not very high in colon cancer (40%), and 52% were physician's choice (Fig. 2F). As for rectal cancer, 74% of cases were concordant with the actual therapeutic regimen, and 12.5% were physician's choice (Fig. 2G).

Gynecological Tumor

Watson was able to support two types of gynecological cancer, cervical cancer and ovarian cancer. Of the 14 patients with cervical cancer, treatment recommendations were concordant only in 50%, and the remaining half were physician's choice. By contrast, the treatment recommendation concordance of ovarian cancer was up to 95.83%, and only 4.17% cases were physician's choice. Concordance regarding the different tumor stages is shown in Figure 2H and I.



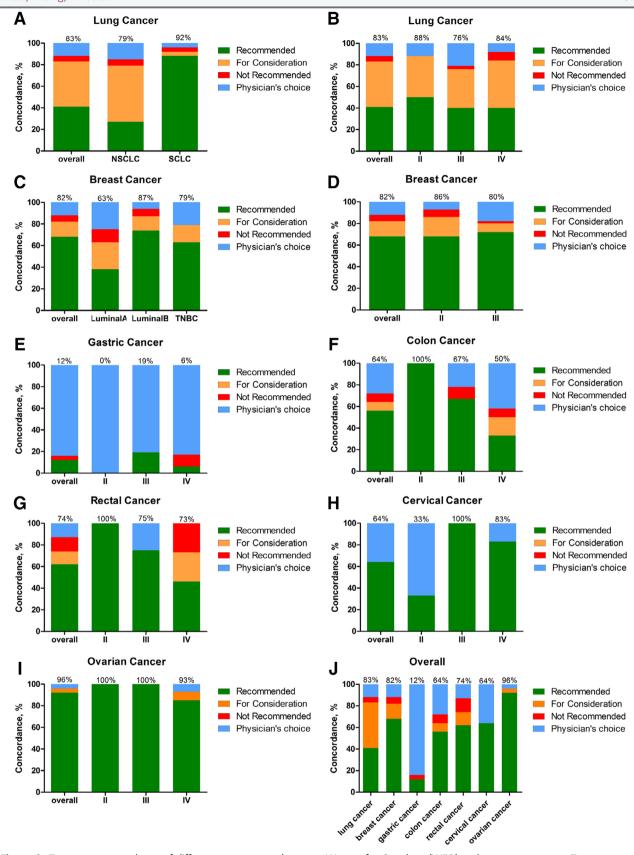


Figure 2. Treatment concordance of different cancer types between Watson for Oncology (WFO) and our cancer center. Treatment was considered concordant if the delivered treatment was rated as either recommended or for consideration by WFO and discordant if the delivered treatment was either not recommended by WFO or was physician's choice (not included in WFO). Concordance of lung cancer by histology (A) and tumor stage (B). Concordance of breast cancer by histology (C) and tumor stage (D). Concordance of gastric cancer (E), colon cancer (F), rectal cancer (G), cervical cancer (H), ovarian cancer (I), and overall (J).

Abbreviations: NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; TNBC, triple-negative breast cancer.

Table 2. Logistic regression model of concordance between different stages and histology of lung cancer

	- 1.1 -1 (a=a) an		
Variable	Odds ratio (95% CI)	p value	
Stage			
II	0.852 (0.094-7.767)	.887	
III	2.482 (0.840-7.333)	.100	
IV (reference)	1.000		
Histology			
NSCLC	3.591 (0.432–17.614)	.115	
SCLC (reference)	1.000		

Abbreviations: CI, confidence interval; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer.

Overall Summary

Overall, different cancer types showed different concordances. As shown in Figure 2J, lung cancer, breast cancer, and ovarian cancer reached a concordance over 80%, and rectal cancer showed a concordance of over 60%. Of note, concordance of gastric cancer was very low, only 12%, and 88% of cases were physicians choice. The logistic regression revealed that only gastric cancers were significantly less likely to be concordant (p < .001; Table 4).

Treatment decisions for gastric cancer are shown in Figure 3. Therapy containing TIJI'AO was the majority, accounting for 62.80%, and therapy containing LIPUSU accounted for 25.57%. These two therapies composed the reason for low concordance, which involved nationally available pharmaceuticals.

Case Report of Reasonable Application of WFO

A 35-year-old lactating woman was referred to our hospital and presented with epigastric pain and dizziness. The upper gastrointestinal endoscopy biopsy results yielded a diagnosis of a low-differentiated adenocarcinoma. A positron emission tomography/computed tomography scan revealed irregular thickening and increased metabolism of the side wall in the lower gastric body, which was unclear with the hepatic left lateral. Moreover, there were highmetabolism and low-density lesions adjacent to the left lobe of liver (43 \times 67 mm). All of the above evidence indicated a diagnosis of gastric cancer involving the left hepatic region.

We input her information into the WFO system and clicked "Ask Watson." Watson recommended two regimens, "Dose Modified DCF" (docetaxel, cisplatin, fluorouracil, and leucovorin) and "FOLFOX" (fluorouracil, leucovorin, and oxaliplatin; Fig. 4A). Considering that the patient was lactating, the DCF regimen would be aggressive and inconvenient for her. We chose paclitaxel liposome 240 mg on day 1 and S-1 60 mg, twice a day on days 1–14, which was improved at the basis of DCF. Paclitaxel liposome is a specific dosage form designed in China and has a slight side effect compared to classical paclitaxel. S-1, which is made in China, has the same efficacy as fluorouracil and leucovorin. Moreover, oral medicine is more convenient than intravenous medicine.

Table 3. Logistic regression model of concordance between different stages and histology of breast cancer

Variable	Odds ratio (95% CI)	p value	
Stage			
II	0.818 (0.279-2.404)	.715	
III (reference)	1.000		
Histology			
Luminal A	2.630 (0.416–16.608)	.304	
Luminal B	0.857 (0.270-2.725)	.794	
TNBC (reference)	1.000		

Abbreviations: CI, confidence interval; TNBC, triple-negative breast cancer.

After three cycles of neoadjuvant chemotherapy, the patient underwent a laparoscopic total gastrectomy plus partial hepatectomy surgery under general anesthesia. Postoperative pathology indicated low differentiated adenocarcinoma in the anterior wall of the gastric body and tumor invasion into the serosa without involvement of the liver tissues; cutting edge and margin of the liver showed no cancer tissue, nor did the omentum tissue. Four weeks after surgery, we updated the patients disease history in the WFO system to ask Watson for the next step and to confirm the effect of the previous choice and chemotherapy. Watson recommended no further treatment, which was within our expectations (Fig. 4B).

Discussion

IBM Watson is one of the leading artificial intelligence (AI) or cognitive technologies and not only can learn to reason and understand the enormous corpus of the literature available to the scientific community but can also make connections among all of the data to answer a complex medical question in a very short amount of time, resulting in evidence-based and personalized treatment options [4]. Watson for Oncology is based on disease history, whereas Watson for Genomics is based on genomic sequencing data. In addition to its tremendous potential and recent advances, AI technology faces many obstacles before it can reach widespread use in medicine. It must be sufficiently accepted by users and integrated into the physician's workflow. The second challenge is to determine the accuracy of the system for diagnoses and treatment recommendations. Another major challenge involves concerns about patient privacy and security [5].

In the process of generalization, the most important factor is the concordance between WFO and physicians, which determines the accuracy of the system for diagnoses and treatment recommendations. MSKCC trained WFO and worked on improving the accuracy of WFO. From 2013 to 2015, MSKCC published the training results at every American Society for Clinical Oncology (ASCO) Annual Meeting. Concordance studies have been performed in various countries. The results of a double-blind study with 638 patients with breast cancer from India indicated that 93% of



Table 4. Logistic regression model of concordance between different cancers

Variable	Odds ratio (95% CI)	p value
Lung cancer	0.606 (0.213–1.728)	.349
Breast cancer	0.719 (0.257–2.013)	.530
Gastric cancer	22.200 (5.968-82.581)	<.001
Colon cancer	1.687 (0.492-5.791)	.406
Gynecological cancer	0.606 (0.213-1.728)	.349
Ovarian cancer	0.130 (0.014–1.183)	.070
Rectal cancer (reference)	1.000	

Abbreviation: CI, confidence interval.

Watson for Oncology's recommendations for standard treatment or consideration were concordant with the recommendations of the tumor board [6]. A study presented at the 2017 ASCO Annual Meeting of 525 patients from Korea showed a 73% concordance rate for colon cancer and a 49% concordance rate for gastric cancer [7]. It appears that the concordance varies by countries and cancer types. China has the largest population in the world and a particular cancer spectrum. Moreover, development of national medicine and different local conditions and customs form different therapeutic experiences and considerations. Therefore, concordance data of Chinese patients are essential for accurate examinations and improvements, which could illustrate the differences in cancer therapy between the Eastern and Western world.

We recruited all cancer types that WFO version17.4 could cover, analyzed their concordance, and discussed the reasons resulting in the difference. As shown in Figure 2J, lung cancer and ovarian cancer were the two most concordant cancer types. Unlike ovarian cancer, the high concordance of lung cancer included one-half "for consideration" options (yellow) for NSCLC (Fig. 2A). The major reason is that following the in-depth study of immunotherapy and approval of PD-1/PD-L1 antibody drugs in the U.S., the WFO would recommend three immunotherapy drugs, pembrolizumab, nivolumab and atezolizumab, for metastatic NSCLC, but chemotherapy was the only regimen available for consideration. Since the China Food and Drug Administration (CFDA) has not approved these immunotherapy drugs, we currently must select chemotherapy. Another reason is that vinorelbine, a type of chemotherapy drug, has not yet been used in China but is used in the U.S. In addition, we selected the n addition, weed"dn adtment (red) for some patients, mainly because patients with the EGFR-resistant mutation could not afford the expensive osimertinib but selected chemotherapy, which was not recommend by the WFO. Breast cancer and lung cancer have nearly the same concordance. It is important to explain what caused the breast cancer discordance. In China, Oncotype Dx (Genomic Health, Inc., Redwood City, CA) has not yet been applied popularly as a regular test to evaluate the necessity of chemotherapy. Thus, we chose chemotherapy as the first-line therapy, whereas WFO recommended endocrine therapy. Another major difference

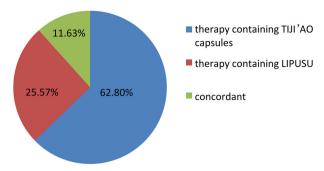


Figure 3. Distribution of actual clinical decisions for gastric cancer.

is the selection of doxorubicin/cyclophosphamide (AC) and docetaxel/cyclophosphamide (TC). We prefer TC for some postmenopausal patients at stage II, which has shown superiority to standard AC (disease-free survival and overall survival) and is tolerable in both older and younger patients [8]. In contrast, WFO recommended AC for these patients, and the explanation is that TC chemotherapy has not yet been prospectively compared with anthracyclineand taxane-containing regimens or to cyclophosphamide/ methotrexate/fluorouracil (CMF). Of all cancer types, gastric cancer showed the significantly lowest concordance (Fig. 2E and J), the incidence of which is obviously different between Eastern and Western countries. Doctors in China have their own experience and habits for treating patients with gastric cancer. WFO recommended the FOL-FOX regimen as the first-line chemotherapy, whereas we selected the SOX regimen (oxaliplatin 85 mg/m² on day 1 and S-1 40 mg/m² on days 1-14) which contained an oral medicine, nationally produced S-1 called TIJIroduced S-1, a combination of three pharmacologic compounds, namely, tegafur, gimeracil, and oteracil potassium (Fig. 3). That is more convenient than intravenous medications. This was also one of the reasons underlying the colon cancer discordance. The other reason is another intravenous chemotherapy drug, raltitrexed, which is produced and used only in China (data not shown). A similar reason could also explain cervical cancer. LIPUSU, an injected paclitaxel liposome, which could greatly reduce the incidence of anaphylaxis and enhance the stability and targeting of paclitaxel, has been developed and applied uniquely in our country.

CONCLUSION

The concordance between WFO and physicians in our cancer center is not as high as previously reported in other countries. However, these data could not deny the accuracy of WFO or the physicians' professionalism. This study illustrates the difference of the present cancer therapy situation between the East and West. The difference can be summarized as follows: (a) the incidence of gastric cancer in China is much higher than that in the U.S., and thus, experience and drug development are more skilled in our country; (b) Chinese patients could not benefit from some advanced new targeted drugs and

Α

Patient Cases / Clinical Information

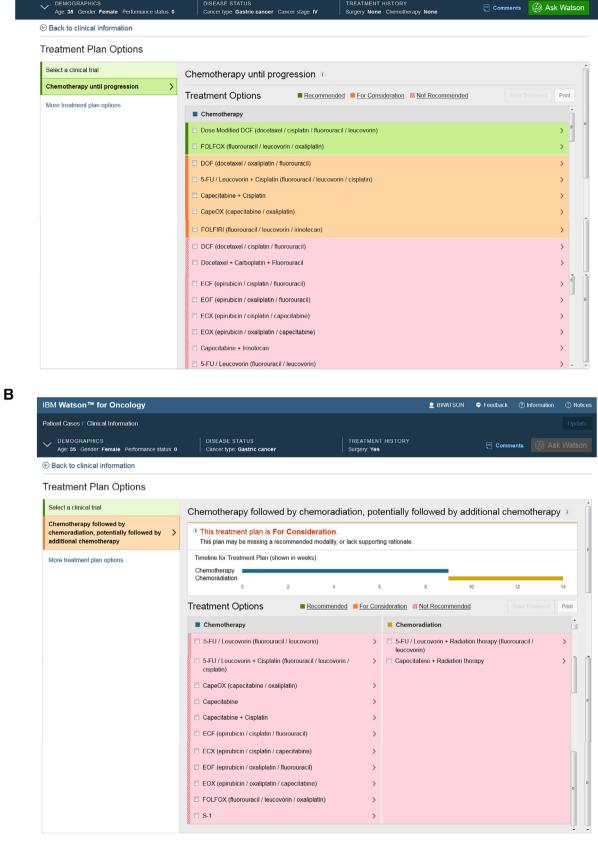


Figure 4. Treatment regimens produced by Watson for Oncology (WFO) for a patient. WFO treatment regimens of the first diagnosis (A) and after clinical therapy (B).



immunotherapy drugs developed in the U.S because of price and CFDA approval; and (c) China has its own national pharmaceuticals and traditional medicine; thus, there will be difference in the selected regimens. Therefore, it is necessary to accelerate the localization of WFO before it can be comprehensively and rapidly applied in China.

ACKNOWLEDGMENTS

This work was supported by a grant obtained from the Taishan Scholar Foundation (tshw201502061 to X.Z.), Qingdao People's Livelihood Science and Technology program (16-6-2-3-nsh to X.Z.) and the Qingdao Leading Innovation Team Foundation.

The authors acknowledge all of the staff at IBM Corporation and Baheal Corporation in China for their support in running the trial and data management.

AUTHOR CONTRIBUTIONS

Conception/design: Na Zhou, Xiao-Chun Zhang

Provision of study material or patients: Chuan-Tao Zhang, Hong-Ying Lv, Tian-Jun Li, Jing-Juan Zhu, Ke-Wei Liu

Collection and/or assembly of data: Chen-Xing Hao, Hua Zhu, Ai-Qin Li, Guo-Qing Zhang. Zi-Bin Tian

Data analysis and interpretation: Na Zhou, Man Jiang, He-Lei Hou, Dong Liu

Manuscript writing: Na Zhou

Final approval of manuscript: Na Zhou, Xiao-Chun Zhang

DISCLOSURES

The authors indicated no financial relationships.

REFERENCES _

- **1.** Makedon F, Karkaletsis V, Maglogiannis I. Overview: Computational analysis and decision support systems in oncology. Oncol Rep 2006; 15:971–974.
- **2.** Somashekhar SP, Sepúlveda MJ, Norden AD et al. Early experience with IBM Watson for Oncology (WFO) cognitive computing system for lung and colorectal cancer treatment. J Clin Oncol 2017;35 (suppl 15):8527A.
- **3.** Zhang XC, Zhou N, Zhang CT 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 (suppl 10): 544PA

- **4.** Curioni-Fontecedro A. A new era of oncology through artificial intelligence. ESMO Open 2017; 2:e000198.
- **5.** Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cariol Rep 2014;16:441.
- **6.** Somashekhar SP, Sepúlveda MJ, Puglielli S et al. Watson for Oncology and breast cancer

treatment recommendations: Agreement with an expert multidisciplinary tumor board. Ann Oncol 2018;29:418–423.

- **7.** Baek JH, Ahn SM, Urman A et al. Use of a cognitive computing system for treatment of colon and gastric cancer in Korea. J Clin Oncol 2017;35 (suppl 15):e18204A.
- **8.** Jones S, Holmes FA, O'Shaughnessy J et al. Doxetaxel with cyclophosphamide is associated with an overall survival benefit compared with doxorubicin and cyclophosphamide: 7-year follow-up of US Oncology Research Trial 9735. J Clin Oncol 2009;27:1177–1183.



See http://www.TheOncologist.com for supplemental material available online.