Can Watson for Oncology replace oncologists: A comparative study between Watson for Oncology and our multidisciplinary tumor board

Chaoyuan Liu¹ MS, Xianling Liu¹ PhD, Fang Wu¹ PhD, Mingxuan Xie² PhD, Yeqian Feng¹ PhD, Chunhong Hu¹ PhD

¹ Department of Oncology, the Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China

² Department of Geriatric Medicine, Xiangya Hospital, Central South University, Changsha, Hunan 410008, P.R. China.

*Corresponding author:
Chunhong Hu
Department of Oncology
the Second Xiangya Hospital
Central South University
No.156 Renmin Road, Changsha, 410011, Hunan Province, P. R. China.

Phone: +86 13508486908
Fax: +86 073185295929
Email: huchunhong@csu.edu.cn
Abstract

Background: Artificial intelligence (AI) is developing quickly in the medical field, which gives convenience to both medical staff and patients. The AI CDSS Watson for Oncology (WFO) is an outstanding representative of AI in medical field, it can give prompt treatment recommendations to cancer patients, just as an excellent oncologist. It is more and more widely used in China, but there is no report about whether it suits Chinese patients. Here we report a retrospective study about the consistency of WFO's recommendations and commendations of the same patient given by our multi-disciplinary team board (MDT) at our center on lung cancer patients' treatment.

Objective: Our objective was to explore practicability of the WFO to the lung cancer cases in China and how to make it more suitable for Chinese patients with lung cancer.

Methods: We selected all the lung cancer patients who were hospitalized and received antitumor treatment in the 2nd Xiangya hospital cancer center for the first time from September to December in 2017. WFO gave treatment recommendations to all qualified cases. If our actual therapeutic regimen (which was given by our MDT) was “recommended” or “for consideration” in WFO, we thought it was of consistency, if it was “not recommended” or WFO didn’t have this option, we
thought it was of no consistency. Blind second round reviews were made by our MDT to reassess the incongruent cases.

**Results:** WFO didn’t support 18% of all the cases (n=182). Of the 149 cases (82%) supported, 66% were consistent with our MDT recommendations. Subgroup analysis showed that pathological type and staging had a significant effect on consistency (p=0.004, p=0, respectively). Age, gender, whether there was gene mutation or not had no effect on consistency. In 81% of the inconsistent cases, our MDT gave two treatments with Chinese characteristics, which were different from recommendations given by WFO but were also with excellent effect. If WFO can bring the two alternative treatments into the recommended or considered range, the overall consistency can be elevated from 66% to 93%.

**Conclusions:** In China, the majority of the treatment recommendations of WFO are consistent with the recommendation of the expert group, but there is still a relatively high proportion of cases which are not supported by WFO. WFO cannot substitute our oncologists currently. As doctors’ assistant, it can improve the efficiency of our work, and it needs to learn local characteristics of patients to become a better assistant.

**Key words:** Watson for Oncology, artificial intelligence, lung cancer, concordance, multidisciplinary tumor board

**Introduction**

The contradiction between high quality medical resources and people’ medical need
is becoming increasingly prominent, which has caused many medical conflicts in China and even attracted the attention of the international community [1-4]. There are several reasons for this. First, junior doctors usually give poor diagnosis and treatment, who can’t keep up with the development in tumor field. As a matter of fact, medical data, papers and guidelines in the field of tumor are growing really too fast nowadays, but doctors’ learning time are limited. In October 2017 alone, the FDA approved 69 drugs for breast cancer alone, and the NCCN guide for lung cancer was updated 9 times only in 2017. Studies have shown that oncologists only spent 4.6 hours a week to learn professional knowledge [5], compared with other ordinary clinical disciplines, specialist needs to be more timely follow up and get a lot of evidence of evidence-based medicine, to support the patients’ individualized treatment plan. Even an oncologist who major in a specialized field can’t master all the knowledge, not mention the doctors at the primary level who should tackle with various tumor types. Second, distribution of medical resources is not balanced. The same patient may have different treatment recommendation in different hospitals; even in the same hospital, different doctors offer different treatment options. A study pointed out the Chinese government needs to strengthen the medical professionalism for young physicians [6]. A device which can help Chinese doctors give right treatment recommendations quickly or help doctors learning is urgently needed.

IBM’s Watson for Oncology (WFO) is a cognitive-support system for oncology therapy selection, which has been described previously by Somashekar SP [7]. In short, WFO
has stored, indexed amounts of literature, protocols, and patient charts, and learned from test cases and experts from Memorial Sloan Kettering Cancer Center (MSKCC) and can use computational reasoning approaches to apply it to a specific case [7], all the information input is verified by the top oncologists in MSKCC. What’s more, it will renew its knowledge according to the cutting edge information every one to two months. The treatment recommendations given by WFO are categorized into three groups: “recommended” which is with strong evidence, “for consideration” which maybe a suitable alternative, “not recommended” which is due to contraindications or strong evidence against its use [7].

It seems that WFO is the right choice to settle our difficulties, it only needs us to input data of a case, then it will tell us the world’s most standard treatment to the specified case in a minute, and display the highly consistent evidences [7]. As a matter of fact, WFO is becoming more and more prevalent in China today, but there is no report about to what extent it suits Chinese patients so far.

The morbidity and mortality of lung cancer is the highest in malignant tumors in the world or in China[8]. In order to explore practicability of the WFO to the lung cancer cases in China as well as how to improve WFO, in other words, how to make it more suitable for Chinese patients with lung cancer, we conducted a retrospective study on lung cancer in the Xiangya 2nd hospital cancer center, Central South University, which compared the treatment advices given by WFO and the actual treatment provided by the MDT in our center. Here, we report the specific content and results of the research.
Methods

Study design

We selected the inpatients with lung cancer who were hospitalized and received antitumor treatment for the first time in September 2017 to December 2017 in the Xiangya 2nd hospital cancer center. A total of 182 cases were included, the pathological types including squamous carcinoma, adenocarcinoma, Adenosquamous carcinoma and large cell carcinoma and small cell carcinoma, the received treatment including postoperative adjuvant therapy, definitive therapy and the best supportive therapy. 33 cases were not supported by Watson, and the remaining 149 patients were included in our comparison study. Treatment recommendations were given by WFO to these cases, we then compared them to the actual treatment we adopted which are recommended by the MDT of our department. The study protocol was reviewed and approved by the Second Xiangya Hospital institutional review board.

Data collection and consistency judgment

Patients’ information and specific treatments were collected from the hospital medical system, a senior physician who was blind to the actual treatments input the information into WFO (IBM Watson 17.1) manually and recorded the recommendation given by WFO. Another doctor compared the WFO’s recommendation and the actual treatment. If our actual therapeutic regimen was "recommended" or "for consideration" in WFO, we marked it with consistency, if it was "not recommended "or the WFO didn't not have this option, we marked it with
no consistency. Our team of specialists in our cancer center reassessed the incongruent cases and gave the reasons why they choose the actual regimens.

Data analysis and statistics

We used Microsoft Excel and SPSS 23.0 to describe and statistic data, and a logistic regression mode was estimated with odds ratios and 95% confidence intervals.

Results

Demographic characteristics of hospitalized lung cancer patients are shown in Table 1. 18% (33/149) of our patients was not supported by WFO. In all qualified cases, non-small cell lung cancer accounted for 85%, small cell lung cancer accounted for 16%, and the proportion was consistent with its lung cancer pathological distribution in the world basically[8]. Median age of our patients was 60, of which 83% were men and 17% were women. Adenocarcinoma with wild-type gene types accounted for 69% (42/61) in all cases, phase III and IV accounted for 81%. After our team of specialists reassessed the incongruent cases, there was no change to the primary concordance.

The general consistency was 66%(shown in Table 2), the consistency for non-metastatic cases was 49%, and for metastatic cases was 89%, for NSCLC was 61%, for SCLC was 85%(shown in Figure 1). Logistic regression analysis showed that age (which was divided into two groups : > 60 years old and 60 years old, p = 0.45), gender (p = 0.3), presence of gene mutation (p = 0.9) had no effect on consistency (shown in Table 3). The following factors shad significant effect on consistency ,they were as
follows: pathological types (p = 0.004), the consistency of small cell carcinoma was 91.3% and the consistency of non-small cell lung cancer was 61%; stage (p = 0), stage I (83%), stage II (59%), stage III (42%), stage IV (89%). There were two major reasons accounted for 81% of the inconsistent cases, firstly, we adopted sequential chemoradiation instead of concurrent chemoradiation; secondly, we adopted Icotinib/endostar instead of the other first generation EGFR-TKI/bevacizumab. If WFO can bring the two alternative treatments into the recommended or considered range, the overall consistency can be elevated from 66% to 93%.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Characteristics of lung cancer cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>149</td>
</tr>
<tr>
<td>male ( % )</td>
<td>124 ( 83 )</td>
</tr>
<tr>
<td>female ( % )</td>
<td>25 ( 17 )</td>
</tr>
<tr>
<td>age ( years ( min, max ) )</td>
<td>60 ( 26-83 )</td>
</tr>
<tr>
<td>pathology n ( % )</td>
<td></td>
</tr>
<tr>
<td>squamous carcinoma</td>
<td>61 ( 41 )</td>
</tr>
<tr>
<td>adenocarcinoma</td>
<td>61 ( 41 )</td>
</tr>
<tr>
<td>Adenosquamous carcinoma</td>
<td>3 ( 2 )</td>
</tr>
<tr>
<td>Small cell carcinoma</td>
<td>23 ( 15.4 )</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>1 ( 0.6 )</td>
</tr>
<tr>
<td>stage ( % )</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>6 ( 4 )</td>
</tr>
<tr>
<td>II</td>
<td>22 ( 15 )</td>
</tr>
<tr>
<td>III</td>
<td>59 ( 40 )</td>
</tr>
<tr>
<td>IV</td>
<td>62 ( 41 )</td>
</tr>
<tr>
<td>gene mutation status ( % )</td>
<td></td>
</tr>
<tr>
<td>mutation</td>
<td>8 ( 5.4 )</td>
</tr>
<tr>
<td>wild type</td>
<td>44 ( 30 )</td>
</tr>
<tr>
<td>unknown</td>
<td>97 ( 64.6 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>MDT and WFO recommendations after the initial reviews</th>
</tr>
</thead>
</table>

### Table 3 Logistic regression model of concordance between WFO and MDT

<table>
<thead>
<tr>
<th>Age</th>
<th>P</th>
<th>OR</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤60 ys</td>
<td>0.45</td>
<td>0.72</td>
<td>0.31-1.70</td>
</tr>
<tr>
<td>&gt;60 ys</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>0.3</th>
<th>0.54</th>
<th>0.17-1.74</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Pathology</th>
<th>NSCLC and SCLC</th>
<th>0.004</th>
<th>0.09</th>
<th>0.02-0.45</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Stage</th>
<th>Reference</th>
<th>0.05</th>
<th>3.51</th>
<th>1.03-12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage4</td>
<td>0</td>
<td>9.5</td>
<td>3.4-26.1</td>
<td></td>
</tr>
</tbody>
</table>

WFO: Watson for Oncology; MDT: multidisciplinary tumor board

Figure 1. Treatment concordance between WFO and the MDT overall and by stage and by pathology categories. MDT, multidisciplinary tumor board; WFO, Watson for Oncology.
Gene mutation (reference)

<table>
<thead>
<tr>
<th>Wild type</th>
<th>0.91</th>
<th>0.91</th>
<th>0.16-5.26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-mesured</td>
<td>0.02</td>
<td>0.32</td>
<td>0.12-0.86</td>
</tr>
</tbody>
</table>

CI: Confidence Interval; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; WFO: Watson for Oncology; MDT: multidisciplinary tumor board

Discussion

This retrospective study suggests that WFO application in lung cancer has a large ascension space. It did not support 18% of the cases, in which 42% (14/33) of cases progressed after targeted therapy. As a matter of fact, there is a huge difference of the gene mutation phenotype of lung cancer between China and western countries. The EGFR mutation rate of lung cancer in European and American countries is about 15%, while probability is 50% or more in China [9,16]. The treatment consistency was 66%, much less than 96.4% which was reported in an abstract on the 2017 ASCO Annual Meeting[17]. There are several reasons for this. First of all, WFO recommend concurrent chemoradiation, while our country adopt sequential chemoradiation (34/51, 67%), because the patient physique of Chinese people is usually weaker, who can’t tolerate concurrent chemoradiation, which is common in China. Secondly, our country uses the drug Icotinib/endostar instead of the other first generation EGFR-TKI/bevacizumab (7/51, 14%). Icotinib and endostar[10-13] are primary research drugs in China, researches have shown that they are as effective as the other first generation EGFR-TKI/ bevacizumab in patients with lung cancer in China[14, 15]. Thirdly, some drugs do not come into the Chinese market, such as immune-targeted therapy drugs with PD1/PD-L1 antibodies as their representatives. In addition,
patients' preferences, prices, and medical insurance will also be taken into
consideration and ultimately affect the inconsistency. What’s more, WFO does not
take some coexistence diseases into consideration. For example, in our study, there
was a patient who was diagnosed with stage III squamous lung carcinoma, who was
also suffering from active tuberculosis, if he received the standard
chemoradiotherapy which was recommended by WFO, TB may spread rapidly, which
may result in a rapid death, so our treatment measure was to take oral anti-TB drugs
prior to chemoradiotherapy. If these individualized information can be incorporated
into WFO, the consistency will be significantly improved.

Although the consistency of WFO in this study is not as high as expected, it cannot
support a part of cases, there is no denying that WFO still has enormous value. First,
it gives evidences to support its advice and arrange these evidences according to
credibility, based on related literature, including works and related data[7]. The
doctor should examine the related evidence to judge whether it can be applied to
the current case. When the doctor selected one treatment, it will also give us survival
rate, incidence of adverse effects and other information related to this solution to
help doctors assess the curative effect of the scheme and risk as a whole. Second, it
can greatly shorten the time of junior doctors to consult relevant literature and
improve their diagnosis and treatment in a short time. Third, as to patients, it can
remove the time to go to various top hospitals and get the best treatment as soon as
possible. Fourth, it can solve the problem of doctor-patient trust, unlike local experts
may give some recommendations out of their own interests, WFO takes no personal
preferences, it is fair, and will win the trust of patients. As a result, cancer patients don’t have to visit a dozen of experts to seek for a regimen which they think fair enough by themselves.

This research is very important. In the artificial intelligence boom, it provides the basis for medical Institutions, medical staff and cancer patients to have a correct cognition of WFO and become rational. Today, with the rapid development of AI, some people worship AI blindly and crazily. It seems that AI can do anything, and it seems to be able to replace doctors rapidly. But our research shows that this is impossible. Medicine is not only a science, but also involves more aspects such as social and psychological factors; doctors need to take individualized measures for different patients, even with the same diagnosis. In the operation of WFO, we can see that WFO needs us to tell it whether the patients can receive radical operation or radiation, whether the patients can tolerate operation or radiation, or whether there is any emergency, all of which must be judged by our oncologists in reality. After WFO gives the treatment recommendations, we need to choose the most suitable treatment plan according to the patient’s physical and mental state, economic situation and the combination of complications and patient’s willing. It is important that WFO can only learn to store existing knowledge and give patients the best treatment in the world at the current level of development. Medical advances still require scientists to keep exploring. In a word, WFO can act as an assistant of doctor. As far as we know, this is the first study about the applying of AI in the treatment of cancer in China.
There are several limitations to our study. Firstly, it is a retrospective observational study with no controls. Secondly, the distribution is imbalance in the groups of patients: fewer patients with stage I, fewer women compared with man, which may be due to the fact that we choose the hospitalized patients, while patients with stage I, after surgery, are only required to receive observation and do not need to be hospitalized for further treatment. The reason for the lack of female patients is that the female patients are mostly with the pathology of adenocarcinoma, most of them are with EGFR mutation, Most of these patients only need to take orally targeted drugs outside the hospital and do not need to be hospitalized either. Thirdly, the bias of input data from experts may cause different treatment recommendations.

AI, on the whole, has strong ability of memorization far stronger than human brain, and can quickly collect and sort the information which it stores, and thus give right conclusions faster than human, such as the application in imaging diagnostic system. However, in order to adapt to the reality of our country, WFO still has a long way to go. We need to standardize the patient's medical data, share the data nationwide, improve the follow-up system to obtain complete information of the patients. In other words, we should build China's unique medical data repository, make WFO study it, and incorporated it with international guidelines and health care system, and then it will reach its potential to serve Chinese people.

Acknowledgements

The authors wish to thank Li Tianle for his support on the free use of WFO, who is
employed by IBM Watson Health.

Conflicts of Interest

None declared.

Disclosure

All authors have declared no conflicts of interest.

References

   PMID: 22579126.


