

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Oncologic big-data platform in promoting digital competencies and professionalism in Chinese medical students: a cross-section study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-061015
Article Type:	Original research
Date Submitted by the Author:	12-Jan-2022
Complete List of Authors:	<p>Liu, Jiahao; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Jiao, Xiaofei; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Zeng, Shaoqing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Li, Huayi; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Jin, Ping; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Chi, Jianhua; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Liu, Xingyu; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Yu, Yang; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Ma, Guanchen; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Zhao, Yingjun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Li, Ming; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Peng, Zikun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Huo, Yabing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p> <p>Gao, Qing-Lei ; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology</p>
Keywords:	MEDICAL EDUCATION & TRAINING, Gynaecological oncology < GYNAECOLOGY, Gynaecological oncology < ONCOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 **1 Oncologic big-data platform in promoting digital competencies and**  
5 **2 professionalism in Chinese medical students: a cross-section study**  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18

3

4 Jiahao Liu<sup>1,2#</sup>, Xiaofei Jiao<sup>1,2#</sup>, Shaoqing Zeng<sup>1,2</sup>, Huayi Li<sup>1,2</sup>, Ping Jin<sup>1,2</sup>, Jianhua Chi<sup>1,2</sup>,  
5 Xingyu Liu<sup>1,2</sup>, Yang Yu<sup>1,2</sup>, Guanchen Ma<sup>1,2</sup>, Yingjun Zhao<sup>1,2</sup>, Ming Li<sup>1,2</sup>, Zikun Peng<sup>1,2</sup>,  
6 Yabing Huo<sup>1,2</sup>, Qinglei Gao<sup>1,2\*</sup>  
7  
8

7

8

9 <sup>1</sup> Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College,  
10 Huazhong University of Science and Technology, Wuhan 430030, Hubei, P. R. China.

11 <sup>2</sup> Cancer Biology Research Center (Key Laboratory of the Ministry of Education),  
12 Tongji Hospital, Tongji Medical College, Huazhong University of Science and  
13 Technology, Wuhan 430030, Hubei, P. R. China.  
14

14

15 # These authors contribute equally.  
16

16

17 \* Correspondence to:

18 Prof. Qinglei Gao, Cancer Biology Research Center (Key Laboratory of the Ministry  
19 of Education), Tongji Hospital, Tongji Medical College, Huazhong University of  
20 Science and Technology, 1095 Jiefang Ave, Wuhan 430030, Hubei, P. R. China. Tel:  
21 +86-27-83662681; Fax: +86-27-83662681; E-mail: qingleigao@hotmail.com.  
22

22

23 Word count (excluding title page, abstract, references, figures and tables): 2,774  
24

24

## 25 **Abstract**

26 **Objectives:** Advancements in big data technology are reshaping the health care system  
27 in China. However, medical students' attitude for big data technology, as well as their  
28 professionalism in the big data age were poorly investigated. This study aims to explore  
29 the role of medical big data in promoting the digital competencies and professionalism  
30 in Chinese medical students.

31 **Design, setting and participants:** This cross-sectional study was performed among  
32 274 medical students who attended a workshop about medical big data conducted on  
33 July 8, 2021 in Tongji Hospital, Tongji Medical College, Huazhong University of  
34 Science and Technology, Wuhan, China. The workshop was based on the first national-  
35 wide multi-function gynecologic oncology medical big data platform in China, the  
36 National Union of Real-World Gynecologic Oncology Research & Patient  
37 Management Platform (NUWA platform).

38 **Outcome measures:** The knowledge and attitude towards big data technology and a  
39 Professionalism Mini-Evaluation Exercise (P-MEX) instrument were collected before  
40 and after the workshop.

41 **Results:** 274 students participated in this workshop and completed all surveys. Before  
42 the workshop, only 27% of them knew the detailed content of medical big data  
43 platforms and 64% knew its potential application. The majority of students believed  
44 that big data technology is practical in their clinical practice (77%), medical educations  
45 (85%), and scientific researches (82%). Over 80% of participants showed positive  
46 attitudes towards big data platforms. They also exhibited sufficient professionalism  
47 before the workshop. Meanwhile, the workshop has significantly promoted students'  
48 knowledge of medical big data ( $P<0.05$ ), and lead to more positive attitudes to big data  
49 platforms and higher levels of professionalism.

50 **Conclusions:** Chinese medical students have primitive acquaintance and positive  
51 attitudes toward big data technologies. They also yield sufficient levels of  
52 professionalism. And the NUWA platform-based workshop was useful in furthering  
53 their understanding of big data and enhancing professionalism.

1  
2  
3  
4 54 **Keywords:** Big data; Chinese medical education; Digital competencies;  
5  
6 55 Professionalism; Workshop.

7  
8 56

9  
10 57 **Strengths and limitations of this study**

11 ● This study depicted the cognition and professionalism of medical big data among  
12 Chinese medical students in the era of big data for the first time.

13  
14  
15 ● This study provided a vivid example of how big data assists medical education,  
16 based on the first nation-wide gynecologic oncology medical big data platform in  
17 China, NUWA platform.  
18  
19

20  
21 ● The limited sample size may restrict the generalization of the conclusion.  
22

23 64  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 65 **Background**

66 Electronification of medical records is the signature of the modern health care system.<sup>1-3</sup>  
67 Massive clinical and omics data were produced to enable more detailed depictions of  
68 patients and diseases. Medical big data is thus reshaping our appreciation of the modern  
69 medical system. Technological advancements in data storage, processing, and analysis  
70 accelerated the clinical application of big data-driven products and contributed to  
71 personalized disease management,<sup>4, 5</sup> early diagnosis,<sup>6-8</sup> and treatment decision.<sup>9, 10</sup>  
72 Especially in the COVID-19 pandemic setting, achieving rapid application for medical  
73 big data would meet the pressing clinical need to predict the progression of diseases by  
74 data characteristics.<sup>11-13</sup>

75 Meanwhile, big data has brought with its new challenges for doctors.<sup>14</sup>  
76 Information overload is posed to every healthcare worker, since they have to acclimate  
77 the nature of big data, including extraordinary value, volume, velocity, variety, and  
78 variability.<sup>15</sup> Challenges can also be expected on medical professionalism in the age of  
79 big data. Because big data valued realistic health-related information more than ever,  
80 the illusion that digital data overweigh face-to-face physician-patient interactions may  
81 motivate doctors to ignore the importance of professionalism.<sup>16, 17</sup> The situation may be  
82 worse in China, because of the explosive development in big data-based technology in  
83 the last years and the well-known heavy workloads of Chinese doctors.<sup>18, 19</sup> So extra  
84 lectures and workshops in obtaining insights into big data and remaining respectful for  
85 patients are necessary.

86 In this study, we carried on a workshop for Chinese clinical and pre-clinical  
87 students to further their knowledge of medical big data and to improve their  
88 professionalism in the big data era. The workshop was based on the National Union of  
89 Real-World Gynecologic Oncology Research & Patient Management Platform  
90 (NUWA platform), which is the first national-wide multi-function gynecologic  
91 oncology medical big data platform in China.

92

93



## 94 **Methods**

### 95 **Study design**

96 We performed a cross-sectional study on Chinese medical students' learning and  
97 application of big data in healthcare. The survey was conducted among students  
98 attending a workshop on medical big data.

### 100 **Workshop design**

101 The workshop was conducted on July 8, 2021 in Tongji Hospital, Tongji medical school,  
102 Huazhong University of Science and Technology, Wuhan, China. It is a part of the  
103 series class of “Medical big data platform learning and applying”.

104 The workshop was delivered by two senior doctors (QLG and SQZ), who have  
105 over 5 years of experience in medical big data platform development and application.  
106 The workshop consisted of three major parts, (i) a lecture on the content and application  
107 of medical big data; (ii) a lecture on professionalism for doctors in the big data era; (iii)  
108 learners are allowed to explore the NUWA platform freely according to their interest.  
109 The first and second sections would last for about 45 minutes, and section three for 30  
110 minutes.

111 In section one, five major parts were included in Dr. Zeng's speech, (i) the  
112 development of medical big data in China and the world, (ii) the application and  
113 potential of big data in clinical practice, medical education, and scientific researches;  
114 (iii) the construction and content of the NUWA platform, (iv) how to use the NUWA  
115 platform; (v) plans for the NUWA platform development. For the second section, Prof.  
116 Gao gave a lecture about (i) the attitude health care workers should have when  
117 communicating with patients and colleagues considering that big data have already  
118 reshaped our medical system, (ii) the importance of detailed and continuous patients'  
119 information for the development of big data platform, (iii) how to protect patient  
120 privacy on the databases and (iv) how to use NUWA platform to satisfy patients' needs.  
121 For section 3, all students would be given a temporary account for the NUWA platform  
122 and have the access to all deidentified patient information. They were allowed to

1  
2  
3  
4 123 explore the medical data freely and view structured health care information for half an  
5  
6 124 hour.

7  
8 125

9 126 **Data collection**

10  
11 127 Clinical and pre-clinical students in Tongji Medical school were invited to attend a 2-  
12  
13 128 hour class by email or roadshows between June 8, 2021 and July 7, 2021. The contents  
14  
15 129 and speakers for the workshop were presented in the email or during the roadshow.  
16  
17 130 Besides, all 5th- (n=50, pre-clinical) and 6th-grade (n=49, clinical) students of 8-year  
18  
19 131 undergraduate education in Tongji Medical School took part in this workshop as an  
20  
21 132 additional course.

22  
23 133 The participants should finish two surveys, one of which should be completed  
24  
25 134 before and the other after the workshop. Both questionnaires contain their basic  
26  
27 135 knowledge and attitude towards big data technology and a Professionalism Mini-  
28  
29 136 Evaluation Exercise (P-MEX) instrument. The attitude towards big data platform was  
30  
31 137 measured by 8 questions, namely, (i) Big data platform could assist future medical  
32  
33 138 education; (ii) Big data platform could assist future medical researches; (iii) Big data  
34  
35 139 platform could assist future clinical practice; (iv) I am willing to learn how to use big  
36  
37 140 data platform; (v) I am willing to use big data platform in the future; (vi) I am willing  
38  
39 141 to recommend big data platform to my colleagues; (vii) Big data platform could benefit  
40  
41 142 my career; (viii) Big data platform could benefit all medical careers. For each question,  
42  
43 143 students should choose from “Totally agree”, “agree”, “disagree”, and “totally  
44  
45 144 disagree”. For the applicability of conclusions, “Totally agree” and “agree” were  
46  
47 145 regarded as “positive attitude”, “disagree” and “totally disagree” were considered as  
48  
49 146 “negative attitude”. The P-MEX consists of 24 items that represent four skill categories,  
50  
51 147 including Doctor-Patient Relationship skills, Reflective skills, Time Management, and  
52  
53 148 Inter-Professional Relationship skills.<sup>20, 21</sup> Besides, the first survey also contained  
54  
55 149 students’ baseline information, and the second one contained a satisfaction  
56  
57 150 questionnaire. Meanwhile, the browsing histories for participants in section 3 were also  
58  
59 151 recorded and analyzed to reflect students’ interest.

1  
2  
3  
4 152 The data collection and analysis plan were acknowledged and agreed by all  
5  
6 153 participants at the beginning of the workshop and the study has been approved by the  
7  
8 154 Research Ethics Commission of Tongji Hospital of Huazhong University of Science  
9  
10 155 and Technology (2020-S201).

### 11 156 12 13 157 **NUWA platform**

14  
15 158 The NUWA platform is the first nationwide Gynecological Oncology data-sharing  
16  
17 159 platform launched by the National Clinical Research Center for Gynecological  
18  
19 160 Oncology in August 2019. This platform integrated inpatient/outpatient clinical data,  
20  
21 161 genomic data, and follow-up data to develop a patient-level longitudinal clinico-  
22  
23 162 genomic database. Information was de-identified and extracted from electronic medical  
24  
25 163 records. A rigorous data quality check was performed to ensure the accuracy of the data  
26  
27 164 entries. Since its foundation in 2019, 17 first-class hospitals from different provinces  
28  
29 165 or cities in China participated in the NUWA platform until August 2021.

### 30 166 31 32 33 167 **Statistics analysis**

34  
35 168 Descriptive statistics were presented by counts and percentages to describe the  
36  
37 169 demographic information. The Chi-square test was used to compare the changes in  
38  
39 170 knowledge and understanding of big data and professionalism before and after the  
40  
41 171 workshop. To acquire more practical results, the categories “not at all important” and  
42  
43 172 “not important” were combined for analysis, as well as “important” and “very  
44  
45 173 important”. *P-values* <0.05 were considered significant. The data were analyzed using  
46  
47 174 R Version 4.03.

### 48 175 49 50 176 **Patient and Public Involvement**

51  
52 177 No patients or members of the public were involved in this study.

### 53 178 54 55 56 179 57 58 180 **Results**

## 181 **Characteristics of students**

182 A total of 274 students participated in this workshop and completed two surveys. All  
183 of them were included in the final analysis. Participants were aged between 22 and 28,  
184 with 148 (54%) pre-clinical students and 126 (46%) clinical ones. Among them, 130  
185 (47.4%) were males and 144 (52.4%) were females (**Table 1**).

186 The majority of participants (207, 75.5%) used to know at least one big data  
187 platform in China or the world. And over two-thirds of them (183, 66.8%) also  
188 acknowledged its application. However, only 17.2% (47) of them have been involved  
189 in any project related to medical big data (**Table 1**).

## 191 **Knowledge for big data platform**

192 Before the workshop, about a quarter of the students (74, 27%) knew the detailed  
193 content of medical big data platforms while near 64% (174) knew the potential  
194 application of it. After attending the lessons, almost all students could understand the  
195 content (253, 92%) and the potential application (253, 92%) of medical big data (**Figure**  
196 **1 and Table S1**).

197 Meanwhile, no matter before or after the workshop, the majority of students  
198 believed that big data technology is practical in medical educations, clinical practice,  
199 and scientific researches (85%, 77%, 82% before the workshop; 99%, 87%, 95% after  
200 the workshop, respectively) (**Figure 1 and Table S1**).

## 202 **Students' attitude towards big data platform**

203 Even before the workshop, most students hold a positive attitude on the potential of the  
204 big data platform in promoting medical educations (237, 86%), medical researches (256,  
205 93%), and clinical practices (247, 90%). And they are also willing to learn and use the  
206 big data platform (250, 91%, and 218, 80% respectively). 74% (202) of them are  
207 enthusiastic to introduce big data platforms to their colleagues. Interestingly, most  
208 students were convinced that big data platform could benefit their careers (248, 91%),  
209 but they were not sure if it would yield the same effect on the others (128, 47%) (**Figure**

210 **2 and Table S2).**

211 When the workshop was finished, almost all students have positive attitudes  
212 towards the big data platform (**Figure 2 and Table S2**). However, there are still 12%  
213 (34) and 18% (48) of students who are not sure if the big data platform could benefit  
214 their and others' medical careers.

215

### 216 **Professionalism**

217 Generally, positive attitudes to all the professionalism items were demonstrated, with  
218 more than three-quarters of the students agree that all professionalism attributes are  
219 "important" or "very important" before the workshop. The three items with most  
220 students thought are "not at all important" or "not important" were maintaining patient  
221 confidentiality (66, 24% of students choose "not important" or "not at all important"),  
222 maintaining appropriate appearance (60, 22% students choose "not important" or "not  
223 at all important") and respecting rules and procedures of the system (55, 20% students  
224 choose "not important" or "not at all important") (**Figure 3A and Table S3**).

225 After the lectures, students exhibited a more positive attitude on most  
226 professionalism items ( $p < 0.05$ ). The most significant improvements happened on  
227 "Maintaining patient confidentiality" (from 76% to 95%), "Listen actively to patient"  
228 (from 89% to 100%), and "Accepting feedback" (from 89% to 98%) (**Figure 3B and**  
229 **Table S3**).

230

### 231 **Interest in medical data**

232 In the free exploration section, most students (253, 92.3%) viewed the patients'  
233 hospitalization logs. 76.3% (209) of them are interested in the history of illness.  
234 Meanwhile, many students were interested in medicine usages (56.2%), surgery reports  
235 (52.9%), and diseases of rare pathologies (58.6%). Only about one-third of students  
236 viewed the image diagnosis (37.6%) and follow-up records (32.5%) (**Table 2**).

237

### 238 **Satisfactory survey**

1  
2  
3  
4 239 When finished, 95.6% (262) of participants were “satisfied” or “extremely satisfied”  
5  
6 240 with this workshop, and only 7 out of 274 students were not at all satisfied. However,  
7  
8 241 we failed to get feedback from them in the next three months. Most students thought  
9  
10 242 the workshop informative (249, 90.9%) and understandable (255, 94.1%). The majority  
11  
12 243 of students were also willing to recommend this seminar to other students (257, 93.8%)  
13  
14 244 and to participant in similar classes in the future (254, 92.7%). As for the duration of  
15  
16 245 this workshop, only 7 (2.6%) of students thought 2 hours was too long (**Table 3**).  
17  
18 246  
19  
20 247

## 21 248 **Discussion**

22  
23  
24 249 Our study demonstrated that Chinese medical students have primitive acquaintance and  
25  
26 250 positive attitudes toward big data technologies. They also yield expertise in  
27  
28 251 professionalism. Meanwhile, workshops based on big data platforms could further  
29  
30 252 strengthen their digital competencies and improve doctor-patient communication  
31  
32 253 capabilities, which would lead to better fitness during the expansion of medical big  
33  
34 254 data.

35  
36 255 In this workshop, the basic knowledge of big data platforms in China was  
37  
38 256 measured. Although a large fraction of students knew about the big data platform and  
39  
40 257 its utility, only a few of them could apprehend its composition or how to use it. This  
41  
42 258 situation may be caused by the rapid development of big data technology in China  
43  
44 259 recently and comparatively lagged relevant educations.<sup>22, 23</sup> Besides, ethical challenges  
45  
46 260 that hinder medical and public health data sharing may also exacerbate the situation.<sup>24</sup>  
47  
48 261 Therefore, aside from conducting big data-related workshops, promoting medical and  
49  
50 262 public health data sharing policy and evolving relevant legal and ethical implications  
51  
52 263 were also of great importance.

53  
54 264 Students’ enthusiasm towards medical big data was also taken into consideration.  
55  
56 265 As expected and consistent with previous reports,<sup>25</sup> the majority of students exhibited  
57  
58 266 positive attitudes towards the big data platform and were willing to acquire further  
59  
60 267 proficiency. They believed that big data would play a vital role in future medical

1  
2  
3  
4 268 education, clinical practice, and scientific researches, which are the main tasks for all  
5  
6 269 Chinese doctors. However, not all of them believed big data could benefit medical  
7  
8 270 careers in any case. The popularity of information technology in China recently may  
9  
10 271 contribute to the big-data-friendly intention in medical students.<sup>26</sup> Meanwhile, their  
11  
12 272 limited understanding of related fields may hamper the belief that it could be used  
13  
14 273 productively.<sup>27-29</sup> In the meantime, we were delighted to see that almost all students  
15  
16 274 recognized the significance of big data after our workshop. We thus believed that proper  
17  
18 275 education may be an efficient way to diminish the misunderstanding and to achieve the  
19  
20 276 full potential in big data platforms.

21  
22 277 Professionalism was related to the development of big data in our study for the  
23  
24 278 first time in China. Under the consideration that big data could reshape medical  
25  
26 279 activities in all aspects including doctors' attitudes towards their colleagues and  
27  
28 280 patients,<sup>30</sup> a lecture about "professionalism in the age of big data" was presented in the  
29  
30 281 current workshop. Before the seminar, students demonstrated sufficient  
31  
32 282 professionalism similar to previous reports, with high scores in almost all items. Several  
33  
34 283 important professionalism levels were significantly increased in many elements after  
35  
36 284 the workshop, indicating that this workshop could be a preliminary attempt for  
37  
38 285 promoting professionalism when the digital data have changed the way we  
39  
40 286 communicate and doctors have to spend more time with electronic records rather than  
41  
42 287 patients. These changes our workshop brought deserve more attention in China since  
43  
44 288 Chinese doctors were well-known to be overburdened.<sup>18,31</sup> These improvements could  
45  
46 289 also help increase the reliability of medical records and produce convincing and  
47  
48 290 effective medical information.

49  
50 291 Another interesting fact in the result is that about 60% of students noticed rare  
51  
52 292 diseases in the free-exploration section, which is hardly involved in routine medical  
53  
54 293 classes.<sup>32,33</sup> We believe this is another strength of big data-based medical education. In  
55  
56 294 traditional medical classes, it is arduous for teachers to grant detailed depictions for  
57  
58 295 every type of rare disease. As a result, students may not have easy access to these  
59  
60 296 exceptional cases, not to mention pay a visit to their medical records thoroughly. The



1  
2  
3  
4 297 big data platform makes it possible for every user to browse cases with rare pathologies  
5  
6 298 conveniently, which would benefit patients as well as reduce misdiagnosis.

7  
8 299 Our workshop has some limitations. The most important one is that the NUWA  
9  
10 300 platform is still under development and the omics data was not included currently.  
11  
12 301 Therefore, another investigation should be performed when the construction of NUWA  
13  
14 302 was completed. Furthermore, narrow geographical distributions of study participants  
15  
16 303 interfere with the generalization of results to other populations. Meanwhile, there are  
17  
18 304 concerns that students who volunteered to participate in the workshop may be more  
19  
20 305 enthusiastic about big data technology than those who did not. We thus included all  
21  
22 306 5th- (pre-clinical) and 6th-grade (clinical) students of 8-year undergraduate education  
23  
24 307 in Tongji Medical School to make the study sample more representative.

25 308

26 309

## 29 310 **Conclusion**

31 311 This study depicted Chinese students' knowledge of medical big data for the first time  
32  
33 312 and the NUWA platform-based workshop was proven useful in promoting their  
34  
35 313 understanding of big data and enhancing professionalism.

36 314

## 37 315 **List of abbreviations**

38  
39 316 P-MEX, Professionalism Mini-Evaluation Exercise instrument; NUWA platform, the  
40  
41 317 National Union of Real-World Gynecologic Oncology Research & Patient  
42  
43 318 Management Platform.

44 319

## 45 320 **Declarations**

### 46 321 *Ethics approval and consent to participate*

47 322 The data collection and analysis plan were acknowledged and agreed by all participants  
48  
49 323 and the study has been approved by the Research Ethics Commission of Tongji Hospital  
50  
51 324 of Huazhong University of Science and Technology (2020-S201).

### 52 325 *Consent for publication*



1  
2  
3  
4 326 Not applicable.

5  
6 327 ***Availability of data and materials***

7 328 The data used in the current study are available from the corresponding author on  
8 329 reasonable request via e-mail (qingleigao@hotmail.com).

9  
10 330 ***Competing interests***

11 331 The authors declare that they have no competing interests.

12  
13  
14 332 ***Funding***

15  
16 333 This work was supported by the Teaching Reform Project for Postgraduates of  
17  
18 334 Huazhong University of Science and Technology (YY202006), and the Major Project  
19  
20 335 of Medical Education (20A0769).

21  
22 336 ***Authors' contributions***

23  
24 337 QLG, JHL and XFJ designed the study and survey. QLG, JHL, XFJ, ML, ZKP, YBH  
25  
26 338 and SQZ conducted the workshop. PJ, HYL, JHC, XYL, YY, and SQZ analyzed and  
27  
28 339 interpreted the data. YJZ and GCM performed interpretation of data and discussion of  
29  
30 340 findings. QLG conceptualized and designed the study, supervised the project, analyzed  
31  
32 341 and interpreted the data, and wrote the paper. All authors approved the final version of  
33  
34 342 the manuscript and agreed to the submission of this manuscript.

35  
36 343 ***Acknowledgements***

37 344 Not applicable.

38  
39  
40 345

41  
42 346 **Reference**

- 43 347 1. Bezemer T, de Groot MC, Blasse E, et al. A Human(e) Factor in Clinical Decision Support Systems.  
44 348 J Med Internet Res. 2019;21(3):e11732-e.  
45 349 2. Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine. Ann Intern Med.  
46 350 2016;164(8):562-3.  
47 351 3. Wu WT, Li YJ, Feng AZ, et al. Data mining in clinical big data: the frequently used databases,  
48 352 steps, and methodological models. Military Medical Research. 2021;8(1):44.  
49 353 4. Olatosi B, Zhang J, Weissman S, et al. Using big data analytics to improve HIV medical care  
50 354 utilisation in South Carolina: A study protocol. #N/A. 2019;9(7):e027688.  
51 355 5. Carra G, Salluh JIF, da Silva Ramos FJ, et al. Data-driven ICU management: Using Big Data and  
52 356 algorithms to improve outcomes. J Crit Care. 2020;60:300-4.  
53 357 6. Kumar NMS, Eswari T, Sampath P, et al. Predictive Methodology for Diabetic Data Analysis in  
54 358 Big Data. In: Vijayakumar V, Neelanarayanan V, editors. Big Data, Cloud and Computing Challenges.  
55 359 Procedia Computer Science. 502015. p. 203-8.

- 360 7. Razavian N, Blecker S, Schmidt AM, et al. Population-Level Prediction of Type 2 Diabetes From  
361 Claims Data and Analysis of Risk Factors. *#N/A*. 2015;3(4):277-87.
- 362 8. Fitzgerald RC. Big data is crucial to the early detection of cancer. *Nat Med*. 2020;26(1):19-20.
- 363 9. Fridley BL, Koestler DC, Godwin AK. Individualizing care for ovarian cancer patients using big  
364 data. *J Natl Cancer Inst*. 2014;106(5).
- 365 10. Cammarota G, Ianiro G, Ahern A, et al. Gut microbiome, big data and machine learning to promote  
366 precision medicine for cancer. *Nat Rev Gastroenterol Hepatol*. 2020;17(10):635-48.
- 367 11. Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19  
368 pneumonia on chest CT using multinational datasets. *Nat Commun*. 2020;11(1):4080.
- 369 12. Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with  
370 COVID-19. *Nat Med*. 2020;26(8):1224-8.
- 371 13. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New  
372 Technology, and Proactive Testing. *Jama*. 2020;323(14):1341-2.
- 373 14. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research  
374 applications and clinical care. *Nature reviews Genetics*. 2012;13(6):395-405.
- 375 15. Andreu-Perez J, Poon CC, Merrifield RD, et al. Big data for health. *IEEE J Biomed Health Inform*.  
376 2015;19(4):1193-208.
- 377 16. O'Connor S, Zhang M, Honey M, et al. Digital professionalism on social media: A narrative review  
378 of the medical, nursing, and allied health education literature. *International journal of medical informatics*.  
379 2021;153:104514.
- 380 17. Attai DJ, Anderson PF, Fisch MJ, et al. Risks and benefits of Twitter use by  
381 hematologists/oncologists in the era of digital medicine. *Seminars in hematology*. 2017;54(4):198-204.
- 382 18. Jingang A. Which future for doctors in China? *Lancet*. 2013;382(9896):936-7.
- 383 19. Wu D, Wang Y, Lam KF, et al. Health system reforms, violence against doctors and job satisfaction  
384 in the medical profession: a cross-sectional survey in Zhejiang Province, Eastern China. *BMJ open*.  
385 2014;4(12):e006431-e.
- 386 20. Cruess R, McIlroy JH, Cruess S, et al. The Professionalism Mini-Evaluation Exercise: A  
387 Preliminary Investigation. *Academic Medicine*. 2006;81(10).
- 388 21. Norcini JJ, Blank LL, Arnold GK, et al. The mini-CEX (clinical evaluation exercise): a preliminary  
389 investigation. *Ann Intern Med*. 1995;123(10):795-9.
- 390 22. Wu J, Wang J, Nicholas S, et al. Application of Big Data Technology for COVID-19 Prevention  
391 and Control in China: Lessons and Recommendations. *J Med Internet Res*. 2020;22(10):e21980.
- 392 23. Shen T, Wang C. Big Data Technology Applications and the Right to Health in China during the  
393 COVID-19 Pandemic. *International journal of environmental research and public health*. 2021;18(14).
- 394 24. Li X, Cong Y. A Systematic Literature Review of Ethical Challenges Related to Medical and Public  
395 Health Data Sharing in China. *Journal of Empirical Research on Human Research Ethics*.  
396 2021:15562646211040299.
- 397 25. Parthiban N, Boland F, Fasil Azim DH, et al. Asian medical students' attitudes towards  
398 professionalism. *Med Educ Online*. 2021;26(1):1927466.
- 399 26. Liang W, Yao J, Chen A, et al. Early triage of critically ill COVID-19 patients using deep learning.  
400 *Nature communications*. 2020;11(1):3543-.
- 401 27. Meo SA, Sattar K, Ullah CH, et al. Progress and prospects of medical education research in Asian  
402 Countries. *Pakistan journal of medical sciences*. 2019;35(6):1475-81.
- 403 28. Schmid RD, Xiong X. Biotech in China 2021, at the beginning of the 14th five-year period ("145").

- 1  
2  
3 404 Applied microbiology and biotechnology. 2021;105(10):3971-85.  
4 405 29. Wu H, Li W, Xie A, et al. Funding of health professional education: China's 20-year process and a  
5 406 global comparison. Medical education. 2021.  
6 407 30. Silva Junior G, Askari M, Oliveira J. Digital Health and Possible Solutions to Improve the Care in  
7 408 the Field of Nephrology. Contributions to nephrology. 2021;199:1-15.  
8 409 31. Liu S, Chen Y, Li S, et al. What Are the Important Factors Influencing the Recruitment and  
9 410 Retention of Doctoral Students in a Public Health Setting? A Discrete Choice Experiment Survey in  
10 411 China. International journal of environmental research and public health. 2021;18(18).  
11 412 32. Walkowiak D, Domaradzki J. Are rare diseases overlooked by medical education? Awareness of  
12 413 rare diseases among physicians in Poland: an explanatory study. Orphanet journal of rare diseases.  
13 414 2021;16(1):400.  
14 415 33. Domaradzki J, Walkowiak D. Knowledge and Attitudes of Future Healthcare Professionals Toward  
15 416 Rare Diseases. Front Genet. 2021;12:639610.

16  
17  
18  
19  
20  
21 417

22  
23 418  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

419 **Tables**420 **Table 1. Baseline characteristics for participants.**

Characteristics	Number of participants
Age, years	
22 (n, %)	27, 9.9%
23 (n, %)	16, 5.8%
24 (n, %)	67, 24.5%
25 (n, %)	66, 24.1%
26 (n, %)	50, 18.2%
27 (n, %)	28, 10.2%
28 (n, %)	20, 7.3%
Gender	
male (n, %)	130, 47.4%
female (n, %)	144, 52.6%
Study stage	
pre-clinical (n, %)	148, 54%
clinical (n, %)	126, 46%
Acknowledgment of any kind of big data platform	
yes (n, %)	207, 75.5%
no (n, %)	67, 24.5%
Know the applications of big data technology	
yes (n, %)	183, 66.8%
no (n, %)	91, 33.2%
Involved in any big data-related projects	
yes (n, %)	47, 17.2%
no (n, %)	227, 82.8%
Total	274, 100%

421 **Table 2. Browser records in the free-exploration section.**

Content	Number of participants
History of Illness	
Yes (n, %)	209, 76.3%
no (n, %)	65, 23.7%
Hospitalization logs	
yes (n, %)	253, 92.3%
no (n, %)	21, 7.7%
Medicine usage	
yes (n, %)	154, 56.2%
no (n, %)	120, 43.8%
Surgery record	
yes (n, %)	145, 52.9%
no (n, %)	129, 47.1%
Image diagnosis	
yes (n, %)	103, 37.6%
no (n, %)	171, 62.4%
Follow-up records	
yes (n, %)	89, 32.5%
no (n, %)	185, 67.5%
Rare pathologies	
yes (n, %)	160, 58.4%
no (n, %)	114, 41.6%
Total (n, %)	274, 100%

422

423

424

425

426

427 **Table 3. Participants 'answers to the workshop satisfaction survey**

	Not at all (n, %)	No (n, %)	Yes (n, %)	Yes, extremely (n, %)
Overall, are you satisfied with this course	7, 2.6%	5, 1.8%	57, 20.8%	205, 74.8%
Did you think the course are informative?	10, 3.6%	15, 5.5%	101, 36.9%	148, 54%
Did you think the duration of this courses is too long	121, 44.2%	118, 43.1%	28, 10.2%	7, 2.6%
Was the course understandable for you	11, 4%	8, 2.9%	121, 44.2%	134, 48.9%
Would you recommend these courses to other students?	6, 2.2%	11, 4%	144, 52.6%	113, 41.2%
Are you willing to take part in similar courses in the future?	6, 2.2%	14, 5.1%	35, 12.8%	219, 79.9%

428

429

430

431

432

433

434

435

436

437

438

1  
2  
3  
4 439 **Figure Legends**

5 440 **Figure 1. Basic knowledge for big data platform (A) before and (B) after the**  
6  
7 441 **workshop.**

8  
9 442 \* means that there is a significant difference before and after the workshop.

10  
11 443 **Figure 2. Students' attitude towards big data platform (A) before and (B) after the**  
12  
13 444 **workshop.**

14  
15 445 \* means that there is a significant difference before and after the workshop.

16  
17 446 **Figure 3. The professionalism for students (A) before and (B) after the workshop.**

18  
19 447 \* means that there is a significant difference before and after the workshop.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

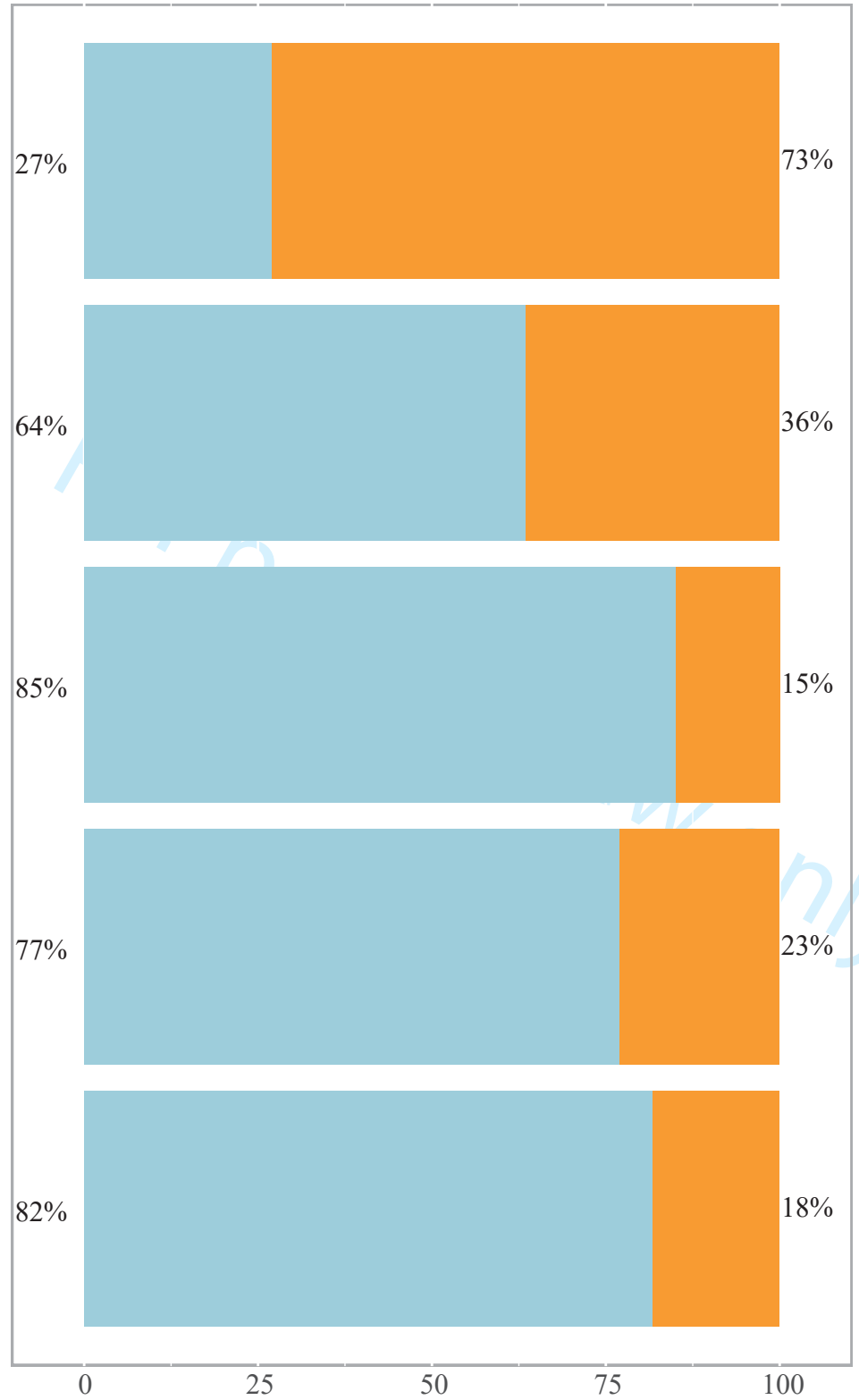
\*I know the content of big data platform

\*I know the potential applications of big data technology

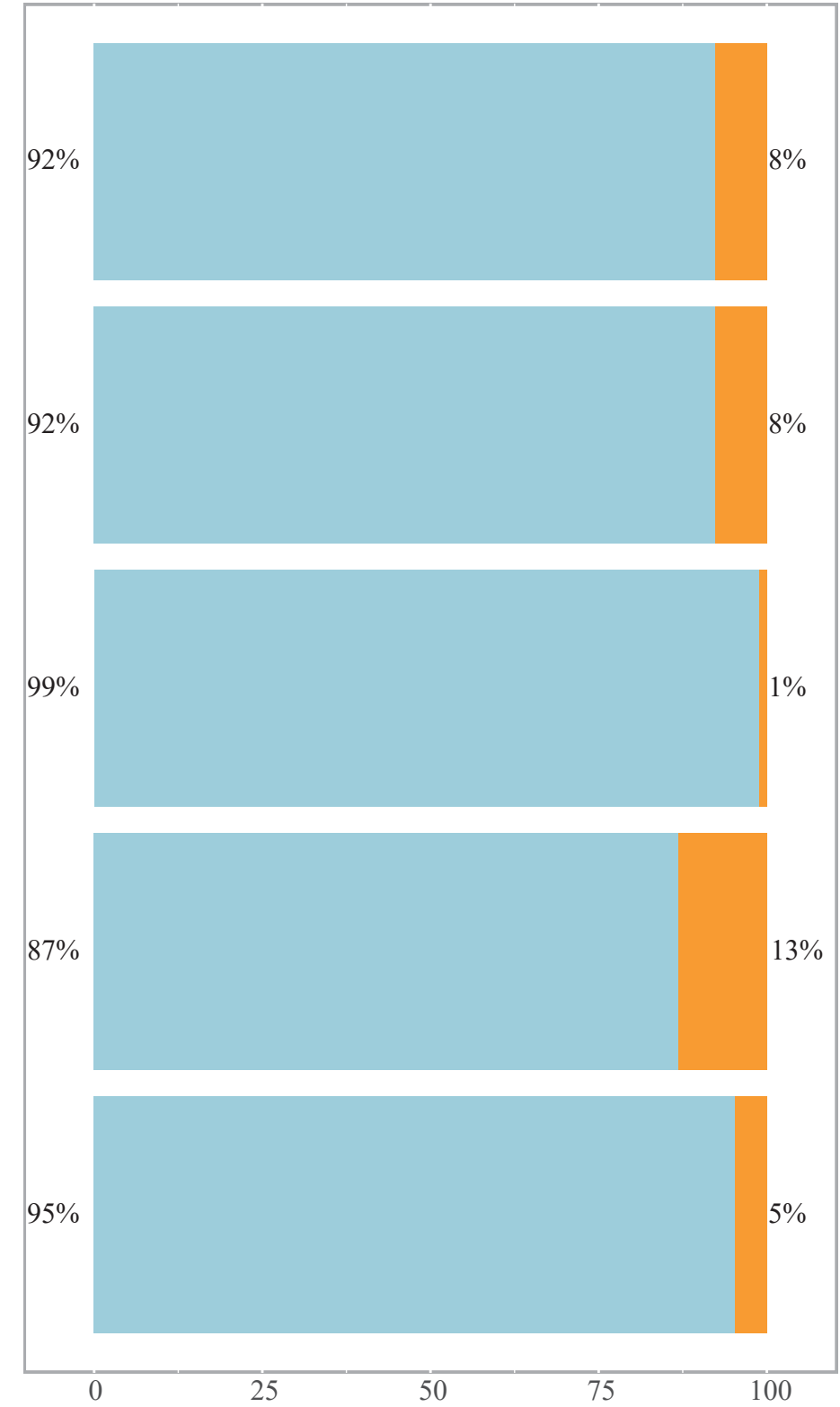
\*Big data technology is practical in medical education

\*Big data technology is practical in clinical practice

\*Big data technology is practical in scientific researches



Percentage

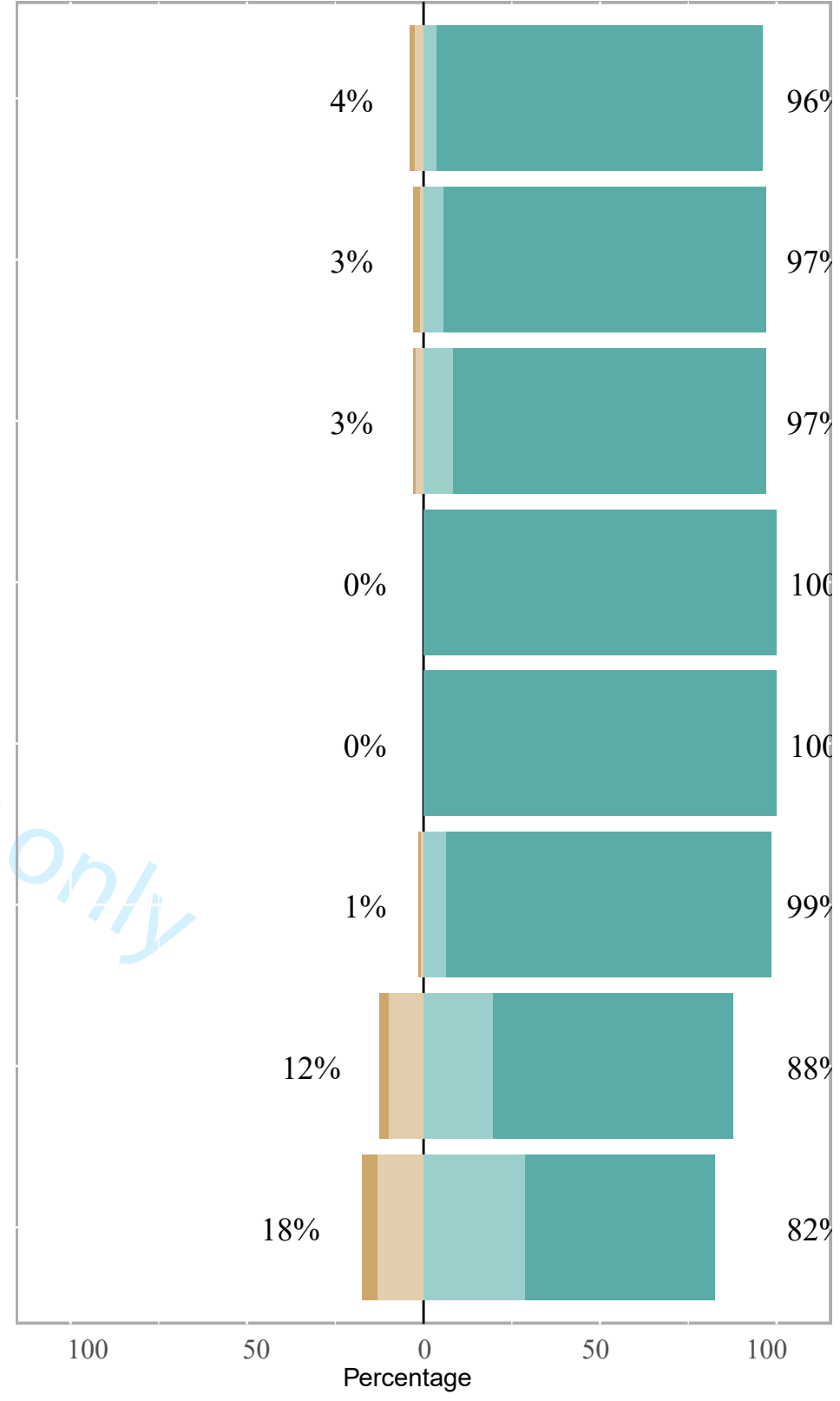
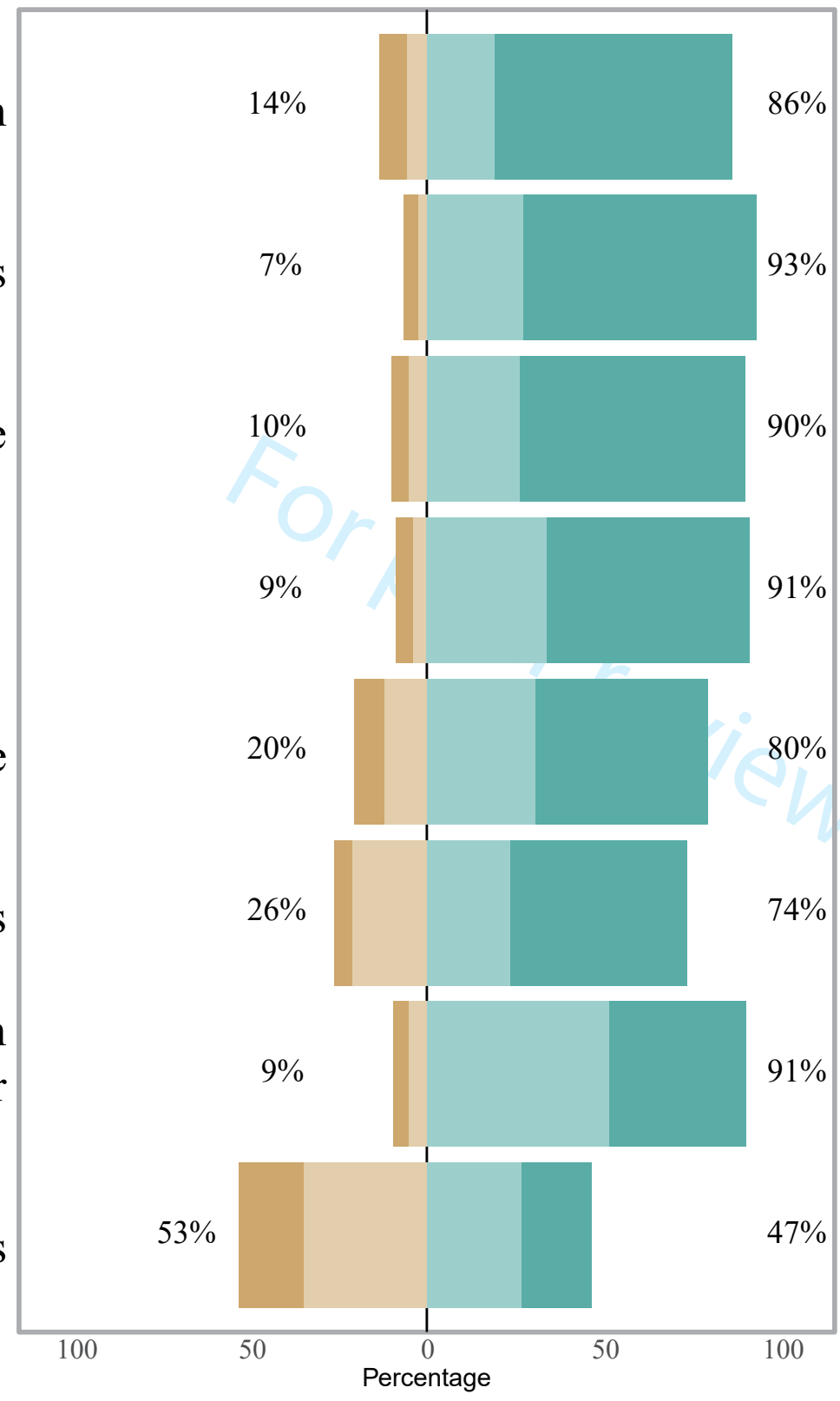


Percentage

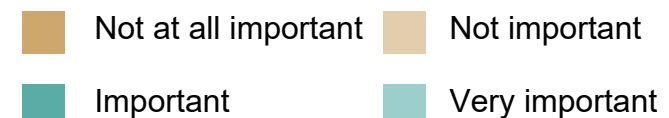
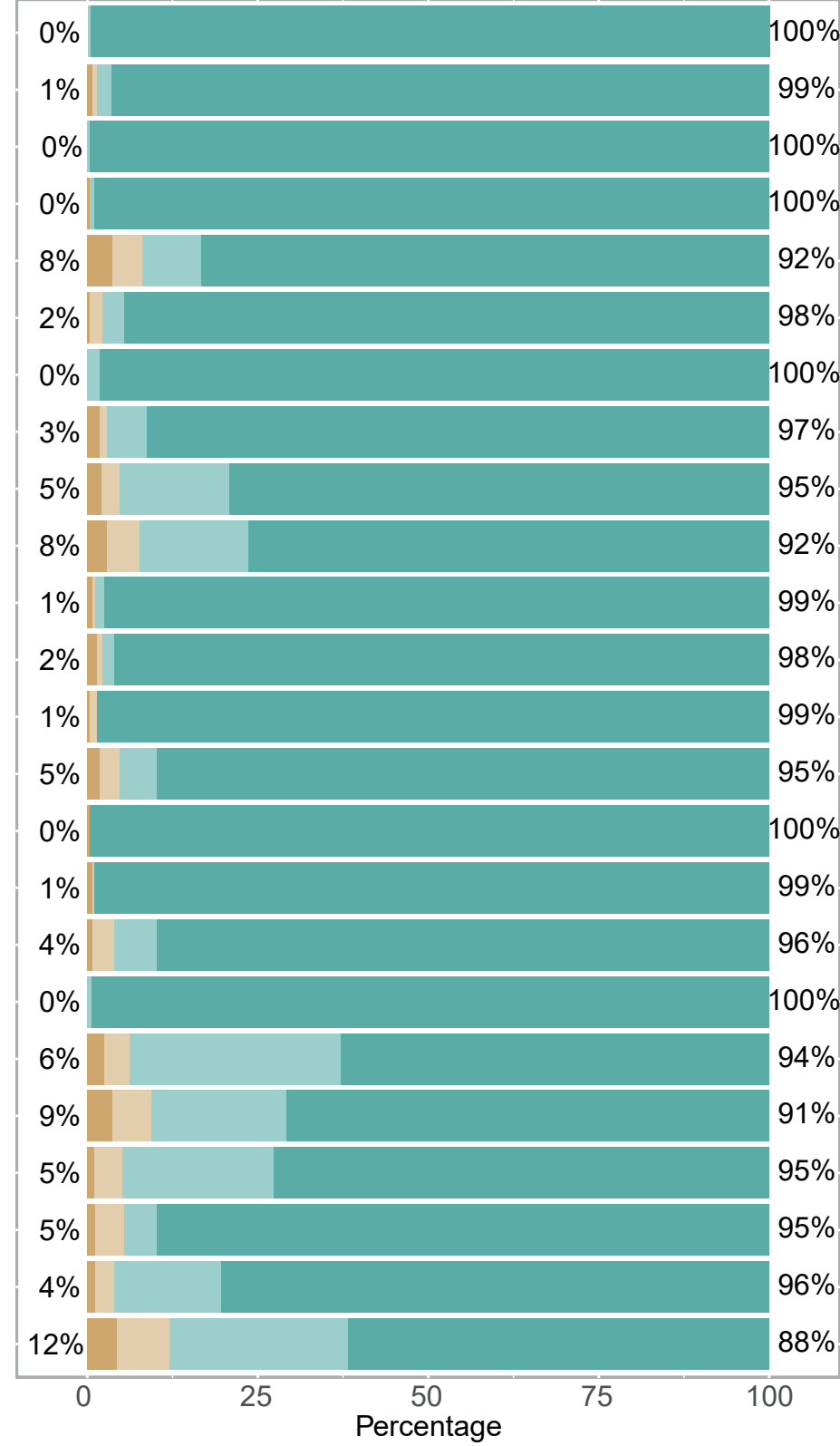
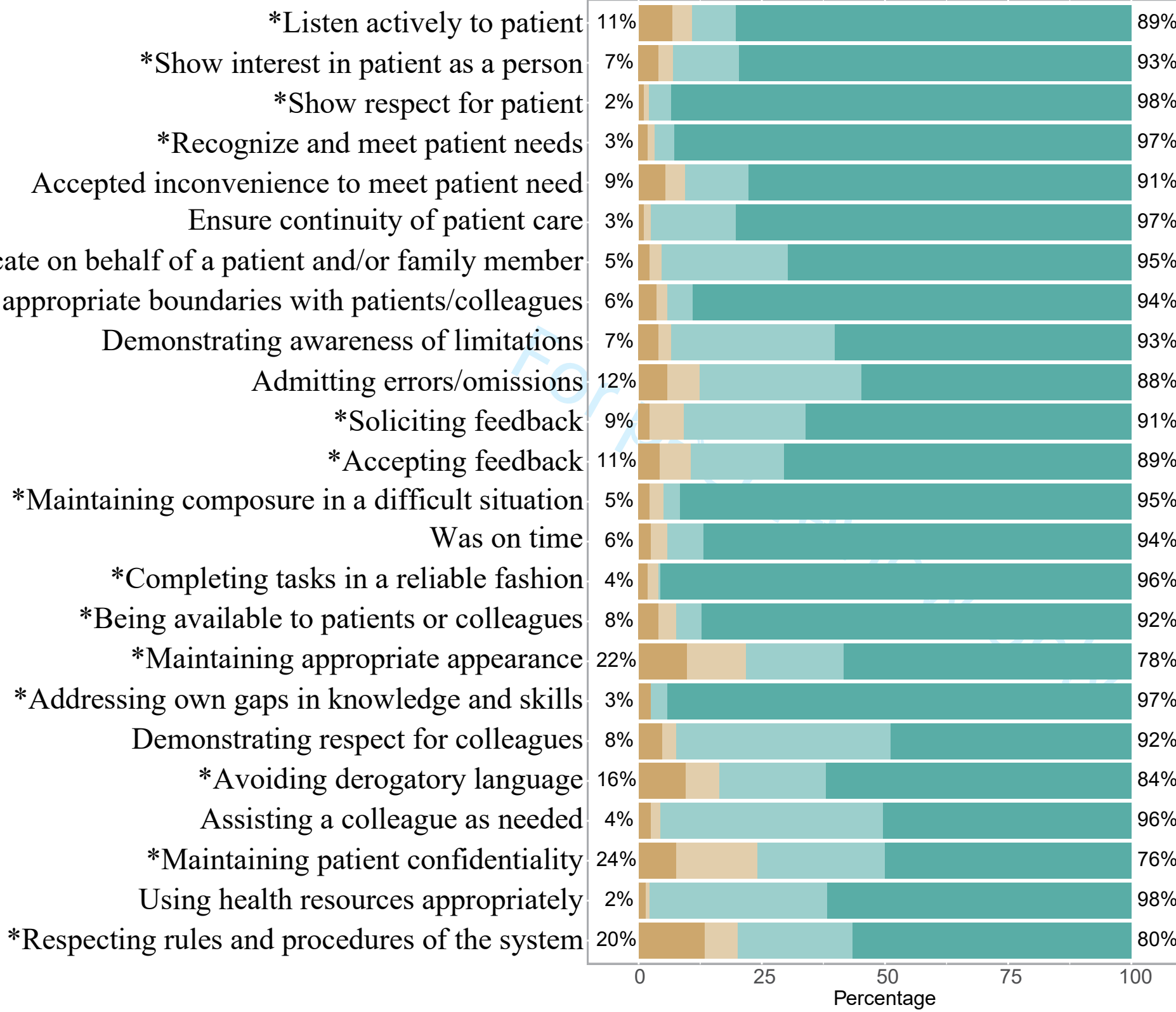
Yes No



- 1 \*Big data platform could assist
- 2 future medical education
- 3
- 4
- 5
- 6 Big data platform could assist
- 7 future medical reserches
- 8
- 9
- 10
- 11
- 12 \*Big data platform could assist
- 13 future clinical practice
- 14
- 15
- 16
- 17
- 18 \*I am willing to learn
- 19 how to use big data platform
- 20
- 21
- 22
- 23
- 24 \*I am willing to use
- 25 big data platform in the future
- 26
- 27
- 28
- 29 \*I am willing to recommanded big
- 30 data platform to my colleges
- 31
- 32
- 33
- 34
- 35
- 36 Big data platform
- 37 could benefit my career
- 38
- 39
- 40
- 41 \*Big data platform could
- 42 benefit all medical careers
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53



## Supplemental tables

Table S1, basic knowledge for big data platform before and after the workshop.

Items	Before Workshop		After Workshop		<i>P</i>
	YES, n	NO, n	YES, n	NO, n	
I know the content of big data platform	74	200	253	21	<0.0001
I know the potential applications of big data technology	174	100	253	21	<0.0001
Big data technology is practical in medical education	233	41	271	3	<0.0001
Big data technology is practical in clinical practice	211	63	238	36	0.0039
Big data technology is practical in scientific researches	224	50	261	13	<0.0001

Table S2, students' attitude towards big data platform before and after the workshop.

Items	Before Workshop				After Workshop				<i>P</i>
	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	
Big data platform could assist future medical education	21	16	53	184	4	7	10	253	0.0002
Big data platform could assist future scientific researches	11	7	75	181	5	3	16	250	0.0705
Big data platform could assist future clinical practice	13	14	72	175	2	6	23	243	0.0017
I am willing to learn how to use big data platform	13	11	93	157	0	0	0	274	<0.0001
I am willing to use big data platform in the future	23	33	84	134	0	0	0	274	<0.0001
I am willing to recommended big data platform to my colleagues	14	58	65	137	2	2	18	252	<0.0001
Big data platform could benefit my career	12	14	142	106	7	27	54	186	0.3382
Big data platform could benefit all medical careers	50	96	74	54	12	36	79	147	<0.0001

Table S3, the professionalism for students before and after the workshop.

Items	Before workshop				After workshop				<i>P</i>
	Not at all important, n	Not important, n	Important, n	Very important, n	Not at all important, n	Not important, n	Important, n	Very important, n	
Doctor-patient relationship skills									
Listen actively to patient	19	11	24	220	0	0	1	273	<0.0001
Show interest in patient as a person	11	8	37	218	2	2	6	264	0.0029
Show respect for patient	3	3	12	256	0	0	1	273	0.0401
Recognize and meet patient needs	5	4	11	254	1	0	2	271	0.0255
Accepted inconvenience to meet patient need	15	11	35	213	10	12	24	228	0.6503
Ensure continuity of patient care	3	4	47	220	1	5	9	259	1
Advocate on behalf of a patient and/or family member	6	7	70	191	0	0	5	269	0.0008
Maintained appropriate boundaries with patients/colleagues	10	6	14	244	5	3	16	250	0.1440
Reflective skills									
Demonstrating awareness of limitations	11	7	91	165	6	7	44	217	0.4595
Admitting errors/omissions	16	18	90	150	8	13	44	209	0.0880
Soliciting feedback	6	19	68	181	2	1	4	267	<0.000

									<i>1</i>
Accepting feedback	12	17	52	193	4	2	5	263	<i>0.0001</i>
Maintaining composure in a difficult situation	6	8	9	251	1	3	0	270	<i>0.0310</i>
Time management									
Was on time	7	9	20	238	5	8	15	246	<i>0.7027</i>
Completing tasks in a reliable fashion	5	6	1	262	1	0	0	273	<i>0.0086</i>
Being available to patients or colleagues	11	10	14	239	2	1	0	271	<i>0.0004</i>
Interprofessional relationship skills									
Maintaining appropriate appearance	27	33	54	160	2	9	17	246	<i>&lt;0.000</i> <i>1</i>
Addressing own gaps in knowledge and skills	7	0	9	258	0	0	2	272	<i>0.0225</i>
Demonstrating respect for colleagues	13	8	119	134	7	10	85	172	<i>0.6139</i>
Avoiding derogatory language	26	19	59	170	10	16	54	194	<i>0.0220</i>
Assisting a colleague as needed	7	5	124	138	3	11	61	199	<i>0.8407</i>
Maintaining patient confidentiality	21	45	71	137	3	12	13	246	<i>&lt;0.000</i> <i>1</i>
Using health resources appropriately	4	2	99	169	3	8	43	220	<i>0.3244</i>
Respecting rules and procedures of the system	37	18	64	155	12	21	72	169	<i>0.0146</i>

**STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies***

Section/Topic	Item #	Recommendation	Reported on page #
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	7
		(b) Describe any methods used to examine subgroups and interactions	NA
		(c) Explain how missing data were addressed	7
		(d) If applicable, describe analytical methods taking account of sampling strategy	7
		(e) Describe any sensitivity analyses	NA
<b>Results</b>			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	All participants were eligible and included in this study
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	Very high response rates with no missing data, as shown in supplementary data.
Outcome data	15*	Report numbers of outcome events or summary measures	8
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	In all tables and figures.
		(b) Report category boundaries when continuous variables were categorized	In all tables and figures.
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
<b>Other information</b>			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

1  
2 \*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.  
3

4 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE  
5 checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at  
6 <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).  
7  
8  
9

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

For peer review only



# BMJ Open

## Oncologic big data platform for promoting digital competencies and professionalism in Chinese medical students: a longitudinal study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-061015.R1
Article Type:	Original research
Date Submitted by the Author:	13-Jul-2022
Complete List of Authors:	Liu, Jiahao; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Jiao, Xiaofei; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Zeng, Shaoqing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Li, Huayi; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Jin, Ping; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Chi, Jianhua; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Liu, Xingyu; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Yu, Yang; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Ma, Guanchen; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Zhao, Yingjun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Li, Ming; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Peng, Zikun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Huo, Yabing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Gao, Qing-Lei ; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Medical education and training
Keywords:	MEDICAL EDUCATION & TRAINING, Gynaecological oncology < GYNAECOLOGY, Gynaecological oncology < ONCOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60





I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 **1 Oncologic big data platform for promoting digital competencies and**  
5 **2 professionalism in Chinese medical students: a longitudinal study**  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

3

4 Jiahao Liu<sup>1,2#</sup>, Xiaofei Jiao<sup>1,2#</sup>, Shaoqing Zeng<sup>1,2</sup>, Huayi Li<sup>1,2</sup>, Ping Jin<sup>1,2</sup>, Jianhua Chi<sup>1,2</sup>,  
5 Xingyu Liu<sup>1,2</sup>, Yang Yu<sup>1,2</sup>, Guanchen Ma<sup>1,2</sup>, Yingjun Zhao<sup>1,2</sup>, Ming Li<sup>1,2</sup>, Zikun Peng<sup>1,2</sup>,  
6 Yabing Huo<sup>1,2</sup>, Qinglei Gao<sup>1,2\*</sup>  
7  
8

7

8

9 <sup>1</sup> Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College,  
10 Huazhong University of Science and Technology, Wuhan 430030, Hubei, P. R. China.

11 <sup>2</sup> Cancer Biology Research Center (Key Laboratory of the Ministry of Education),  
12 Tongji Hospital, Tongji Medical College, Huazhong University of Science and  
13 Technology, Wuhan 430030, Hubei, P. R. China.  
14

14

15 # These authors contribute equally.  
16

16

17 \* Correspondence to:

18 Prof. Qinglei Gao, Cancer Biology Research Center (Key Laboratory of the Ministry  
19 of Education), Tongji Hospital, Tongji Medical College, Huazhong University of  
20 Science and Technology, 1095 Jiefang Ave, Wuhan 430030, Hubei, P. R. China. Tel:  
21 +86-27-83662681; Fax: +86-27-83662681; E-mail: qingleigao@hotmail.com.  
22

22

23 Word count (excluding title page, abstract, references, figures and tables): 3,162  
24

24

## 25 **Abstract**

26 **Objectives:** Advancements in big data technology are reshaping the health care system  
27 in China. This study aims to explore the role of medical big data in promoting digital  
28 competencies and professionalism among Chinese medical students.

29 **Design, setting and participants:** This study was conducted among 274 medical  
30 students who attended a workshop on medical big data conducted on July 8, 2021 in  
31 Tongji Hospital. The workshop was based on the first nationwide multifunction  
32 gynecologic oncology medical big data platform in China, at the National Union of  
33 Real-World Gynecologic Oncology Research & Patient Management Platform  
34 (NUWA platform).

35 **Outcome measures:** Data on knowledge, attitudes toward big data technology and  
36 professionalism were collected before and after the workshop. We have measured the  
37 four skill categories: doctor–patient relationship skills, reflective skills, time  
38 management, and interprofessional relationship skills using the Professionalism  
39 Mini–Evaluation Exercise (P-MEX) as a reflection for professionalism.

40 **Results:** A total of 274 students participated in this workshop and completed all the  
41 surveys. Before the workshop, only 27% of them knew the detailed content of medical  
42 big data platforms, and 64% knew the potential application of medical big data. The  
43 majority of the students believed that big data technology is practical in their clinical  
44 practice (77%), medical education (85%), and scientific research (82%). Over 80% of  
45 the participants showed positive attitudes toward big data platforms. They also  
46 exhibited sufficient professionalism before the workshop. Meanwhile, the workshop  
47 significantly promoted students' knowledge of medical big data ( $P < 0.05$ ), and led to  
48 more positive attitudes toward big data platforms and higher levels of professionalism.

49 **Conclusions:** Chinese medical students have primitive acquaintance and positive  
50 attitudes toward big data technologies. The NUWA platform-based workshop may  
51 potentially promote their understanding of big data and enhance professionalism,  
52 according to the self-measured P-MEX scale.

53 **Keywords:** Big data, Chinese medical education, Digital competencies,

1  
2  
3  
4 54 Professionalism, Workshop

5  
6 55

7  
8 56 **Strengths and limitations of this study**

9  
10 57 ● This study depicted the knowledge and professionalism of medical big data among  
11  
12 58 Chinese medical students in the era of big data for the first time.

13  
14 59 ● This study provided a vivid example of how big data can assist medical education,  
15  
16 60 based on the first nationwide gynecologic oncology medical big data platform in  
17  
18 61 China (NUWA platform).

19  
20 62 ● The generalization of the findings is limited due to the sample representation and  
21  
22 63 longitudinal continuity of this study.

23  
24 64

## 65 **Background**

66       Electronification of medical records is the signature of the modern health care  
67 system.<sup>[1-3]</sup> Massive clinical and omics data have been produced to enable more detailed  
68 depictions of patients and diseases. Medical big data is thus reshaping our appreciation  
69 of the modern medical system. Technological advancements in data storage, processing,  
70 and analysis accelerated the clinical application of big data-driven products and  
71 contributed to personalized disease management,<sup>[4, 5]</sup> early diagnosis,<sup>[6-8]</sup> and treatment  
72 decision.<sup>[9, 10]</sup> Especially in the COVID-19 pandemic setting, achieving rapid  
73 application of medical big data would meet the pressing clinical need to predict the  
74 progression of diseases using data characteristics.<sup>[11-13]</sup>

75       Big data has brought new challenges for doctors.<sup>[14]</sup> Information overload is a  
76 challenge for every healthcare worker, since they have to acclimate to the nature of big  
77 data, including extraordinary value, volume, velocity, variety, and variability.<sup>[15]</sup>  
78 Challenges can also be expected regarding medical professionalism in the age of big  
79 data. Because big data values realistic health-related information more than ever, the  
80 illusion that digital data outweigh face-to-face physician–patient interactions may  
81 motivate doctors to ignore the importance of professionalism.<sup>[16, 17]</sup> There is no precise  
82 definition of medical professionalism, which is reflected in the attitudes and behaviors  
83 directly related to clinical practice. Epstein RM et al. proposed the definition of  
84 professional competence as wisely using communication, knowledge, technical skills,  
85 clinical reasoning, emotions, values, and reflection in daily practice.<sup>[18]</sup> And Cain J et  
86 al. summarized professionalism as a series of attributes, such as altruism, respect,  
87 honesty, and so on.<sup>[19]</sup> Some studies have shown that digital medical education based  
88 on big data plays a positive effect in promoting professionalism.<sup>[16, 20, 21]</sup> However, there  
89 is a lack of research on the attitude and professionalism of Chinese healthcare workers  
90 regarding medical big data. Therefore, extra lectures and workshops about obtaining  
91 insights into big data and remaining respectful to patients are necessary.

92       In this study, we aimed at depicting the knowledge of medical big data and  
93 professionalism in the era of big data among Chinese medical students. What's more,

1  
2  
3  
4 94 we conducted a workshop for Chinese clinical and preclinical students to improve their  
5  
6 95 knowledge of medical big data and their professionalism in the big data era, based on  
7  
8 96 the National Union of Real-World Gynecologic Oncology Research & Patient  
9  
10 97 Management Platform (NUWA platform), which is the first nationwide multifunction  
11  
12 98 gynecologic oncology medical big data platform in China.

13  
14 99

## 100 **Methods**

### 101 **Study design**

102 A workshop on the introduction of medical big data was conducted among clinical  
103 and preclinical students in Tongji Medical School. We conducted this study on Chinese  
104 medical students' learning and the application of big data in healthcare using pre- and  
105 post-course questionnaires.

106

### 107 **Workshop design**

108 The workshop was conducted on July 8, 2021 in Tongji Hospital, Tongji Medical  
109 School, Huazhong University of Science and Technology, Wuhan, China. It was part  
110 of the series class of “Medical big data platform learning and applying”.

111 The workshop was delivered by two senior doctors (QLG and SQZ), who have over  
112 5 years of experience in medical big data platform development and application. The  
113 workshop consisted of three major parts, (i) a lecture on the content and application of  
114 medical big data; (ii) a lecture on professionalism for doctors in the big data era; and  
115 (iii) learners being allowed to explore the NUWA platform freely according to their  
116 interest. The first and second sections lasted for approximately 45 minutes, and the third  
117 section lasted for 30 minutes.

118 In section one, five major parts were included in lectures given by the senior doctors,  
119 (i) the development of medical big data in China and the world, (ii) the application and  
120 potential of big data in clinical practice, medical education, and scientific research, (iii)  
121 the construction and content of the NUWA platform, (iv) how to use the NUWA



1  
2  
3  
4 122 platform, and (v) plans for the NUWA platform development. For the second section,  
5  
6 123 the following four items were discussed: (i) the attitude health care workers should have  
7  
8 124 when communicating with patients and colleagues considering that big data have  
9  
10 125 already reshaped our medical system, (ii) the importance of detailed and continuous  
11  
12 126 patient information for the development of big data platforms, (iii) how to protect  
13  
14 127 patient privacy on the databases, and (iv) how to use the NUWA platform to satisfy  
15  
16 128 patients' needs. For Section 3, all students were given a temporary account for the  
17  
18 129 NUWA platform and have access to all deidentified patient information. They were  
19  
20 130 allowed to explore the medical data freely and view structured health care information  
21  
22 131 for half an hour.

23  
24  
25 132

### 25 133 **Data collection**

26  
27 134 Clinical and preclinical students in Tongji Medical School were invited to attend a  
28  
29 135 2-hour class by email or roadshows between June 8, 2021, and July 7, 2021. The  
30  
31 136 contents and speakers for the workshop were presented in the email or during the  
32  
33 137 roadshow. In addition, all 5th- (n=50, preclinical) and 6th-grade (n=49, clinical)  
34  
35 138 students of 8-year undergraduate education in Tongji Medical School took part in this  
36  
37 139 workshop as an additional course.

38  
39 140 The participants completed two surveys, one before the workshop and the other after  
40  
41 141 the workshop. Both questionnaires contain their basic knowledge and attitude toward  
42  
43 142 big data technology and a Professionalism Mini-Evaluation Exercise (P-MEX)  
44  
45 143 instrument. The P-MEX is developed from the mini-Clinical Examination Exercise  
46  
47 144 (mini-CEX) format by Cruess R et al. in 2006 to evaluate professionalism in clinical  
48  
49 145 training.<sup>[22]</sup> It consists of 24 items representing four skill categories: doctor-patient  
50  
51 146 relationship skills, reflective skills, time management, and interprofessional  
52  
53 147 relationship skills skills.<sup>[23, 24]</sup> And the reliability and validity of P-MEX have been  
54  
55 148 confirmed in both Eastern and Western cultural backgrounds.<sup>[25-27]</sup> The attitude scale is  
56  
57 149 a self-created scale focused on measuring participants' pedagogic evaluation,  
58  
59 150 acceptance and expectation of the big data platform. It consisted of 8 questions: (i) Big  
60

1  
2  
3  
4 151 data platform could assist future medical education, (ii) Big data platform could assist  
5  
6 152 future medical research, (iii) Big data platform could assist future clinical practice, (iv)  
7  
8 153 I am willing to learn how to use big data platform, (v) I am willing to use big data  
9  
10 154 platform in the future, (vi) I am willing to recommend big data platform to my  
11  
12 155 colleagues, (vii) Big data platform could benefit my career, and (viii) Big data platform  
13  
14 156 could benefit all medical careers. For each question, students chose from “totally agree”,  
15  
16 157 “agree”, “disagree”, and “totally disagree”. For the applicability of conclusions, “totally  
17  
18 158 agree” and “agree” were regarded as “positive attitude”, and “disagree” and “totally  
19  
20 159 disagree” were considered as “negative attitude”. In addition, the first survey also  
21  
22 160 contained students’ baseline information, and the second survey contained a  
23  
24 161 satisfaction questionnaire. Meanwhile, the browsing histories for participants in Section  
25  
26 162 3 were also recorded and analyzed to reflect students’ interest.

27 163 The data collection and analysis plan were acknowledged and agreed upon by all  
28  
29 164 participants at the beginning of the workshop, and the study was approved by the  
30  
31 165 Research Ethics Commission of Tongji Hospital of Huazhong University of Science  
32  
33 166 and Technology (2020-S201).

34  
35 167

### 36 37 168 **NUWA platform**

38  
39 169 The NUWA platform is the first nationwide Gynecological Oncology data-sharing  
40  
41 170 platform launched by the National Clinical Research Center for Gynecological  
42  
43 171 Oncology in August 2019. This platform integrated inpatient/outpatient clinical data,  
44  
45 172 genomic data, and follow-up data to develop a patient-level longitudinal clinico-  
46  
47 173 genomic database. Information was deidentified and extracted from electronic medical  
48  
49 174 records. A rigorous data quality check was performed to ensure the accuracy of the data  
50  
51 175 entries. Since its foundation in 2019, 17 first-class hospitals from different provinces  
52  
53 176 or cities in China have participated in the NUWA platform until August 2021.

54  
55 177

### 56 57 178 **Statistical analysis**

58  
59 179 Descriptive statistics were presented by counts and percentages to describe the  
60

180 demographic information. The Chi-square test was used to compare the changes in  
 181 knowledge and understanding of big data and professionalism before and after the  
 182 workshop. To acquire more practical results, the categories “not at all important” and  
 183 “not important” were combined for the analysis, as well as “important” and “very  
 184 important”. *P values* <0.05 were considered statistically significant. The data were  
 185 analyzed using R Version 4.03.

186

### 187 Patient and Public Involvement

188 No patients or members of the public were involved in this study.

189

## 190 Results

### 191 Characteristics of students

192 A total of 274 students participated in this workshop and completed two surveys. All  
 193 of them were included in the final analysis. Participants were aged between 22 and 28,  
 194 with 148 (54.0%) preclinical students and 126 (46.0%) clinical students. Among them,  
 195 130 (47.4%) were males and 144 (52.4%) were females (**Table 1**).

196 The majority of participants (207, 75.5%) knew of at least one big data platform in  
 197 China or the world. Over two-thirds of them (183, 66.8%) also acknowledged its  
 198 application. However, only 17.2% (47) of them had been involved in any project related  
 199 to medical big data (**Table 1**).

200

201 **Table 1. Baseline characteristics for participants.**

Characteristics	Number of participants
Age, years	
22 (n, %)	27, 9.9%
23 (n, %)	16, 5.8%
24 (n, %)	67, 24.5%
25 (n, %)	66, 24.1%
26 (n, %)	50, 18.2%

	27 (n, %)	28, 10.2%
	28 (n, %)	20, 7.3%
Gender		
	male (n, %)	130, 47.4%
	female (n, %)	144, 52.6%
Study stage		
	pre-clinical (n, %)	148, 54.0%
	clinical (n, %)	126, 46.0%
Acknowledgment of any kind of big data platform		
	yes (n, %)	207, 75.5%
	no (n, %)	67, 24.5%
Know the applications of big data technology		
	yes (n, %)	183, 66.8%
	no (n, %)	91, 33.2%
Involved in any big data-related projects		
	yes (n, %)	47, 17.2%
	no (n, %)	227, 82.8%
Total		274, 100%

202

### 203 Knowledge of big data platforms

204 Before the workshop, approximately a quarter of the students (74, 27%) knew  
 205 detailed content of at least one medical big data platform, while nearly 64% (174) knew  
 206 the potential application of medical big data platforms. After attending the lessons,  
 207 almost all students could understand the content (253, 92%) and the potential  
 208 application (253, 92%) of medical big data (**Figure 1 and Table S1**).

209 Regardless of the survey completed (pre- or post-course), the majority of students  
 210 believed that big data technology is practical in medical education, clinical practice,

1  
2  
3  
4 211 and scientific research (85%, 77%, 82% before the workshop; 99%, 87%, 95% after the  
5  
6 212 workshop, respectively) (**Figure 1 and Table S1**).

7  
8 213

#### 9 214 **Students' attitudes toward big data platforms**

10  
11 215 Even before the workshop, most students hold a positive attitude on the potential of  
12  
13 216 the big data platform in promoting medical education (237, 86%), medical research  
14  
15 217 (256, 93%), and clinical practices (247, 90%). They were also willing to learn about  
16  
17 218 and use the big data platform (250, 91%, and 218, 80% respectively). Of them, 74%  
18  
19 219 (202) were enthusiastic to introduce big data platforms to their colleagues. Interestingly,  
20  
21 220 most students were convinced that the big data platform could benefit their careers (248,  
22  
23 221 91%), but they were not sure if it would yield the same effect on the others (128, 47%)  
24  
25 222 (**Figure 2 and Table S2**).

26  
27 223 When the workshop was completed, almost all students had positive attitudes toward  
28  
29 224 the big data platform (**Figure 2 and Table S2**). However, there were still 12% (34) and  
30  
31 225 18% (48) of students who were not sure if the big data platform could benefit their and  
32  
33 226 others' medical careers, respectively.

34  
35 227

#### 36 228 **Professionalism**

37  
38 229 Generally, positive attitudes toward all the professionalism items were demonstrated,  
39  
40 230 with more than three-quarters of the students agreeing that all professionalism attributes  
41  
42 231 were "important" or "very important" before the workshop. The three items that most  
43  
44 232 students thought were "not at all important" or "not important" were maintaining  
45  
46 233 patient confidentiality (66, 24% of students chose "not important" or "not at all  
47  
48 234 important"), maintaining appropriate appearance (60, 22% students chose "not  
49  
50 235 important" or "not at all important") and respecting rules and procedures of the system  
51  
52 236 (55, 20% students chose "not important" or "not at all important") (**Figure 3A and**  
53  
54 237 **Table S3**).

55  
56 238 After the lectures, students exhibited a more positive attitude on most  
57  
58 239 professionalism items ( $p < 0.05$ ). The most significant improvements occurred for  
59  
60

240 “Maintaining patient confidentiality” (from 76% to 95%), “Listen actively to patient”  
 241 (from 89% to 100%), and “Accepting feedback” (from 89% to 98%) (**Figure 3B and**  
 242 **Table S3**).

243

#### 244 **Interest in medical data**

245 In the free exploration section, most students (253, 92.3%) viewed the patients’  
 246 hospitalization logs. A total of 76.3% (209) of students were interested in the patients’  
 247 history of illness. Meanwhile, many students were interested in medicine usages  
 248 (56.2%), surgery reports (52.9%), and diseases of rare pathologies (58.6%). Only  
 249 approximately one-third of students viewed image diagnosis (37.6%) and follow-up  
 250 records (32.5%) (**Table 2**).

251

252 **Table 2. Browser records in the free-exploration section.**

Content	Number of participants
History of Illness	
Yes (n, %)	209, 76.3%
no (n, %)	65, 23.7%
Hospitalization logs	
yes (n, %)	253, 92.3%
no (n, %)	21, 7.7%
Medicine usage	
yes (n, %)	154, 56.2%
no (n, %)	120, 43.8%
Surgery record	
yes (n, %)	145, 52.9%
no (n, %)	129, 47.1%
Image diagnosis	
yes (n, %)	103, 37.6%
no (n, %)	171, 62.4%

Follow-up records	
yes (n, %)	89, 32.5%
no (n, %)	185, 67.5%
Rare pathologies	
yes (n, %)	160, 58.4%
no (n, %)	114, 41.6%
Total (n, %)	274, 100%

253

254 **Satisfactory survey**

255 Upon completing the survey, 95.6% (262) of participants were “satisfied” or  
 256 “extremely satisfied” with this workshop, and only 7 out of 274 students were not at all  
 257 satisfied. However, we failed to obtain feedback from them in the next three months.  
 258 Most students thought that the workshop was informative (249, 90.9%) and  
 259 understandable (255, 94.1%). The majority of students were also willing to recommend  
 260 this seminar to other students (257, 93.8%) and to participate in similar classes in the  
 261 future (254, 92.7%). Regarding the duration of this workshop, only 7 (2.6%) students  
 262 thought that 2 hours was too long (**Table 3**).

263

264 **Table 3. Participants ‘answers to the workshop satisfaction survey**

	Not at all (n, %)	No (n, %)	Yes (n, %)	Yes, extremely (n, %)	Mean	SD
Overall, are you satisfied with this course	7, 2.6%	5, 1.8%	57, 20.8%	205, 74.8%	3.68	0.640
Did you think the course are informative?	10, 3.6%	15, 5.5%	101, 36.9%	148, 54%	3.41	0.757
Did you think the duration of this courses is too long	121, 44.2%	118, 43.1%	28, 10.2%	7, 2.6%	2.07	0.954
Was the course	11, 4%	8, 2.9%	121, 44.2%	134, 48.9%	3.38	0.733

understandable for you								
Would you recommend these courses to other students?	6, 2.2%	11, 4%	144, 52.6%	113, 41.2%	3.33	0.659		
Are you willing to take part in similar courses in the future?	6, 2.2%	14, 5.1%	35, 12.8%	219, 79.9%	3.70	0.666		

265 SD, Standard Deviation.

266

## 267 Discussion

268 Our study demonstrated that Chinese medical students have little knowledge of and  
 269 positive attitudes toward big data technologies. They also yield expertise in  
 270 professionalism. Furthermore, workshops based on big data platforms could further  
 271 strengthen their digital competencies and improve doctor–patient communication  
 272 capabilities, which would lead to better fitness during the expansion of medical big  
 273 data.

274 In this workshop, the basic knowledge of big data platforms in China was measured.  
 275 Although a large fraction of students knew about the big data platform and its utility,  
 276 only a few of them could apprehend its composition or how to use it. This situation may  
 277 be caused by the recent rapid development of big data technology in China and  
 278 comparatively lagging relevant education.<sup>[28, 29]</sup> In addition, ethical challenges that  
 279 hinder medical and public health data sharing may have also exacerbated the  
 280 situation.<sup>[30]</sup> Therefore, aside from conducting big data-related workshops, promoting  
 281 medical and public health data sharing policies and evolving relevant legal and ethical  
 282 implications are also of great importance.

283 Students' enthusiasm toward medical big data was also taken into consideration. As  
 284 expected and consistent with previous reports,<sup>[25]</sup> the majority of students exhibited  
 285 positive attitudes toward the big data platform and were willing to acquire further  
 286 proficiency. They believed that big data would play a vital role in future medical



1  
2  
3  
4 287 education, clinical practice, and scientific research, which are the main tasks for all  
5  
6 288 Chinese doctors. However, not all of them believed that big data could benefit all  
7  
8 289 medical careers. The popularity of information technology of China recently may  
9  
10 290 contribute to the big data-friendly intention in medical students.<sup>[31]</sup> Meanwhile, their  
11  
12 291 limited understanding of related fields may hamper the belief that it could be used  
13  
14 292 productively.<sup>[32-34]</sup> In the meantime, we were delighted to see that almost all students  
15  
16 293 recognized the significance of big data after our workshop. We thus believe that proper  
17  
18 294 education may be an efficient way to diminish misunderstandings and to achieve full  
19  
20 295 potential in big data platforms.

21 296 Professionalism was related to the development of big data in our study for the first  
22  
23 297 time in China. Considering that big data can reshape medical activities in all aspects,  
24  
25 298 including doctors' attitudes toward their colleagues and patients,<sup>[35]</sup> a lecture about  
26  
27 299 "professionalism in the age of big data" was presented in the current workshop. Before  
28  
29 300 the seminar, students demonstrated sufficient professionalism similar to the findings of  
30  
31 301 previous reports, with high scores on the majority of these items. Nevertheless, it is  
32  
33 302 worth noting that nearly 1/4 of the participants did not pay enough attention to  
34  
35 303 maintaining patient confidentiality. Protecting patients' privacy is an essential  
36  
37 304 embodiment of medical ethics and humanities.<sup>[36]</sup> Participants mostly answered these  
38  
39 305 questions from the perspective of big data users before our workshop, ignoring that  
40  
41 306 those data represented thousands of actual patients. After we emphasized the  
42  
43 307 importance of data privacy in the workshop, students realized that respect for patients  
44  
45 308 is the foundation for improving their medical professionalism. Meanwhile, to enhance  
46  
47 309 patient-data privacy protection, the privacy information, such as ID number, was  
48  
49 310 privatized and converted into an alternative ID number using a hashing algorithm  
50  
51 311 before being uploaded to the NUWA platform. Therefore, personal information of all  
52  
53 312 included patients in NUWA platform is confidential and unreachable for all users.  
54  
55 313 Several important professionalism levels were also significantly increased in many  
56  
57 314 other elements after the workshop, indicating that this workshop could be a preliminary  
58  
59 315 attempt to promote professionalism when the prominence of digital data has changed  
60

1  
2  
3  
4 316 the way we communicate and when doctors have to spend more time with electronic  
5  
6 317 records than with patients. These changes our workshop brought deserve more attention  
7  
8 318 in China since Chinese doctors are well-known to be overburdened.<sup>[37, 38]</sup> These  
9  
10 319 improvements could also help increase the reliability of medical records and produce  
11  
12 320 convincing and effective medical information.

13 321 Another interesting fact in the results is that approximately 60% of students noticed  
14  
15 322 rare diseases in the free-exploration section, which is hardly involved in routine medical  
16  
17 323 classes.<sup>[39, 40]</sup> We believe this is another strength of big data-based medical education.  
18  
19 324 In traditional medical classes, it is arduous for teachers to grant detailed depictions for  
20  
21 325 every type of rare disease. As a result, students may not have easy access to these  
22  
23 326 exceptional cases and therefore are often not able to diagnose rare diseases when  
24  
25 327 reviewing patients' medical records. The big data platform makes it possible for every  
26  
27 328 user to conveniently browse through cases with rare pathologies, which would benefit  
28  
29 329 patients as well as reduce misdiagnoses.

30  
31 330 Our workshop had some limitations. The most important one is that the NUWA  
32  
33 331 platform is still under development, and the omics data are not currently included.  
34  
35 332 Therefore, another investigation should be conducted when the construction of NUWA  
36  
37 333 has been completed. Furthermore, narrow geographical distributions of study  
38  
39 334 participants interfered with the generalization of results to other populations.  
40  
41 335 Meanwhile, there are concerns that students who volunteered to participate in the  
42  
43 336 workshop may be more enthusiastic about big data technology than those who did not.  
44  
45 337 We thus included all 5th- (preclinical) and 6th-grade (clinical) students of 8-year  
46  
47 338 undergraduate education in Tongji Medical School to make the study sample more  
48  
49 339 representative. As for the measurement of attitudes and professionalism, a qualitative  
50  
51 340 approach would have been more appropriate and should be considered in the future.  
52  
53 341 Meanwhile, this workshop is too short to produce fundamental improvements in  
54  
55 342 students' attitudes and professionalism. In the future, more long-term studies are wanted  
56  
57 343 to draw a clear conclusion.

58 344  
59  
60

## 345 **Conclusion**

346 This study depicted Chinese students' knowledge of medical big data for the first  
347 time and the NUWA platform-based workshop had potential to improve their  
348 understanding of big data and enhance professionalism.

## 350 **List of abbreviations**

351 P-MEX, Professionalism Mini-Evaluation Exercise instrument. NUWA platform, the  
352 National Union of Real-World Gynecologic Oncology Research & Patient  
353 Management Platform.

## 355 **Contributorship**

356 QLG, JHL and XFJ designed the study and survey. QLG, JHL, XFJ, ML, ZKP, YBH  
357 and SQZ conducted the workshop. PJ, HYL, JHC, XYL, YY, and SQZ analyzed and  
358 interpreted the data. YJZ and GCM performed interpretation of data and discussion of  
359 findings. QLG conceptualized and designed the study, supervised the project, analyzed  
360 and interpreted the data, and wrote the paper. All authors approved the final version of  
361 the manuscript and agreed to the submission of this manuscript.

## 363 **Funding statement**

364 This work was supported by the Teaching Reform Project for Postgraduates of  
365 Huazhong University of Science and Technology (YY202006), the Major Project of  
366 Medical Education (20A0769), and Teaching Research Project of The Second Clinical  
367 School, Tongji Medical College, Huazhong University of Science and Technology  
368 (2021022).

## 370 **Competing of interests**

371 The authors declare that they have no competing interests.

## 373 **Ethics approval**

1  
2  
3  
4 374 The data collection and analysis plan were acknowledged and agreed by all participants  
5  
6 375 and the study has been approved by the Research Ethics Commission of Tongji Hospital  
7  
8 376 of Huazhong University of Science and Technology (2020-S201).  
9

10 377

### 11 378 **Data sharing**

12  
13 379 The data used in the current study are available from the corresponding author on  
14  
15 380 reasonable request via e-mail ([qingleigao@hotmail.com](mailto:qingleigao@hotmail.com)).  
16

17 381

### 18 382 **Consent for publication**

19 383 Not applicable.  
20

21 384

### 22 385 **Acknowledgements**

23 386 Not applicable.  
24

25 387

### 26 388 **Reference**

- 27 389 [1] Bezemer T, de Groot MC, Blasse E, et al. A Human(e) Factor in Clinical Decision Support  
28 390 Systems [J]. *J Med Internet Res*. 2019;21(3):e11732-e.  
29 391 [2] Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine [J]. *Ann Intern Med*.  
30 392 2016;164(8):562-3.  
31 393 [3] Wu WT, Li YJ, Feng AZ, et al. Data mining in clinical big data: the frequently used databases,  
32 394 steps, and methodological models [J]. *Military Medical Research*. 2021;8(1):44.  
33 395 [4] Olatosi B, Zhang J, Weissman S, et al. Using big data analytics to improve HIV medical care  
34 396 utilisation in South Carolina: A study protocol [J]. #N/A. 2019;9(7):e027688.  
35 397 [5] Carra G, Salluh JIF, da Silva Ramos FJ, et al. Data-driven ICU management: Using Big Data  
36 398 and algorithms to improve outcomes [J]. *J Crit Care*. 2020;60:300-4.  
37 399 [6] Kumar NMS, Eswari T, Sampath P, et al. Predictive Methodology for Diabetic Data Analysis  
38 400 in Big Data In: Vijayakumar V, Neelanarayanan V, editors. *Big Data, Cloud and Computing*  
39 401 *Challenges* [M]. *Procedia Computer Science*. 502015. p. 203-8.  
40 402 [7] Razavian N, Blecker S, Schmidt AM, et al. Population-Level Prediction of Type 2 Diabetes  
41 403 From Claims Data and Analysis of Risk Factors [J]. #N/A. 2015;3(4):277-87.  
42 404 [8] Fitzgerald RC. Big data is crucial to the early detection of cancer [J]. *Nat Med*. 2020;26(1):19-  
43 405 20.  
44 406 [9] Fridley BL, Koestler DC, Godwin AK. Individualizing care for ovarian cancer patients using  
45 407 big data [J]. *J Natl Cancer Inst*. 2014;106(5).  
46 408 [10] Cammarota G, Ianiro G, Ahern A, et al. Gut microbiome, big data and machine learning to  
47 409 promote precision medicine for cancer [J]. *Nat Rev Gastroenterol Hepatol*. 2020;17(10):635-  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 410 48.  
4 411 [11] Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19  
5 412 pneumonia on chest CT using multinational datasets [J]. *Nat Commun.* 2020;11(1):4080.  
6 413 [12] Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with  
7 414 COVID-19 [J]. *Nat Med.* 2020;26(8):1224-8.  
8 415 [13] Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New  
9 416 Technology, and Proactive Testing [J]. *Jama.* 2020;323(14):1341-2.  
10 417 [14] Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research  
11 418 applications and clinical care [J]. *Nature reviews Genetics.* 2012;13(6):395-405.  
12 419 [15] Andreu-Perez J, Poon CC, Merrifield RD, et al. Big data for health [J]. *IEEE J Biomed Health*  
13 420 *Inform.* 2015;19(4):1193-208.  
14 421 [16] O'Connor S, Zhang M, Honey M, et al. Digital professionalism on social media: A narrative  
15 422 review of the medical, nursing, and allied health education literature [J]. *Int J Med Inf.*  
16 423 2021;153:104514.  
17 424 [17] Attai DJ, Anderson PF, Fisch MJ, et al. Risks and benefits of Twitter use by  
18 425 hematologists/oncologists in the era of digital medicine [J]. *Seminars in hematology.*  
19 426 2017;54(4):198-204.  
20 427 [18] Epstein RM, Hundert EM. Defining and assessing professional competence [J]. *Jama.*  
21 428 2002;287(2):226-35.  
22 429 [19] Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age [J]. *Currents in*  
23 430 *Pharmacy Teaching and Learning.* 2009;1(2):66-70.  
24 431 [20] Hao X, Peng X, Ding X, et al. Application of digital education in undergraduate nursing and  
25 432 medical interns during the COVID-19 pandemic: A systematic review [J]. *Nurse Educ Today.*  
26 433 2022;108:105183.  
27 434 [21] Kailey BS, Dev D, Sado DM, et al. CardioWebinar: the evolution of digital education during  
28 435 the COVID-19 pandemic [J]. *Heart.* 2021;107(24):2004-5.  
29 436 [22] Cruess R, McIlroy JH, Cruess S, et al. The Professionalism Mini-Evaluation Exercise: A  
30 437 Preliminary Investigation [J]. 2006;81(10):S74-S8.  
31 438 [23] Cruess R, McIlroy JH, Cruess S, et al. The Professionalism Mini-Evaluation Exercise: A  
32 439 Preliminary Investigation [J]. *Academic Medicine.* 2006;81(10).  
33 440 [24] Norcini JJ, Blank LL, Arnold GK, et al. The mini-CEX (clinical evaluation exercise): a  
34 441 preliminary investigation [J]. *Ann Intern Med.* 1995;123(10):795-9.  
35 442 [25] Parthiban N, Boland F, Fadil Azim DH, et al. Asian medical students' attitudes towards  
36 443 professionalism [J]. *Med Educ Online.* 2021;26(1):1927466.  
37 444 [26] Tsugawa Y, Ohbu S, Cruess R, et al. Introducing the Professionalism Mini-Evaluation Exercise  
38 445 (P-MEX) in Japan: results from a multicenter, cross-sectional study [J]. *Acad Med.*  
39 446 2011;86(8):1026-31.  
40 447 [27] Bajwa NM, Nendaz MR, Posfay-Barbe KM, et al. A Meaningful and Actionable  
41 448 Professionalism Assessment: Validity Evidence for the Professionalism Mini-Evaluation  
42 449 Exercise (P-MEX) Across 8 Years [J]. *Acad Med.* 2021;96(11s):S151-s7.  
43 450 [28] Wu J, Wang J, Nicholas S, et al. Application of Big Data Technology for COVID-19 Prevention  
44 451 and Control in China: Lessons and Recommendations [J]. *J Med Internet Res.*  
45 452 2020;22(10):e21980.  
46 453 [29] Shen T, Wang C. Big Data Technology Applications and the Right to Health in China during  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 454 the COVID-19 Pandemic [J]. *International journal of environmental research and public health*.  
4 455 2021;18(14).  
5  
6 456 [30] Li X, Cong Y. A Systematic Literature Review of Ethical Challenges Related to Medical and  
7 457 Public Health Data Sharing in China [J]. *Journal of Empirical Research on Human Research*  
8 458 *Ethics*. 2021:15562646211040299.  
9  
10 459 [31] Liang W, Yao J, Chen A, et al. Early triage of critically ill COVID-19 patients using deep  
11 460 learning [J]. *Nature communications*. 2020;11(1):3543-.  
12 461 [32] Meo SA, Sattar K, Ullah CH, et al. Progress and prospects of medical education research in  
13 462 Asian Countries [J]. *Pakistan journal of medical sciences*. 2019;35(6):1475-81.  
14  
15 463 [33] Schmid RD, Xiong X. Biotech in China 2021, at the beginning of the 14th five-year period  
16 464 ("145") [J]. *Applied microbiology and biotechnology*. 2021;105(10):3971-85.  
17  
18 465 [34] Wu H, Li W, Xie A, et al. Funding of health professional education: China's 20-year process  
19 466 and a global comparison [J]. *Medical education*. 2021.  
20 467 [35] Silva Junior G, Askari M, Oliveira J. Digital Health and Possible Solutions to Improve the Care  
21 468 in the Field of Nephrology [J]. *Contributions to nephrology*. 2021;199:1-15.  
22  
23 469 [36] Doukas DJ, McCullough LB, Wear S. Perspective: Medical education in medical ethics and  
24 470 humanities as the foundation for developing medical professionalism [J]. *Acad Med*.  
25 471 2012;87(3):334-41.  
26  
27 472 [37] Liu S, Chen Y, Li S, et al. What Are the Important Factors Influencing the Recruitment and  
28 473 Retention of Doctoral Students in a Public Health Setting? A Discrete Choice Experiment  
29 474 Survey in China [J]. *International journal of environmental research and public health*.  
30 475 2021;18(18).  
31  
32 476 [38] Jingang A. Which future for doctors in China? [J]. *Lancet*. 2013;382(9896):936-7.  
33 477 [39] Walkowiak D, Domaradzki J. Are rare diseases overlooked by medical education? Awareness  
34 478 of rare diseases among physicians in Poland: an explanatory study [J]. *Orphanet journal of rare*  
35 479 *diseases*. 2021;16(1):400.  
36  
37 480 [40] Domaradzki J, Walkowiak D. Knowledge and Attitudes of Future Healthcare Professionals  
38 481 Toward Rare Diseases [J]. *Front Genet*. 2021;12:639610.  
39  
40 482  
41  
42 483

1  
2  
3  
4 484 **Figure Legends**

5 485 **Figure 1. Basic knowledge for big data platform (A) before and (B) after the**  
6  
7 486 **workshop.**

8  
9 487 \* means that there is a significant difference before and after the workshop.

10  
11 488 **Figure 2. Students' attitude towards big data platform (A) before and (B) after the**  
12  
13 489 **workshop.**

14  
15 490 \* means that there is a significant difference before and after the workshop.

16  
17 491 **Figure 3. The professionalism for students (A) before and (B) after the workshop.**

18  
19 492 \* means that there is a significant difference before and after the workshop.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

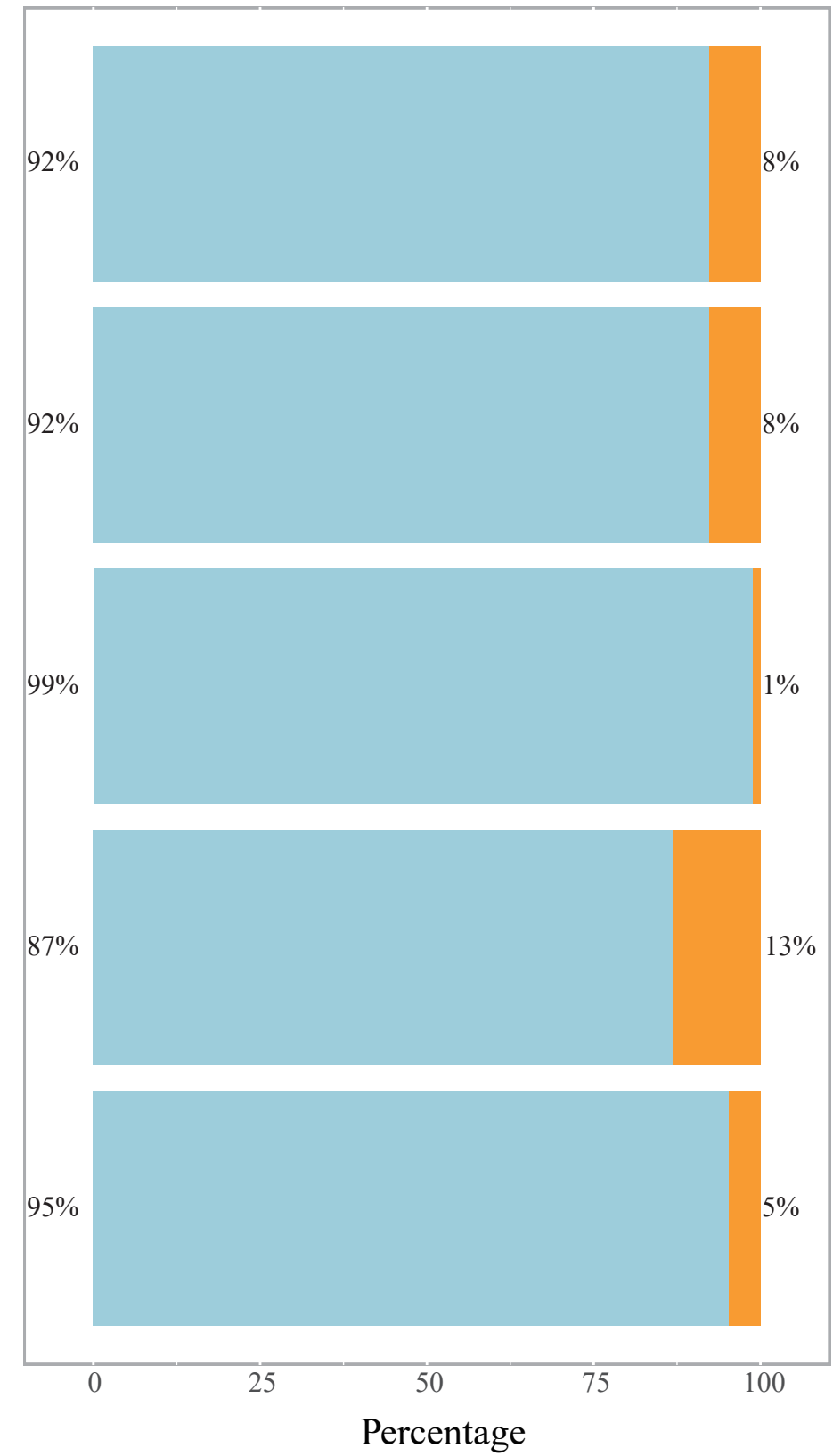
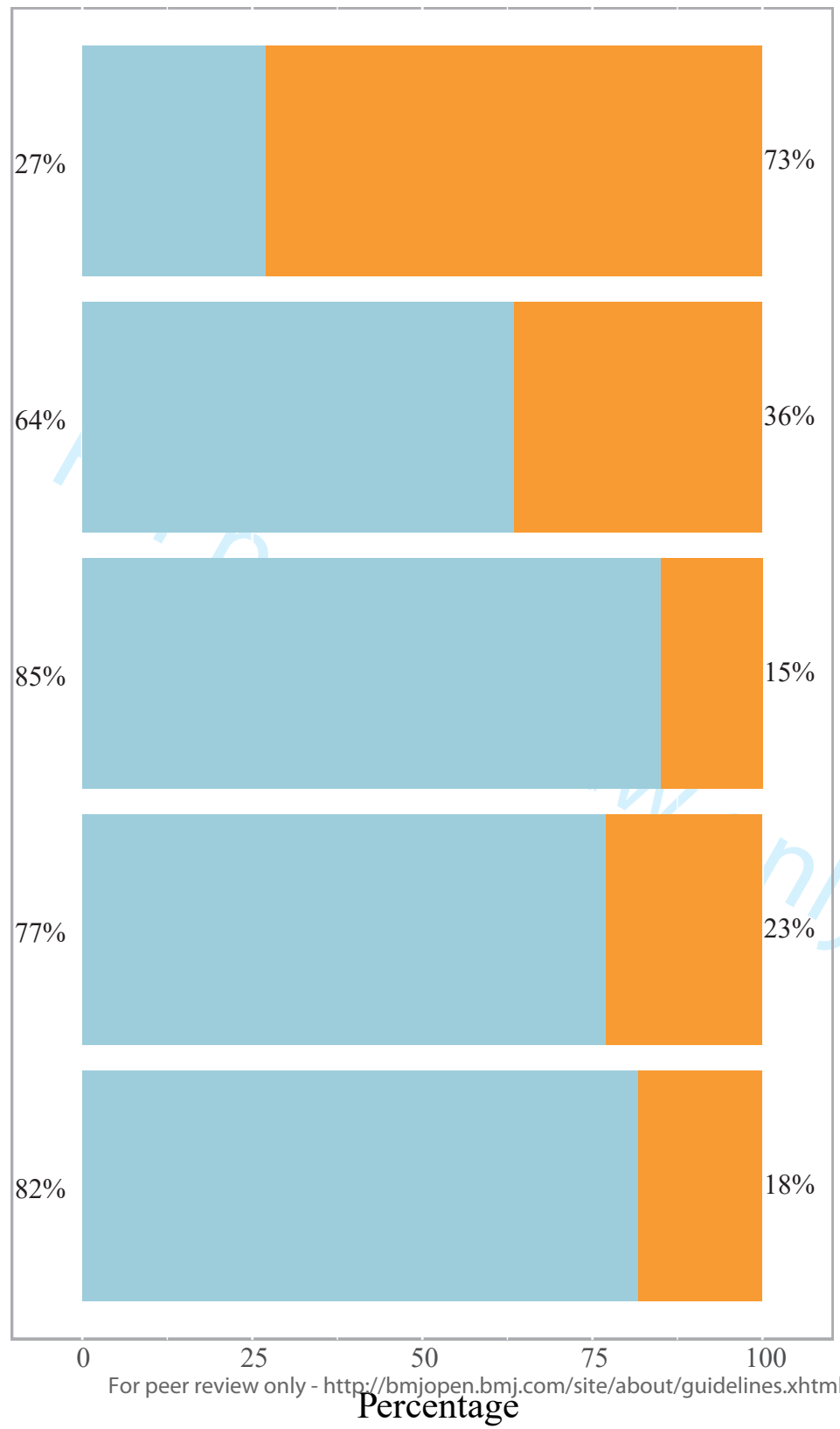
\*I know the content of big data platform

\*I know the potential applications of big data technology

\*Big data technology is practical in medical education

\*Big data technology is practical in clinical practice

\*Big data technology is practical in scientific research





A

BMJ Open

B

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

\*Big data platform could assist future medical education

Big data platform could assist future medical reserch

\*Big data platform could assist future clinical practice

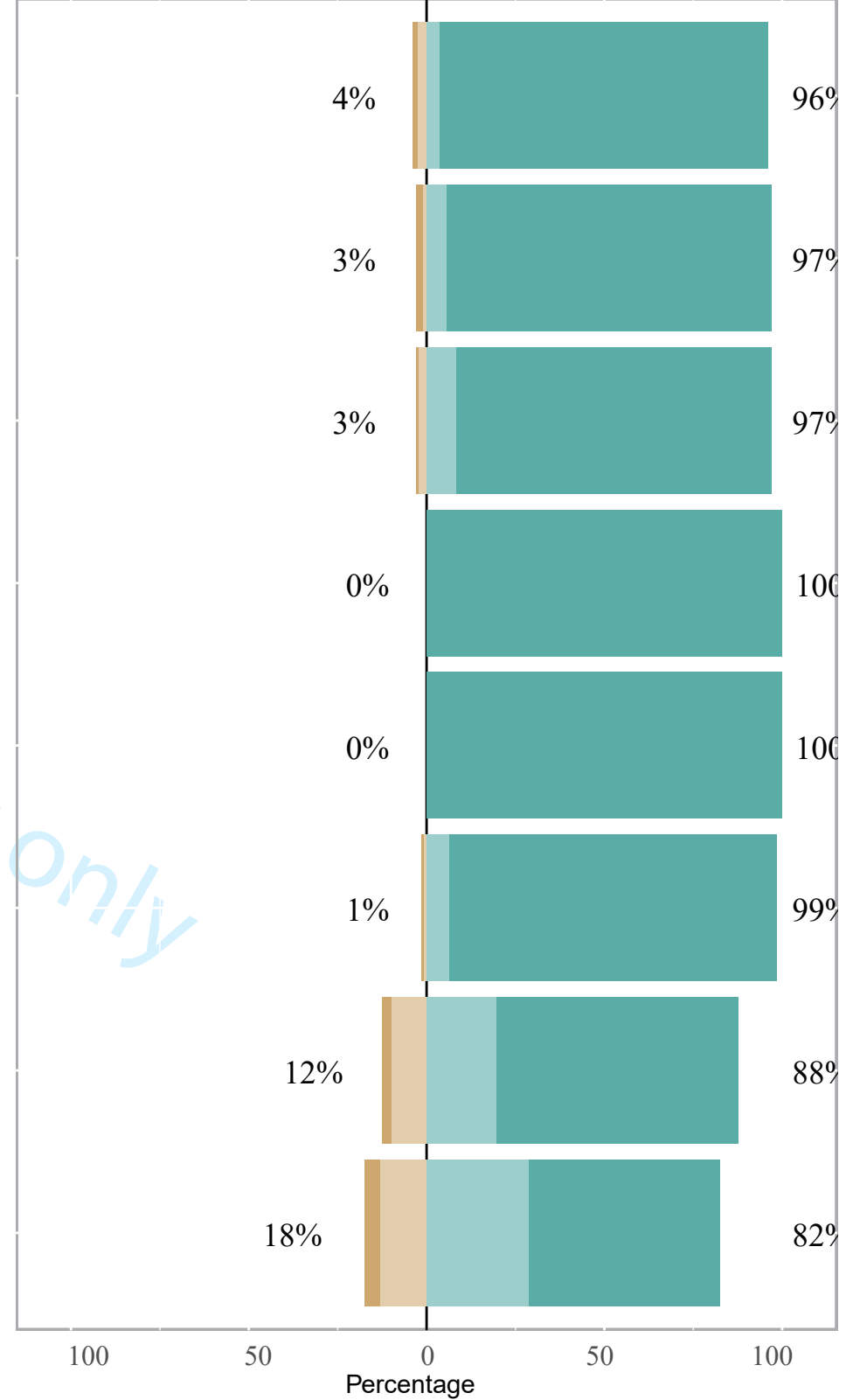
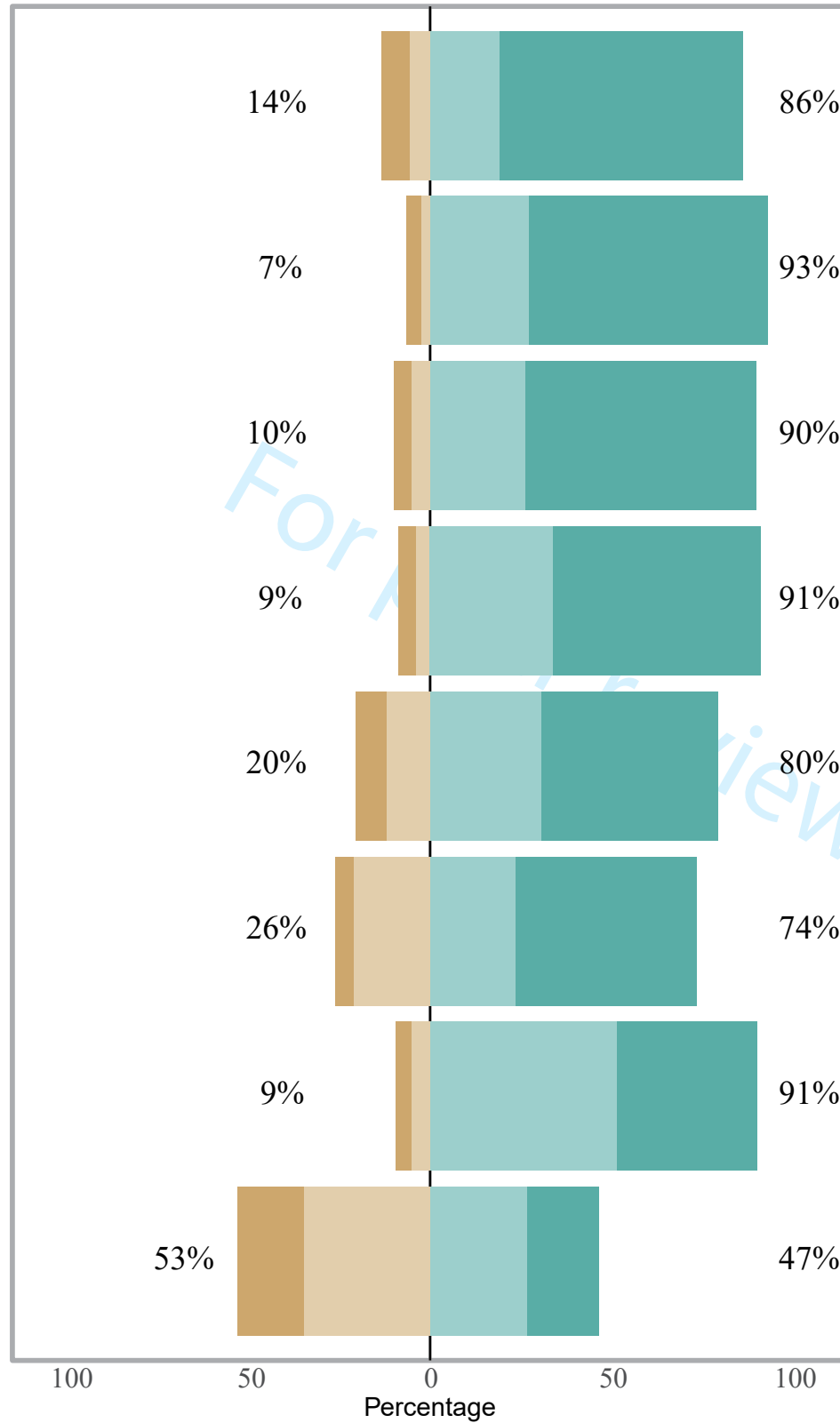
\*I am willing to learn how to use big data platform

\*I am willing to use big data platform in the future

\*I am willing to recommanded big data platform to my colleges

Big data platform could benefit my career

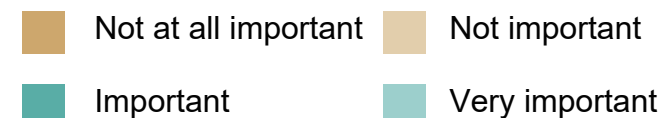
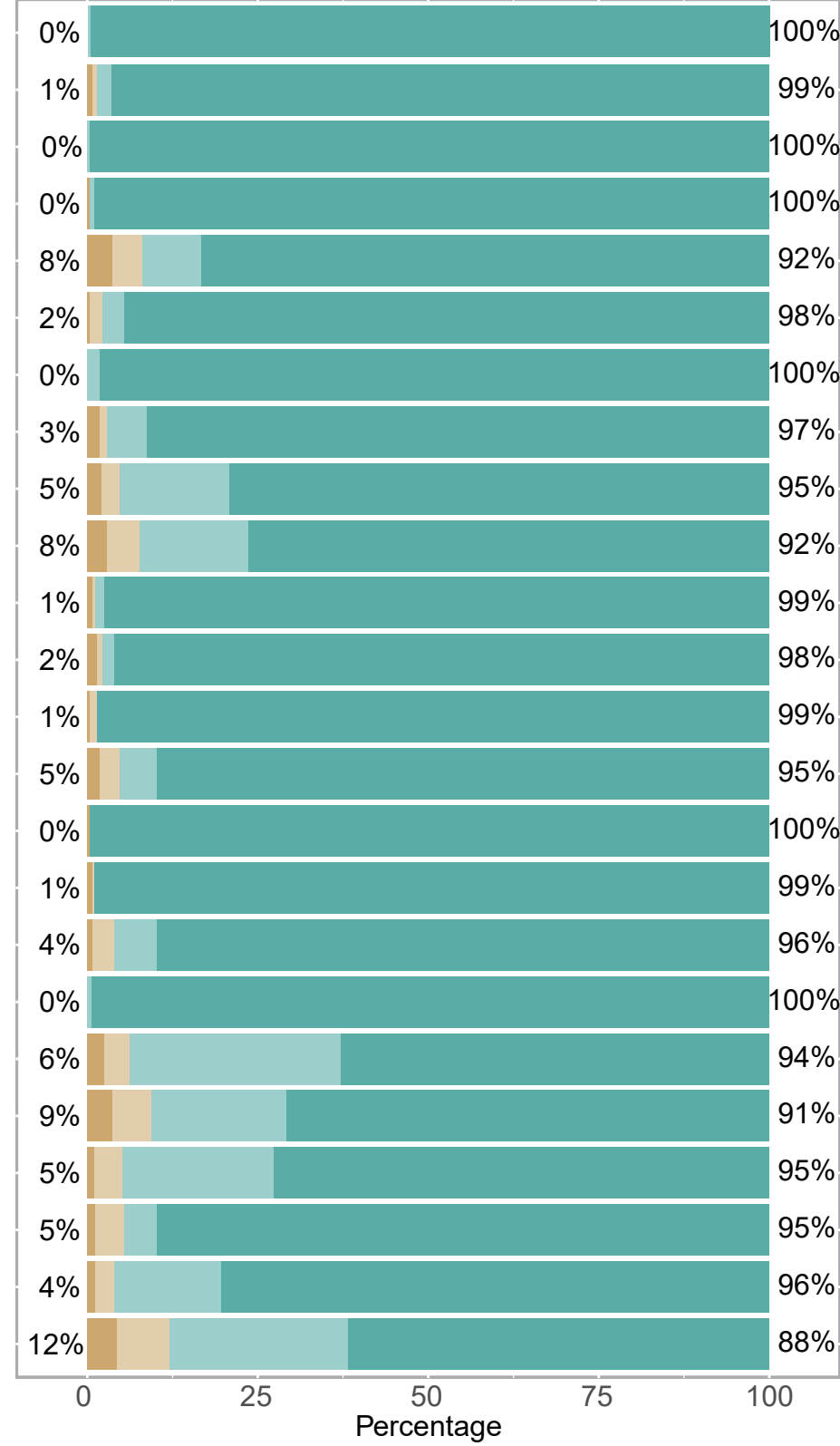
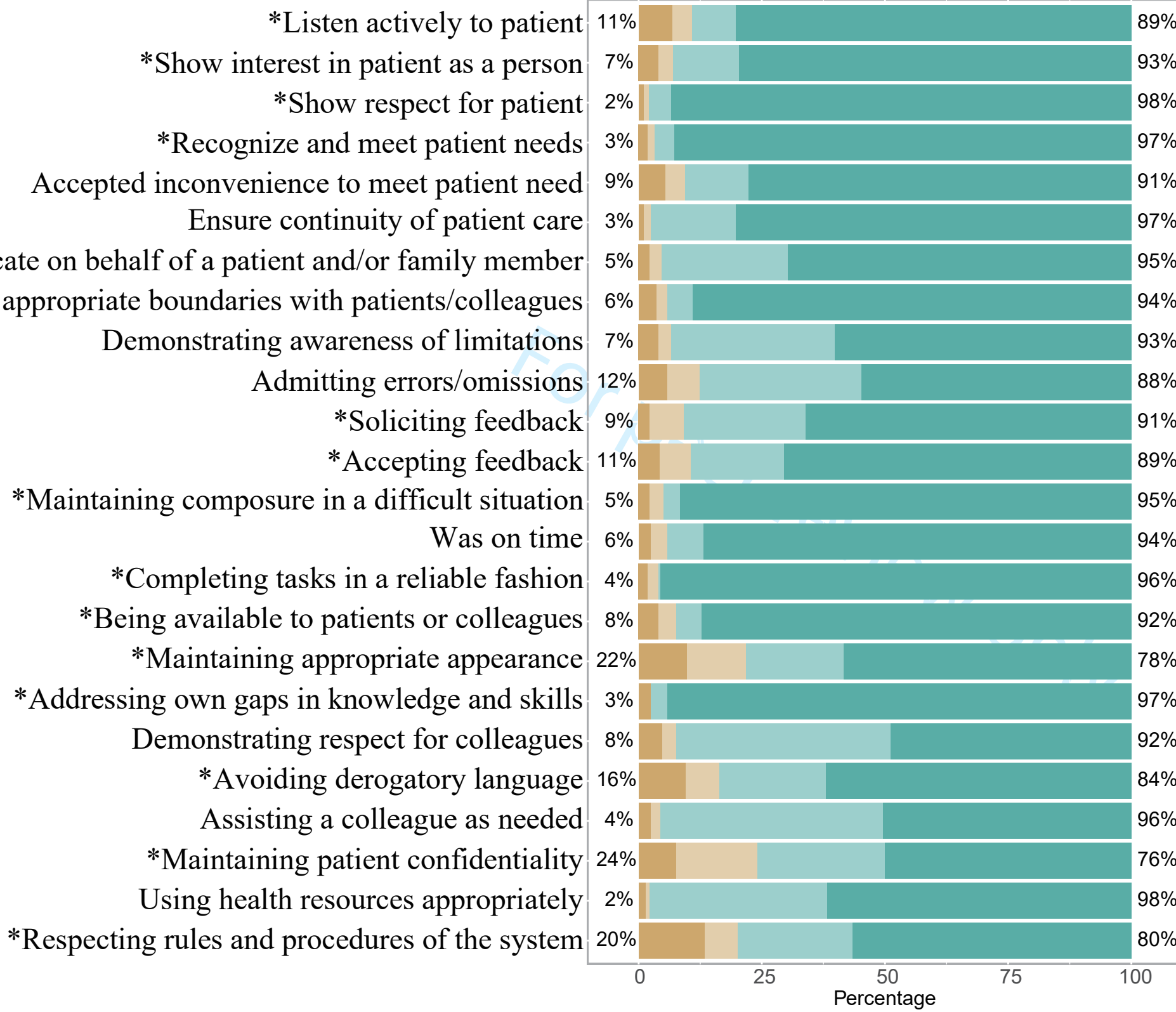
\*Big data platform could benefit all medical careers



Totally disagree Disagree

Totally agree Agree

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53



## Supplemental tables

Table S1, basic knowledge for big data platform before and after the workshop.

Items	Before Workshop		After Workshop		<i>P</i>
	YES, n	NO, n	YES, n	NO, n	
I know the content of big data platform	74	200	253	21	<0.0001
I know the potential applications of big data technology	174	100	253	21	<0.0001
Big data technology is practical in medical education	233	41	271	3	<0.0001
Big data technology is practical in clinical practice	211	63	238	36	0.0039
Big data technology is practical in scientific research	224	50	261	13	<0.0001

Table S2, students' attitude towards big data platform before and after the workshop.

Items	Before Workshop						After Workshop						<i>P</i>
	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	Mean	SD	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	Mean	SD	
Big data platform could assist future medical education	21	16	53	184	3.50	0.910	4	7	10	253	3.87	0.504	0.0002
Big data platform could assist future scientific research	11	7	75	181	3.55	0.735	5	3	16	250	3.86	0.499	0.0705
Big data platform could assist future clinical practice	13	14	72	175	3.49	0.799	2	6	23	243	3.85	0.464	0.0017
I am willing to learn how to use big data platform	13	11	93	157	3.43	0.783	0	0	0	274	4	0	<0.0001

I am willing to use big data platform in the future	23	33	84	134	3.20	0.953	0	0	0	274	4	0	<0.0001
I am willing to recommended big data platform to my colleagues	14	58	65	137	3.19	0.941	2	2	18	252	3.90	0.388	<0.0001
Big data platform could benefit my career	12	14	142	106	3.25	0.744	7	27	54	186	3.53	0.776	0.3382
Big data platform could benefit all medical careers	50	96	74	54	2.48	1.006	12	36	79	147	3.32	0.863	<0.0001

SD, Standard Deviation.

Table S3, the professionalism for students before and after the workshop.

Items	Before workshop				After workshop				P
	Not at all important, n	Not important, n	Important, n	Very important, n	Not at all important, n	Not important, n	Important, n	Very important, n	
Doctor-patient relationship skills									
Listen actively to patient	19	11	24	220	0	0	1	273	<0.0001
Show interest in patient as a person	11	8	37	218	2	2	6	264	0.0029
Show respect for patient	3	3	12	256	0	0	1	273	0.0401
Recognize and meet patient needs	5	4	11	254	1	0	2	271	0.0255
Accepted inconvenience to meet patient need	15	11	35	213	10	12	24	228	0.6503
Ensure continuity of patient care	3	4	47	220	1	5	9	259	1

1	Advocate on behalf of a patient and/or family member	6	7	70	191	0	0	5	269	0.0008
2	Maintained appropriate boundaries with									
3	patients/colleagues	10	6	14	244	5	3	16	250	0.1440
4										
5										
6	Reflective skills									
7										
8	Demonstrating awareness of limitations	11	7	91	165	6	7	44	217	0.4595
9										
10	Admitting errors/omissions	16	18	90	150	8	13	44	209	0.0880
11										
12	Soliciting feedback	6	19	68	181	2	1	4	267	<0.0001
13										
14										
15										
16	Accepting feedback	12	17	52	193	4	2	5	263	0.0001
17										
18	Maintaining composure in a difficult situation	6	8	9	251	1	3	0	270	0.0310
19										
20	Time management									
21										
22	Was on time	7	9	20	238	5	8	15	246	0.7027
23										
24	Completing tasks in a reliable fashion	5	6	1	262	1	0	0	273	0.0086
25										
26	Being available to patients or colleagues	11	10	14	239	2	1	0	271	0.0004
27										
28	Interprofessional relationship skills									
29										
30	Maintaining appropriate appearance	27	33	54	160	2	9	17	246	<0.0001
31										
32										
33										
34	Addressing own gaps in knowledge and skills	7	0	9	258	0	0	2	272	0.0225
35										
36	Demonstrating respect for colleagues	13	8	119	134	7	10	85	172	0.6139
37										
38	Avoiding derogatory language	26	19	59	170	10	16	54	194	0.0220
39										
40	Assisting a colleague as needed	7	5	124	138	3	11	61	199	0.8407
41										
42										
43										
44										
45										
46										

1	Maintaining patient confidentiality	21	45	71	137	3	12	13	246	<0.000
2										<i>I</i>
3										
4	Using health resources appropriately	4	2	99	169	3	8	43	220	0.3244
5										
6	Respecting rules and procedures of the system	37	18	64	155	12	21	72	169	0.0146
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										
39										
40										
41										
42										
43										
44										
45										
46										

For peer review only

## STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	5
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	7-8
		(b) Describe any methods used to examine subgroups and interactions	NA
		(c) Explain how missing data were addressed	7-8
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	7-8
		(e) Describe any sensitivity analyses	NA

Continued on next page

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

<b>Results</b>			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	All participants were eligible and included in this study.
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8-9
		(b) Indicate number of participants with missing data for each variable of interest	Very high response rates with no missing data, as shown in supplementary data.
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	NA
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	NA
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	8
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	In all tables and figures.
		(b) Report category boundaries when continuous variables were categorized	In all tables and figures.
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-16
Generalisability	21	Discuss the generalisability (external validity) of the study results	15-16

**Other information**



1  
2 Funding 22 Give the source of funding and the role of the funders for the present  
3 study and, if applicable, for the original study on which the present article  
4 is based  
5

18

6  
7 \*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and  
8 unexposed groups in cohort and cross-sectional studies.  
9

10  
11 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and  
12 published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely  
13 available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at  
14 <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is  
15 available at [www.strobe-statement.org](http://www.strobe-statement.org).  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Open

## Oncologic big data platform for promoting digital competencies and professionalism in Chinese medical students

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-061015.R2
Article Type:	Original research
Date Submitted by the Author:	18-Aug-2022
Complete List of Authors:	Liu, Jiahao; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Jiao, Xiaofei; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Zeng, Shaoqing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Li, Huayi; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Jin, Ping; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Chi, Jianhua; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Liu, Xingyu; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Yu, Yang; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Ma, Guanchen; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Zhao, Yingjun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Li, Ming; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Peng, Zikun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Huo, Yabing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Gao, Qing-Lei ; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Medical education and training
Keywords:	MEDICAL EDUCATION & TRAINING, Gynaecological oncology < GYNAECOLOGY, Gynaecological oncology < ONCOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60





I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 **1 Oncologic big data platform for promoting digital competencies and**  
5 **2 professionalism in Chinese medical students**  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18

3  
4 Jiahao Liu<sup>1,2#</sup>, Xiaofei Jiao<sup>1,2#</sup>, Shaoqing Zeng<sup>1,2</sup>, Huayi Li<sup>1,2</sup>, Ping Jin<sup>1,2</sup>, Jianhua Chi<sup>1,2</sup>,  
5 Xingyu Liu<sup>1,2</sup>, Yang Yu<sup>1,2</sup>, Guanchen Ma<sup>1,2</sup>, Yingjun Zhao<sup>1,2</sup>, Ming Li<sup>1,2</sup>, Zikun Peng<sup>1,2</sup>,  
6 Yabing Huo<sup>1,2</sup>, Qinglei Gao<sup>1,2\*</sup>  
7  
8

9 <sup>1</sup> Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College,  
10 Huazhong University of Science and Technology, Wuhan 430030, Hubei, P. R. China.

11 <sup>2</sup> Cancer Biology Research Center (Key Laboratory of the Ministry of Education),  
12 Tongji Hospital, Tongji Medical College, Huazhong University of Science and  
13 Technology, Wuhan 430030, Hubei, P. R. China.  
14

15 # These authors contribute equally.  
16

17 \* Correspondence to:

18 Prof. Qinglei Gao, Cancer Biology Research Center (Key Laboratory of the Ministry  
19 of Education), Tongji Hospital, Tongji Medical College, Huazhong University of  
20 Science and Technology, 1095 Jiefang Ave, Wuhan 430030, Hubei, P. R. China. Tel:  
21 +86-27-83662681; Fax: +86-27-83662681; E-mail: qingleigao@hotmail.com.  
22

23 Word count (excluding title page, abstract, references, figures and tables): 3,198  
24

## 25 **Abstract**

26 **Objectives:** Advancements in big data technology are reshaping the health care system  
27 in China. This study aims to explore the role of medical big data in promoting digital  
28 competencies and professionalism among Chinese medical students.

29 **Design, setting and participants:** This study was conducted among 274 medical  
30 students who attended a workshop on medical big data conducted on July 8, 2021 in  
31 Tongji Hospital. The workshop was based on the first nationwide multifunction  
32 gynecologic oncology medical big data platform in China, at the National Union of  
33 Real-World Gynecologic Oncology Research & Patient Management Platform  
34 (NUWA platform).

35 **Outcome measures:** Data on knowledge, attitudes toward big data technology and  
36 professionalism were collected before and after the workshop. We have measured the  
37 four skill categories: doctor–patient relationship skills, reflective skills, time  
38 management, and interprofessional relationship skills using the Professionalism  
39 Mini–Evaluation Exercise (P-MEX) as a reflection for professionalism.

40 **Results:** A total of 274 students participated in this workshop and completed all the  
41 surveys. Before the workshop, only 27% of them knew the detailed content of medical  
42 big data platforms, and 64% knew the potential application of medical big data. The  
43 majority of the students believed that big data technology is practical in their clinical  
44 practice (77%), medical education (85%), and scientific research (82%). Over 80% of  
45 the participants showed positive attitudes toward big data platforms. They also  
46 exhibited sufficient professionalism before the workshop. Meanwhile, the workshop  
47 significantly promoted students' knowledge of medical big data ( $P < 0.05$ ), and led to  
48 more positive attitudes toward big data platforms and higher levels of professionalism.

49 **Conclusions:** Chinese medical students have primitive acquaintance and positive  
50 attitudes toward big data technologies. The NUWA platform-based workshop may  
51 potentially promote their understanding of big data and enhance professionalism,  
52 according to the self-measured P-MEX scale.

53 **Keywords:** Big data, Chinese medical education, Digital competencies,

1  
2  
3  
4 54 Professionalism, Workshop

5  
6 55

7  
8 56 **Strengths and limitations of this study**

9  
10 57 ● This study depicted the knowledge and professionalism of medical big data among  
11  
12 58 Chinese medical students in the era of big data for the first time.

13  
14 59 ● This study provided a vivid example of how big data can assist medical education,  
15  
16 60 based on the first nationwide gynecologic oncology medical big data platform in  
17  
18 61 China (NUWA platform).

19  
20 62 ● The generalization of the findings is limited by the sample representation. The  
21  
22 63 method of measuring attitudes and professionalism may be improved in the future.  
23  
24 64 We need longer time to observe the change of digital competencies and  
25  
26 65 professionalism in Chinese medical students.

27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60 66

## 67 **Background**

68       Electronification of medical records is the signature of the modern health care  
69 system.<sup>[1-3]</sup> Massive clinical and omics data have been produced to enable more detailed  
70 depictions of patients and diseases. Medical big data is thus reshaping our appreciation  
71 of the modern medical system. Technological advancements in data storage, processing,  
72 and analysis accelerated the clinical application of big data-driven products and  
73 contributed to personalized disease management,<sup>[4, 5]</sup> early diagnosis,<sup>[6-8]</sup> and treatment  
74 decision.<sup>[9, 10]</sup> Especially in the COVID-19 pandemic setting, achieving rapid  
75 application of medical big data would meet the pressing clinical need to predict the  
76 progression of diseases using data characteristics.<sup>[11-13]</sup>

77       Big data has brought new challenges for doctors.<sup>[14]</sup> Information overload is a  
78 challenge for every healthcare worker, since they have to acclimate to the nature of big  
79 data, including extraordinary value, volume, velocity, variety, and variability.<sup>[15]</sup>  
80 Challenges can also be expected regarding medical professionalism in the age of big  
81 data. Because big data values realistic health-related information more than ever, the  
82 illusion that digital data outweigh face-to-face physician–patient interactions may  
83 motivate doctors to ignore the importance of professionalism.<sup>[16, 17]</sup> There is no precise  
84 definition of medical professionalism, which is reflected in the attitudes and behaviors  
85 directly related to clinical practice. Epstein RM et al. proposed the definition of  
86 professional competence as wisely using communication, knowledge, technical skills,  
87 clinical reasoning, emotions, values, and reflection in daily practice.<sup>[18]</sup> And Cain J et  
88 al. summarized professionalism as a series of attributes, such as altruism, respect,  
89 honesty, and so on.<sup>[19]</sup> Some studies have shown that digital medical education based  
90 on big data plays a positive effect in promoting professionalism.<sup>[16, 20, 21]</sup> However, there  
91 is a lack of research on the attitude and professionalism of Chinese healthcare workers  
92 regarding medical big data. Therefore, extra lectures and workshops about obtaining  
93 insights into big data and remaining respectful to patients are necessary.

94       In this study, we aimed at depicting the knowledge of medical big data and  
95 professionalism in the era of big data among Chinese medical students. What's more,



1  
2  
3  
4 96 we conducted a workshop for Chinese clinical and preclinical students to improve their  
5  
6 97 knowledge of medical big data and their professionalism in the big data era, based on  
7  
8 98 the National Union of Real-World Gynecologic Oncology Research & Patient  
9  
10 99 Management Platform (NUWA platform), which is the first nationwide multifunction  
11  
12 100 gynecologic oncology medical big data platform in China.

101

## 102 **Methods**

### 103 **Study design**

104 A workshop on the introduction of medical big data was conducted among clinical  
105 and preclinical students in Tongji Medical School. We conducted this study on Chinese  
106 medical students' learning and the application of big data in healthcare using pre- and  
107 post-course questionnaires.

108

### 109 **Workshop design**

110 The workshop was conducted on July 8, 2021 in Tongji Hospital, Tongji Medical  
111 School, Huazhong University of Science and Technology, Wuhan, China. It was part  
112 of the series class of “Medical big data platform learning and applying”.

113 The workshop was delivered by two senior doctors (QLG and SQZ), who have over  
114 5 years of experience in medical big data platform development and application. The  
115 workshop consisted of three major parts, (i) a lecture on the content and application of  
116 medical big data; (ii) a lecture on professionalism for doctors in the big data era; and  
117 (iii) learners being allowed to explore the NUWA platform freely according to their  
118 interest. The first and second sections lasted for approximately 45 minutes, and the third  
119 section lasted for 30 minutes.

120 In section one, five major parts were included in lectures given by the senior doctors,  
121 (i) the development of medical big data in China and the world, (ii) the application and  
122 potential of big data in clinical practice, medical education, and scientific research, (iii)  
123 the construction and content of the NUWA platform, (iv) how to use the NUWA

1  
2  
3  
4 124 platform, and (v) plans for the NUWA platform development. For the second section,  
5  
6 125 the following four items were discussed: (i) the attitude health care workers should have  
7  
8 126 when communicating with patients and colleagues considering that big data have  
9  
10 127 already reshaped our medical system, (ii) the importance of detailed and continuous  
11  
12 128 patient information for the development of big data platforms, (iii) how to protect  
13  
14 129 patient privacy on the databases, and (iv) how to use the NUWA platform to satisfy  
15  
16 130 patients' needs. For Section 3, all students were given a temporary account for the  
17  
18 131 NUWA platform and have access to all deidentified patient information. They were  
19  
20 132 allowed to explore the medical data freely and view structured health care information  
21  
22 133 for half an hour.

23  
24 134

### 25 135 **Data collection**

26  
27 136 Clinical and preclinical students in Tongji Medical School were invited to attend a  
28  
29 137 2-hour class by email or roadshows between June 8, 2021, and July 7, 2021. The  
30  
31 138 contents and speakers for the workshop were presented in the email or during the  
32  
33 139 roadshow. In addition, all 5th- (n=50, preclinical) and 6th-grade (n=49, clinical)  
34  
35 140 students of 8-year undergraduate education in Tongji Medical School took part in this  
36  
37 141 workshop as an additional course.

38  
39 142 The participants completed two surveys, one before the workshop and the other after  
40  
41 143 the workshop. Both questionnaires contain their basic knowledge and attitude toward  
42  
43 144 big data technology and a Professionalism Mini-Evaluation Exercise (P-MEX)  
44  
45 145 instrument. The P-MEX is developed from the mini-Clinical Examination Exercise  
46  
47 146 (mini-CEX) format by Cruess R et al. in 2006 to evaluate professionalism in clinical  
48  
49 147 training.<sup>[22]</sup> It consists of 24 items representing four skill categories: doctor-patient  
50  
51 148 relationship skills, reflective skills, time management, and interprofessional  
52  
53 149 relationship skills skills.<sup>[23, 24]</sup> And the reliability and validity of P-MEX have been  
54  
55 150 confirmed in both Eastern and Western cultural backgrounds.<sup>[25-27]</sup> The attitude scale is  
56  
57 151 a self-created scale focused on measuring participants' pedagogic evaluation,  
58  
59 152 acceptance and expectation of the big data platform. It consisted of 8 questions: (i) Big  
60

1  
2  
3  
4 153 data platform could assist future medical education, (ii) Big data platform could assist  
5  
6 154 future medical research, (iii) Big data platform could assist future clinical practice, (iv)  
7  
8 155 I am willing to learn how to use big data platform, (v) I am willing to use big data  
9  
10 156 platform in the future, (vi) I am willing to recommend big data platform to my  
11  
12 157 colleagues, (vii) Big data platform could benefit my career, and (viii) Big data platform  
13  
14 158 could benefit all medical careers. For each question, students chose from “totally agree”,  
15  
16 159 “agree”, “disagree”, and “totally disagree”. For the applicability of conclusions, “totally  
17  
18 160 agree” and “agree” were regarded as “positive attitude”, and “disagree” and “totally  
19  
20 161 disagree” were considered as “negative attitude”. In addition, the first survey also  
21  
22 162 contained students’ baseline information, and the second survey contained a  
23  
24 163 satisfaction questionnaire. Meanwhile, the browsing histories for participants in Section  
25  
26 164 3 were also recorded and analyzed to reflect students’ interest.

27 165 The data collection and analysis plan were acknowledged and agreed upon by all  
28  
29 166 participants at the beginning of the workshop, and the study was approved by the  
30  
31 167 Research Ethics Commission of Tongji Hospital of Huazhong University of Science  
32  
33 168 and Technology (2020-S201).

34  
35 169

### 36 37 170 **NUWA platform**

38  
39 171 The NUWA platform is the first nationwide Gynecological Oncology data-sharing  
40  
41 172 platform launched by the National Clinical Research Center for Gynecological  
42  
43 173 Oncology in August 2019. This platform integrated inpatient/outpatient clinical data,  
44  
45 174 genomic data, and follow-up data to develop a patient-level longitudinal clinico-  
46  
47 175 genomic database. Information was deidentified and extracted from electronic medical  
48  
49 176 records. A rigorous data quality check was performed to ensure the accuracy of the data  
50  
51 177 entries. Since its foundation in 2019, 17 first-class hospitals from different provinces  
52  
53 178 or cities in China have participated in the NUWA platform until August 2021.

54  
55 179

### 56 57 180 **Statistical analysis**

58  
59 181 Descriptive statistics were presented by counts and percentages to describe the  
60

182 demographic information. The Chi-square test was used to compare the changes in  
 183 knowledge and understanding of big data and professionalism before and after the  
 184 workshop. To acquire more practical results, the categories “not at all important” and  
 185 “not important” were combined for the analysis, as well as “important” and “very  
 186 important”. *P values* <0.05 were considered statistically significant. The data were  
 187 analyzed using R Version 4.03.

188

### 189 Patient and Public Involvement

190 No patients or members of the public were involved in this study.

191

## 192 Results

### 193 Characteristics of students

194 A total of 274 students participated in this workshop and completed two surveys. All  
 195 of them were included in the final analysis. Participants were aged between 22 and 28,  
 196 with 148 (54.0%) preclinical students and 126 (46.0%) clinical students. Among them,  
 197 130 (47.4%) were males and 144 (52.4%) were females (**Table 1**).

198 The majority of participants (207, 75.5%) knew of at least one big data platform in  
 199 China or the world. Over two-thirds of them (183, 66.8%) also acknowledged its  
 200 application. However, only 17.2% (47) of them had been involved in any project related  
 201 to medical big data (**Table 1**).

202

203 **Table 1. Baseline characteristics for participants.**

Characteristics	Number of participants, n (%)
Age, years	
22	27 (9.9%)
23	16 (5.8%)
24	67 (24.5%)
25	66 (24.1%)
26	50 (18.2%)

	27	28 (10.2%)
	28	20 (7.3%)
Gender		
	male	130 (47.4%)
	female	144 (52.6%)
Study stage		
	pre-clinical	148 (54.0%)
	clinical	126 (46.0%)
Acknowledgment of any kind of big data platform		
	yes	207 (75.5%)
	no	67 (24.5%)
Know the applications of big data technology		
	yes	183 (66.8%)
	no	91 (33.2%)
Involved in any big data-related projects		
	yes	47 (17.2%)
	no	227 (82.8%)
Total		274 (100%)

204

### 205 Knowledge of big data platforms

206 Before the workshop, approximately a quarter of the students (74, 27%) knew  
 207 detailed content of at least one medical big data platform, while nearly 64% (174) knew  
 208 the potential application of medical big data platforms. After attending the lessons,  
 209 almost all students could understand the content (253, 92%) and the potential  
 210 application (253, 92%) of medical big data (**Figure 1 and Table S1**).

211 Regardless of the survey completed (pre- or post-course), the majority of students

1  
2  
3  
4 212 believed that big data technology is practical in medical education, clinical practice,  
5  
6 213 and scientific research (85%, 77%, 82% before the workshop; 99%, 87%, 95% after the  
7  
8 214 workshop, respectively) (**Figure 1 and Table S1**).

9  
10 215

### 11 216 **Students' attitudes toward big data platforms**

12  
13 217 Even before the workshop, most students hold a positive attitude on the potential of  
14  
15 218 the big data platform in promoting medical education (237, 86%), medical research  
16  
17 219 (256, 93%), and clinical practices (247, 90%). They were also willing to learn about  
18  
19 220 and use the big data platform (250, 91%, and 218, 80% respectively). Of them, 74%  
20  
21 221 (202) were enthusiastic to introduce big data platforms to their colleagues. Interestingly,  
22  
23 222 most students were convinced that the big data platform could benefit their careers (248,  
24  
25 223 91%), but they were not sure if it would yield the same effect on the others (128, 47%)  
26  
27 224 (**Figure 2 and Table S2**).

28  
29 225 When the workshop was completed, almost all students had positive attitudes toward  
30  
31 226 the big data platform (**Figure 2 and Table S2**). However, there were still 12% (34) and  
32  
33 227 18% (48) of students who were not sure if the big data platform could benefit their and  
34  
35 228 others' medical careers, respectively.

36  
37 229

### 38 230 **Professionalism**

39  
40 231 Generally, positive attitudes toward all the professionalism items were demonstrated,  
41  
42 232 with more than three-quarters of the students agreeing that all professionalism attributes  
43  
44 233 were "important" or "very important" before the workshop. The three items that most  
45  
46 234 students thought were "not at all important" or "not important" were maintaining  
47  
48 235 patient confidentiality (66, 24% of students chose "not important" or "not at all  
49  
50 236 important"), maintaining appropriate appearance (60, 22% students chose "not  
51  
52 237 important" or "not at all important") and respecting rules and procedures of the system  
53  
54 238 (55, 20% students chose "not important" or "not at all important") (**Figure 3A and**  
55  
56 239 **Table S3**).

57  
58 240 After the lectures, students exhibited a more positive attitude on most  
59  
60

241 professionalism items ( $p < 0.05$ ). The most significant improvements occurred for  
 242 “Maintaining patient confidentiality” (from 76% to 95%), “Listen actively to patient”  
 243 (from 89% to 100%), and “Accepting feedback” (from 89% to 98%) (**Figure 3B and**  
 244 **Table S3**).

#### 246 **Interest in medical data**

247 In the free exploration section, most students (253, 92.3%) viewed the patients’  
 248 hospitalization logs. A total of 76.3% (209) of students were interested in the patients’  
 249 history of illness. Meanwhile, many students were interested in medicine usages  
 250 (56.2%), surgery reports (52.9%), and diseases of rare pathologies (58.6%). Only  
 251 approximately one-third of students viewed image diagnosis (37.6%) and follow-up  
 252 records (32.5%) (**Table 2**).

254 **Table 2. Browser records in the free-exploration section.**

Content	Number of participants, n (%)
History of Illness	
yes	209 (76.3%)
no	65 (23.7%)
Hospitalization logs	
yes	253 (92.3%)
no	21 (7.7%)
Medicine usage	
yes	154 (56.2%)
no	120 (43.8%)
Surgery record	
yes	145 (52.9%)
no	129 (47.1%)
Image diagnosis	
yes	103 (37.6%)

	no	171 (62.4%)
Follow-up records	yes	89 (32.5%)
	no	185 (67.5%)
Rare pathologies	yes	160 (58.4%)
	no	114 (41.6%)
Total		274 (100%)

255

256 **Satisfactory survey**

257 Upon completing the survey, 95.6% (262) of participants were “satisfied” or  
 258 “extremely satisfied” with this workshop, and only 7 out of 274 students were not at all  
 259 satisfied. However, we failed to obtain feedback from them in the next three months.  
 260 Most students thought that the workshop was informative (249, 90.9%) and  
 261 understandable (255, 94.1%). The majority of students were also willing to recommend  
 262 this seminar to other students (257, 93.8%) and to participate in similar classes in the  
 263 future (254, 92.7%). Regarding the duration of this workshop, only 7 (2.6%) students  
 264 thought that 2 hours was too long (**Table 3**).

265

266 **Table 3. Participants ‘answers to the workshop satisfaction survey**

	Not at all, n (%)	No, n (%)	Yes, n (%)	Yes, extremely, n (%)	Mean	SD
Overall, are you satisfied with this course	7 (2.6%)	5 (1.8%)	57 (20.8%)	205 (74.8%)	3.68	0.640
Did you think the course are informative?	10 (3.6%)	15 (5.5%)	101 (36.9%)	148 (54%)	3.41	0.757
Did you think the duration of this courses is too long	121 (44.2%)	118 (43.1%)	28 (10.2%)	7 (2.6%)	2.07	0.954



Was the course understandable for you	11 (4%)	8 (2.9%)	121 (44.2%)	134 (48.9%)	3.38	0.733
Would you recommend these courses to other students?	6 (2.2%)	11 (4%)	144 (52.6%)	113 (41.2%)	3.33	0.659
Are you willing to take part in similar courses in the future?	6 (2.2%)	14 (5.1%)	35 (12.8%)	219 (79.9%)	3.70	0.666

267 SD, Standard Deviation.

268

## 269 Discussion

270 Our study demonstrated that Chinese medical students have little knowledge of and  
 271 positive attitudes toward big data technologies. They also yield expertise in  
 272 professionalism. Furthermore, workshops based on big data platforms could further  
 273 strengthen their digital competencies and improve doctor–patient communication  
 274 capabilities, which would lead to better fitness during the expansion of medical big  
 275 data.

276 In this workshop, the basic knowledge of big data platforms in China was measured.  
 277 Although a large fraction of students knew about the big data platform and its utility,  
 278 only a few of them could apprehend its composition or how to use it. This situation may  
 279 be caused by the recent rapid development of big data technology in China and  
 280 comparatively lagging relevant education.<sup>[28, 29]</sup> In addition, ethical challenges that  
 281 hinder medical and public health data sharing may have also exacerbated the  
 282 situation.<sup>[30]</sup> Therefore, aside from conducting big data-related workshops, promoting  
 283 medical and public health data sharing policies and evolving relevant legal and ethical  
 284 implications are also of great importance.

285 Students' enthusiasm toward medical big data was also taken into consideration. As  
 286 expected and consistent with previous reports,<sup>[25]</sup> the majority of students exhibited  
 287 positive attitudes toward the big data platform and were willing to acquire further

1  
2  
3  
4 288 proficiency. They believed that big data would play a vital role in future medical  
5  
6 289 education, clinical practice, and scientific research, which are the main tasks for all  
7  
8 290 Chinese doctors. However, not all of them believed that big data could benefit all  
9  
10 291 medical careers. The popularity of information technology of China recently may  
11  
12 292 contribute to the big data-friendly intention in medical students.<sup>[31]</sup> Meanwhile, their  
13  
14 293 limited understanding of related fields may hamper the belief that it could be used  
15  
16 294 productively.<sup>[32-34]</sup> In the meantime, we were delighted to see that almost all students  
17  
18 295 recognized the significance of big data after our workshop. We thus believe that proper  
19  
20 296 education may be an efficient way to diminish misunderstandings and to achieve full  
21  
22 297 potential in big data platforms.

23  
24 298 Professionalism was related to the development of big data in our study for the first  
25  
26 299 time in China. Considering that big data can reshape medical activities in all aspects,  
27  
28 300 including doctors' attitudes toward their colleagues and patients,<sup>[35]</sup> a lecture about  
29  
30 301 "professionalism in the age of big data" was presented in the current workshop. Before  
31  
32 302 the seminar, students demonstrated sufficient professionalism similar to the findings of  
33  
34 303 previous reports, with high scores on the majority of these items. Nevertheless, it is  
35  
36 304 worth noting that nearly 1/4 of the participants did not pay enough attention to  
37  
38 305 maintaining patient confidentiality. Protecting patients' privacy is an essential  
39  
40 306 embodiment of medical ethics and humanities.<sup>[36]</sup> Participants mostly answered these  
41  
42 307 questions from the perspective of big data users before our workshop, ignoring that  
43  
44 308 those data represented thousands of actual patients. After we emphasized the  
45  
46 309 importance of data privacy in the workshop, students realized that respect for patients  
47  
48 310 is the foundation for improving their medical professionalism. Meanwhile, to enhance  
49  
50 311 patient-data privacy protection, the privacy information, such as ID number, was  
51  
52 312 privatized and converted into an alternative ID number using a hashing algorithm  
53  
54 313 before being uploaded to the NUWA platform. Therefore, personal information of all  
55  
56 314 included patients in NUWA platform is confidential and unreachable for all users.  
57  
58 315 Several important professionalism levels were also significantly increased in many  
59  
60 316 other elements after the workshop, indicating that this workshop could be a preliminary

1  
2  
3  
4 317 attempt to promote professionalism when the prominence of digital data has changed  
5  
6 318 the way we communicate and when doctors have to spend more time with electronic  
7  
8 319 records than with patients. These changes our workshop brought deserve more attention  
9  
10 320 in China since Chinese doctors are well-known to be overburdened.<sup>[37, 38]</sup> These  
11  
12 321 improvements could also help increase the reliability of medical records and produce  
13  
14 322 convincing and effective medical information.

15 323 Another interesting fact in the results is that approximately 60% of students noticed  
16  
17 324 rare diseases in the free-exploration section, which is hardly involved in routine medical  
18  
19 325 classes.<sup>[39, 40]</sup> We believe this is another strength of big data-based medical education.  
20  
21 326 In traditional medical classes, it is arduous for teachers to grant detailed depictions for  
22  
23 327 every type of rare disease. As a result, students may not have easy access to these  
24  
25 328 exceptional cases and therefore are often not able to diagnose rare diseases when  
26  
27 329 reviewing patients' medical records. The big data platform makes it possible for every  
28  
29 330 user to conveniently browse through cases with rare pathologies, which would benefit  
30  
31 331 patients as well as reduce misdiagnoses.

32  
33 332 Our workshop had some limitations. The most important one is that the NUWA  
34  
35 333 platform is still under development, and the omics data are not currently included.  
36  
37 334 Therefore, another investigation should be conducted when the construction of NUWA  
38  
39 335 has been completed. Meanwhile, we did not calculate the sample size before the  
40  
41 336 conduction of this study, which may lead to potential bias. And the limited sample size  
42  
43 337 may be another reason that restricts the popularization of our conclusion. Furthermore,  
44  
45 338 narrow geographical distributions of study participants interfered with the  
46  
47 339 generalization of results to other populations. Meanwhile, there are concerns that  
48  
49 340 students who volunteered to participate in the workshop may be more enthusiastic about  
50  
51 341 big data technology than those who did not. We thus included all 5th- (preclinical) and  
52  
53 342 6th-grade (clinical) students of 8-year undergraduate education in Tongji Medical  
54  
55 343 School to make the study sample more representative. As for the measurement of  
56  
57 344 attitudes and professionalism, a qualitative approach would have been more appropriate  
58  
59 345 and should be considered in the future. Meanwhile, this workshop is too short to  
60

1  
2  
3  
4 346 produce fundamental improvements in students' attitudes and professionalism. In the  
5  
6 347 future, more long-term studies are wanted to draw a clear conclusion.  
7  
8 348

## 9 349 **Conclusion**

10 350 This study depicted Chinese students' knowledge of medical big data for the first  
11  
12 351 time and the NUWA platform-based workshop had potential to improve their  
13  
14 352 understanding of big data and enhance professionalism.  
15  
16 353

## 17 354 **List of abbreviations**

18 355 P-MEX, Professionalism Mini-Evaluation Exercise instrument. NUWA platform, the  
19  
20 356 National Union of Real-World Gynecologic Oncology Research & Patient  
21  
22 357 Management Platform.  
23  
24 358

## 25 359 **Contributorship**

26 360 QLG, JHL and XFJ designed the study and survey. QLG, JHL, XFJ, ML, ZKP, YBH  
27  
28 361 and SQZ conducted the workshop. PJ, HYL, JHC, XYL, YY, and SQZ analyzed and  
29  
30 362 interpreted the data. YJZ and GCM performed interpretation of data and discussion of  
31  
32 363 findings. QLG conceptualized and designed the study, supervised the project, analyzed  
33  
34 364 and interpreted the data, and wrote the paper. All authors approved the final version of  
35  
36 365 the manuscript and agreed to the submission of this manuscript.  
37  
38 366

## 39 367 **Funding statement**

40 368 This work was supported by the Teaching Reform Project for Postgraduates of  
41  
42 369 Huazhong University of Science and Technology (YY202006), the Major Project of  
43  
44 370 Medical Education (20A0769), and Teaching Research Project of The Second Clinical  
45  
46 371 School, Tongji Medical College, Huazhong University of Science and Technology  
47  
48 372 (2021022).  
49  
50 373

## 51 374 **Competing of interests**

1  
2  
3  
4 375 The authors declare that they have no competing interests.

5  
6 376

7  
8 377 **Ethics approval**

9 378 The data collection and analysis plan were acknowledged and agreed by all participants  
10  
11 379 and the study has been approved by the Research Ethics Commission of Tongji Hospital  
12  
13 380 of Huazhong University of Science and Technology (2020-S201).  
14

15 381

16  
17 382 **Data sharing**

18  
19 383 The data used in the current study are available from the corresponding author on  
20  
21 384 reasonable request via e-mail ([qingleigao@hotmail.com](mailto:qingleigao@hotmail.com)).  
22

23 385

24  
25 386 **Consent for publication**

26  
27 387 Not applicable.

28  
29 388

30  
31 389 **Acknowledgements**

32  
33 390 Not applicable.

34  
35 391

36  
37 392 **Reference**

38 393 [1] Bezemer T, de Groot MC, Blasse E, et al. A Human(e) Factor in Clinical Decision Support  
39 394 Systems [J]. *J Med Internet Res*. 2019;21(3):e11732-e.

40 395 [2] Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine [J]. *Ann Intern Med*.  
41 396 2016;164(8):562-3.

42 397 [3] Wu WT, Li YJ, Feng AZ, et al. Data mining in clinical big data: the frequently used databases,  
43 398 steps, and methodological models [J]. *Military Medical Research*. 2021;8(1):44.

44 399 [4] Olatosi B, Zhang J, Weissman S, et al. Using big data analytics to improve HIV medical care  
45 400 utilisation in South Carolina: A study protocol [J]. #N/A. 2019;9(7):e027688.

46 401 [5] Carra G, Salluh JIF, da Silva Ramos FJ, et al. Data-driven ICU management: Using Big Data  
47 402 and algorithms to improve outcomes [J]. *J Crit Care*. 2020;60:300-4.

48 403 [6] Kumar NMS, Eswari T, Sampath P, et al. Predictive Methodology for Diabetic Data Analysis  
49 404 in Big Data In: Vijayakumar V, Neelanarayanan V, editors. *Big Data, Cloud and Computing*  
50 405 *Challenges* [M]. *Procedia Computer Science*. 502015. p. 203-8.

51 406 [7] Razavian N, Blecker S, Schmidt AM, et al. Population-Level Prediction of Type 2 Diabetes  
52 407 From Claims Data and Analysis of Risk Factors [J]. #N/A. 2015;3(4):277-87.

53 408 [8] Fitzgerald RC. Big data is crucial to the early detection of cancer [J]. *Nat Med*. 2020;26(1):19-  
54 409 20.  
55  
56  
57  
58  
59  
60

- 1  
2  
3 410 [9] Fridley BL, Koestler DC, Godwin AK. Individualizing care for ovarian cancer patients using  
4 411 big data [J]. *J Natl Cancer Inst.* 2014;106(5).
- 6 412 [10] Cammarota G, Ianiro G, Ahern A, et al. Gut microbiome, big data and machine learning to  
7 413 promote precision medicine for cancer [J]. *Nat Rev Gastroenterol Hepatol.* 2020;17(10):635-  
8 414 48.
- 10 415 [11] Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19  
11 416 pneumonia on chest CT using multinational datasets [J]. *Nat Commun.* 2020;11(1):4080.
- 12 417 [12] Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with  
13 418 COVID-19 [J]. *Nat Med.* 2020;26(8):1224-8.
- 15 419 [13] Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New  
16 420 Technology, and Proactive Testing [J]. *Jama.* 2020;323(14):1341-2.
- 18 421 [14] Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research  
19 422 applications and clinical care [J]. *Nature reviews Genetics.* 2012;13(6):395-405.
- 20 423 [15] Andreu-Perez J, Poon CC, Merrifield RD, et al. Big data for health [J]. *IEEE J Biomed Health*  
21 424 *Inform.* 2015;19(4):1193-208.
- 23 425 [16] O'Connor S, Zhang M, Honey M, et al. Digital professionalism on social media: A narrative  
24 426 review of the medical, nursing, and allied health education literature [J]. *Int J Med Inf.*  
25 427 2021;153:104514.
- 27 428 [17] Attai DJ, Anderson PF, Fisch MJ, et al. Risks and benefits of Twitter use by  
28 429 hematologists/oncologists in the era of digital medicine [J]. *Seminars in hematology.*  
29 430 2017;54(4):198-204.
- 31 431 [18] Epstein RM, Hundert EM. Defining and assessing professional competence [J]. *Jama.*  
32 432 2002;287(2):226-35.
- 33 433 [19] Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age [J]. *Currents in*  
34 434 *Pharmacy Teaching and Learning.* 2009;1(2):66-70.
- 36 435 [20] Hao X, Peng X, Ding X, et al. Application of digital education in undergraduate nursing and  
37 436 medical interns during the COVID-19 pandemic: A systematic review [J]. *Nurse Educ Today.*  
38 437 2022;108:105183.
- 40 438 [21] Kailey BS, Dev D, Sado DM, et al. CardioWebinar: the evolution of digital education during  
41 439 the COVID-19 pandemic [J]. *Heart.* 2021;107(24):2004-5.
- 43 440 [22] Cruess R, McIlroy JH, Cruess S, et al. The Professionalism Mini-Evaluation Exercise: A  
44 441 Preliminary Investigation [J]. 2006;81(10):S74-S8.
- 45 442 [23] Cruess R, McIlroy JH, Cruess S, et al. The Professionalism Mini-Evaluation Exercise: A  
46 443 Preliminary Investigation [J]. *Academic Medicine.* 2006;81(10).
- 47 444 [24] Norcini JJ, Blank LL, Arnold GK, et al. The mini-CEX (clinical evaluation exercise): a  
48 445 preliminary investigation [J]. *Ann Intern Med.* 1995;123(10):795-9.
- 50 446 [25] Parthiban N, Boland F, Fadil Azim DH, et al. Asian medical students' attitudes towards  
51 447 professionalism [J]. *Med Educ Online.* 2021;26(1):1927466.
- 53 448 [26] Tsugawa Y, Ohbu S, Cruess R, et al. Introducing the Professionalism Mini-Evaluation Exercise  
54 449 (P-MEX) in Japan: results from a multicenter, cross-sectional study [J]. *Acad Med.*  
55 450 2011;86(8):1026-31.
- 57 451 [27] Bajwa NM, Nendaz MR, Posfay-Barbe KM, et al. A Meaningful and Actionable  
58 452 Professionalism Assessment: Validity Evidence for the Professionalism Mini-Evaluation  
59 453 Exercise (P-MEX) Across 8 Years [J]. *Acad Med.* 2021;96(11s):S151-s7.



- 1  
2  
3 454 [28] Wu J, Wang J, Nicholas S, et al. Application of Big Data Technology for COVID-19 Prevention  
4 455 and Control in China: Lessons and Recommendations [J]. *J Med Internet Res.*  
5 456 2020;22(10):e21980.  
6  
7 457 [29] Shen T, Wang C. Big Data Technology Applications and the Right to Health in China during  
8 458 the COVID-19 Pandemic [J]. *International journal of environmental research and public health.*  
9 459 2021;18(14).  
10  
11 460 [30] Li X, Cong Y. A Systematic Literature Review of Ethical Challenges Related to Medical and  
12 461 Public Health Data Sharing in China [J]. *Journal of Empirical Research on Human Research*  
13 462 *Ethics.* 2021:15562646211040299.  
14  
15 463 [31] Liang W, Yao J, Chen A, et al. Early triage of critically ill COVID-19 patients using deep  
16 464 learning [J]. *Nature communications.* 2020;11(1):3543-.  
17  
18 465 [32] Meo SA, Sattar K, Ullah CH, et al. Progress and prospects of medical education research in  
19 466 Asian Countries [J]. *Pakistan journal of medical sciences.* 2019;35(6):1475-81.  
20  
21 467 [33] Schmid RD, Xiong X. Biotech in China 2021, at the beginning of the 14th five-year period  
22 468 ("145") [J]. *Applied microbiology and biotechnology.* 2021;105(10):3971-85.  
23  
24 469 [34] Wu H, Li W, Xie A, et al. Funding of health professional education: China's 20-year process  
25 470 and a global comparison [J]. *Medical education.* 2021.  
26  
27 471 [35] Silva Junior G, Askari M, Oliveira J. Digital Health and Possible Solutions to Improve the Care  
28 472 in the Field of Nephrology [J]. *Contributions to nephrology.* 2021;199:1-15.  
29  
30 473 [36] Doukas DJ, McCullough LB, Wear S. Perspective: Medical education in medical ethics and  
31 474 humanities as the foundation for developing medical professionalism [J]. *Acad Med.*  
32 475 2012;87(3):334-41.  
33  
34 476 [37] Liu S, Chen Y, Li S, et al. What Are the Important Factors Influencing the Recruitment and  
35 477 Retention of Doctoral Students in a Public Health Setting? A Discrete Choice Experiment  
36 478 Survey in China [J]. *International journal of environmental research and public health.*  
37 479 2021;18(18).  
38  
39 480 [38] Jingang A. Which future for doctors in China? [J]. *Lancet.* 2013;382(9896):936-7.  
40  
41 481 [39] Walkowiak D, Domaradzki J. Are rare diseases overlooked by medical education? Awareness  
42 482 of rare diseases among physicians in Poland: an explanatory study [J]. *Orphanet journal of rare*  
43 483 *diseases.* 2021;16(1):400.  
44  
45 484 [40] Domaradzki J, Walkowiak D. Knowledge and Attitudes of Future Healthcare Professionals  
46 485 Toward Rare Diseases [J]. *Front Genet.* 2021;12:639610.  
47  
48 486  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 488 **Figure Legends**

5 489 **Figure 1. Basic knowledge for big data platform (A) before and (B) after the**  
6  
7 **workshop.**  
8

9 491 \* means that there is a significant difference before and after the workshop.

10  
11 492 **Figure 2. Students' attitude towards big data platform (A) before and (B) after the**  
12  
13 **workshop.**  
14

15 494 \* means that there is a significant difference before and after the workshop.

16  
17 495 **Figure 3. The professionalism for students (A) before and (B) after the workshop.**

18  
19 496 \* means that there is a significant difference before and after the workshop.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



A

BMJ Open

B

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

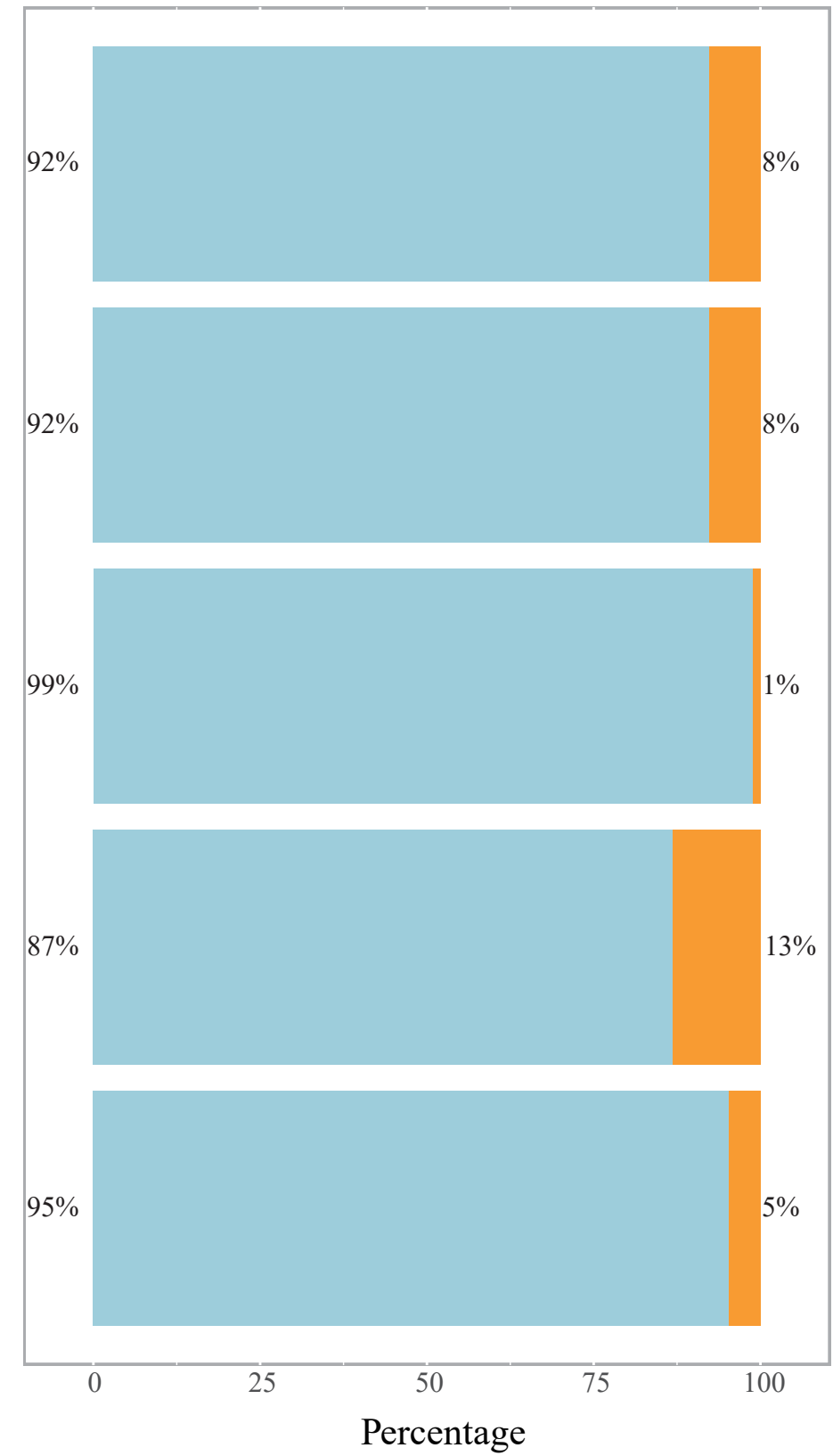
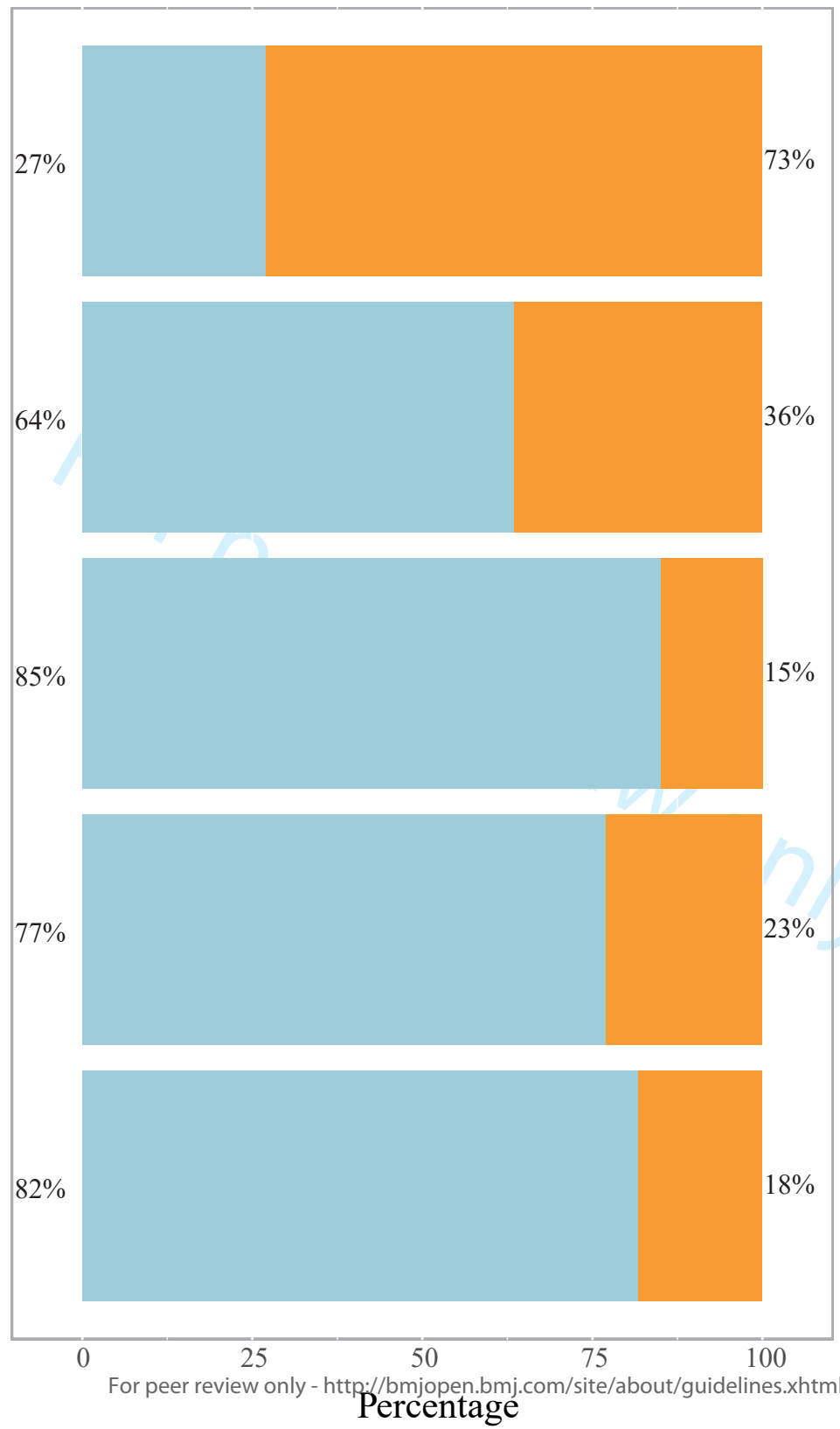
\*I know the content of big data platform

\*I know the potential applications of big data technology

\*Big data technology is practical in medical education

\*Big data technology is practical in clinical practice

\*Big data technology is practical in scientific research



Percentage

Percentage

Yes No

A

BMJ Open

B

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

\*Big data platform could assist future medical education

Big data platform could assist future medical reserch

\*Big data platform could assist future clinical practice

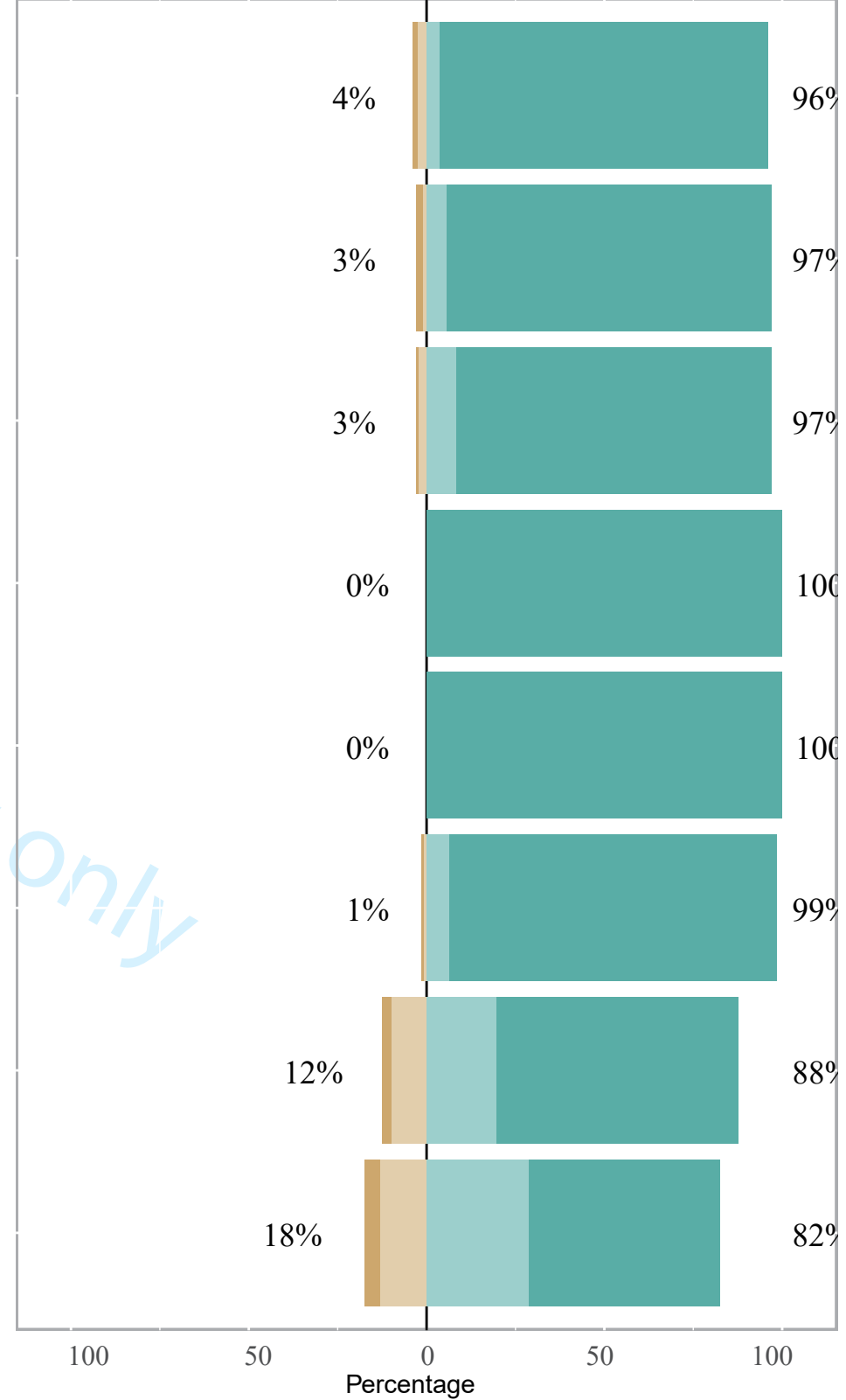
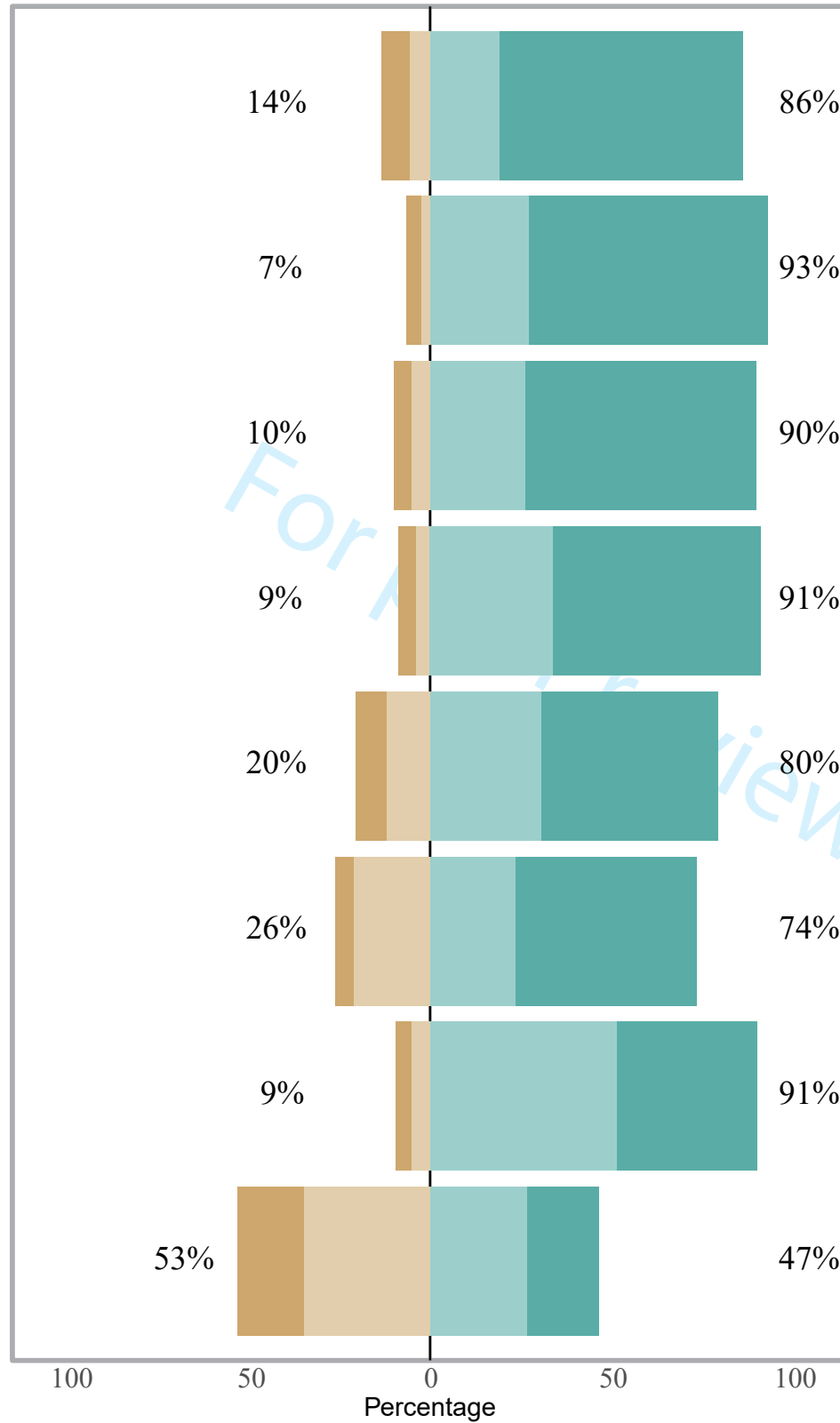
\*I am willing to learn how to use big data platform

\*I am willing to use big data platform in the future

\*I am willing to recommanded big data platform to my colleges

Big data platform could benefit my career

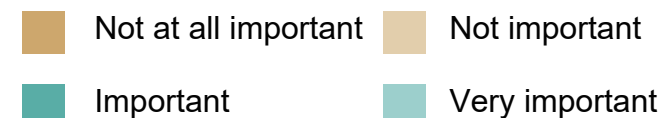
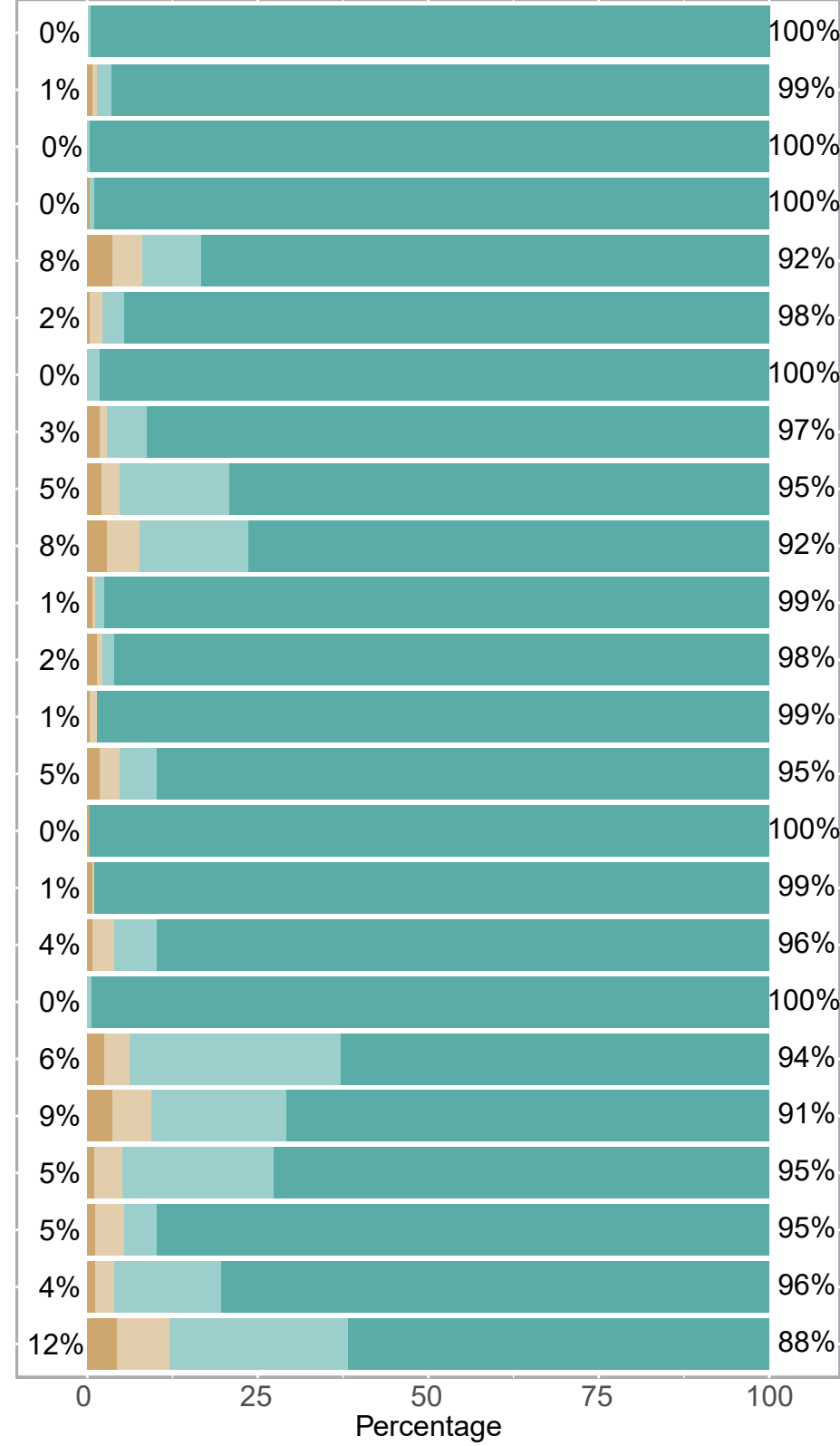
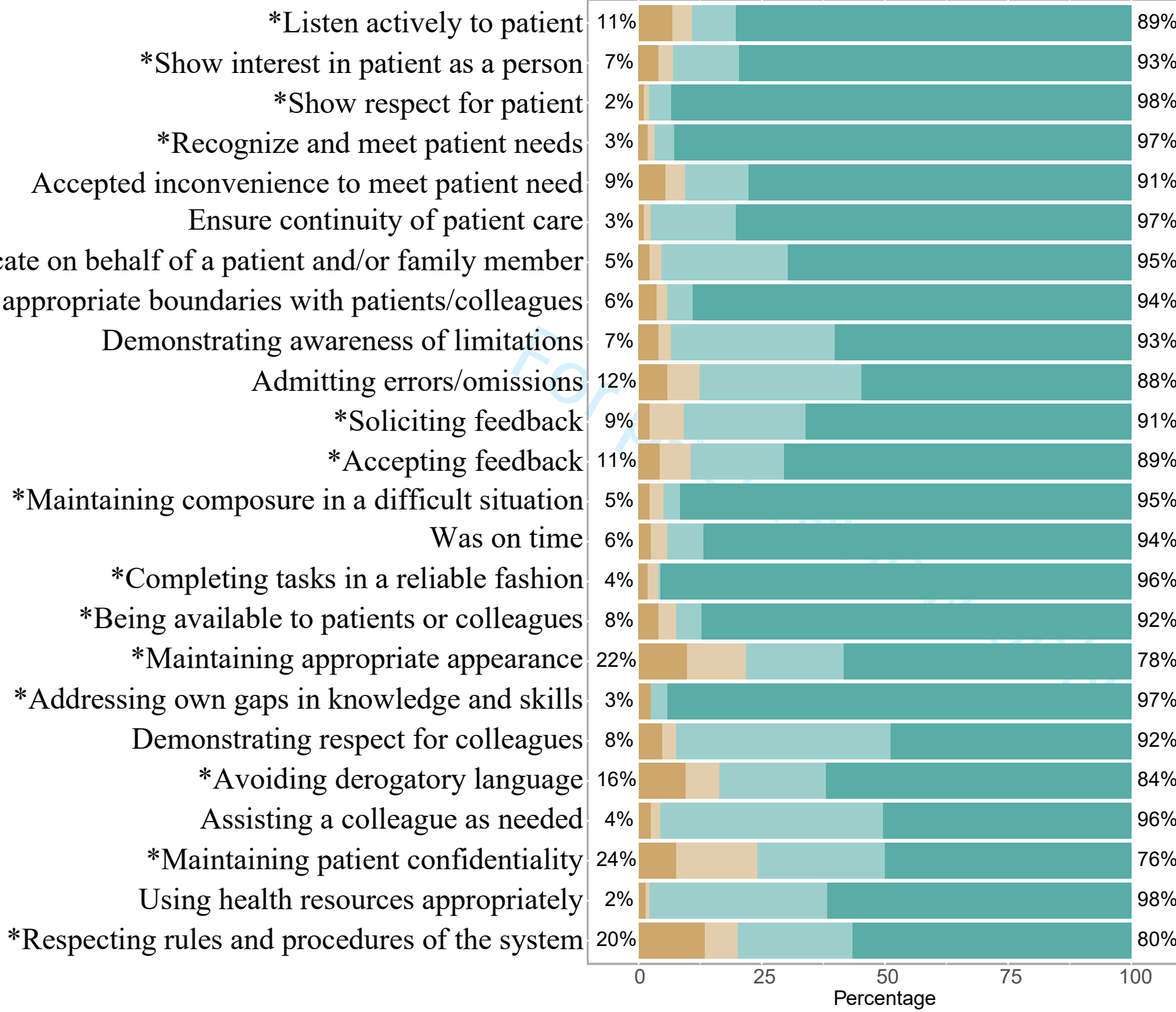
\*Big data platform could benefit all medical careers



Totally disagree Disagree

Totally agree Agree

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53



## Supplemental tables

Table S1, basic knowledge for big data platform before and after the workshop.

Items	Before Workshop		After Workshop		<i>P</i>
	YES, n	NO, n	YES, n	NO, n	
I know the content of big data platform	74	200	253	21	<0.0001
I know the potential applications of big data technology	174	100	253	21	<0.0001
Big data technology is practical in medical education	233	41	271	3	<0.0001
Big data technology is practical in clinical practice	211	63	238	36	0.0039
Big data technology is practical in scientific research	224	50	261	13	<0.0001

Table S2, students' attitude towards big data platform before and after the workshop.

Items	Before Workshop						After Workshop						<i>P</i>
	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	Mean	SD	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	Mean	SD	
Big data platform could assist future medical education	21	16	53	184	3.50	0.910	4	7	10	253	3.87	0.504	0.0002
Big data platform could assist future scientific research	11	7	75	181	3.55	0.735	5	3	16	250	3.86	0.499	0.0705
Big data platform could assist future clinical practice	13	14	72	175	3.49	0.799	2	6	23	243	3.85	0.464	0.0017
I am willing to learn how to use big data platform	13	11	93	157	3.43	0.783	0	0	0	274	4	0	<0.0001

I am willing to use big data platform in the future	23	33	84	134	3.20	0.953	0	0	0	274	4	0	<0.0001
I am willing to recommended big data platform to my colleagues	14	58	65	137	3.19	0.941	2	2	18	252	3.90	0.388	<0.0001
Big data platform could benefit my career	12	14	142	106	3.25	0.744	7	27	54	186	3.53	0.776	0.3382
Big data platform could benefit all medical careers	50	96	74	54	2.48	1.006	12	36	79	147	3.32	0.863	<0.0001

SD, Standard Deviation.

Table S3, the professionalism for students before and after the workshop.

Items	Before workshop				After workshop				P
	Not at all important, n	Not important, n	Important, n	Very important, n	Not at all important, n	Not important, n	Important, n	Very important, n	
Doctor-patient relationship skills									
Listen actively to patient	19	11	24	220	0	0	1	273	<0.0001
Show interest in patient as a person	11	8	37	218	2	2	6	264	0.0029
Show respect for patient	3	3	12	256	0	0	1	273	0.0401
Recognize and meet patient needs	5	4	11	254	1	0	2	271	0.0255
Accepted inconvenience to meet patient need	15	11	35	213	10	12	24	228	0.6503
Ensure continuity of patient care	3	4	47	220	1	5	9	259	1

1	Advocate on behalf of a patient and/or family member	6	7	70	191	0	0	5	269	0.0008
2	3 Maintained appropriate boundaries with 4 patients/colleagues	10	6	14	244	5	3	16	250	0.1440
5										
6	7 Reflective skills									
8	Demonstrating awareness of limitations	11	7	91	165	6	7	44	217	0.4595
9	Admitting errors/omissions	16	18	90	150	8	13	44	209	0.0880
10	12 Soliciting feedback	6	19	68	181	2	1	4	267	<0.000 1
13										
14	Accepting feedback	12	17	52	193	4	2	5	263	0.0001
15	Maintaining composure in a difficult situation	6	8	9	251	1	3	0	270	0.0310
16	20 Time management									
17	Was on time	7	9	20	238	5	8	15	246	0.7027
18	Completing tasks in a reliable fashion	5	6	1	262	1	0	0	273	0.0086
19	Being available to patients or colleagues	11	10	14	239	2	1	0	271	0.0004
20	28 Interprofessional relationship skills									
21	30 Maintaining appropriate appearance	27	33	54	160	2	9	17	246	<0.000 1
22										
23	Addressing own gaps in knowledge and skills	7	0	9	258	0	0	2	272	0.0225
24	Demonstrating respect for colleagues	13	8	119	134	7	10	85	172	0.6139
25	Avoiding derogatory language	26	19	59	170	10	16	54	194	0.0220
26	Assisting a colleague as needed	7	5	124	138	3	11	61	199	0.8407

1	Maintaining patient confidentiality	21	45	71	137	3	12	13	246	<i>&lt;0.000</i>
2										<i>1</i>
3										
4	Using health resources appropriately	4	2	99	169	3	8	43	220	<i>0.3244</i>
5										
6	Respecting rules and procedures of the system	37	18	64	155	12	21	72	169	<i>0.0146</i>
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										
39										
40										
41										
42										
43										
44										
45										
46										

For peer review only

## STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	5
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	7-8
		(b) Describe any methods used to examine subgroups and interactions	NA
		(c) Explain how missing data were addressed	7-8
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	7-8
		(e) Describe any sensitivity analyses	NA

Continued on next page



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

<b>Results</b>			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	All participants were eligible and included in this study.
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8-9
		(b) Indicate number of participants with missing data for each variable of interest	Very high response rates with no missing data, as shown in supplementary data.
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	NA
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	NA
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	8
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	In all tables and figures.
		(b) Report category boundaries when continuous variables were categorized	In all tables and figures.
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16
Generalisability	21	Discuss the generalisability (external validity) of the study results	15
<b>Other information</b>			

1  
2 Funding 22 Give the source of funding and the role of the funders for the present study 16  
3 and, if applicable, for the original study on which the present article is  
4 based  
5

6  
7 \*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and  
8 unexposed groups in cohort and cross-sectional studies.  
9

10  
11 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and  
12 published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely  
13 available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at  
14 <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is  
15 available at [www.strobe-statement.org](http://www.strobe-statement.org).  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Open

## Oncologic big data platform for promoting digital competencies and professionalism in Chinese medical students: a cross-section study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-061015.R3
Article Type:	Original research
Date Submitted by the Author:	24-Aug-2022
Complete List of Authors:	Liu, Jiahao; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Jiao, Xiaofei; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Zeng, Shaoqing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Li, Huayi; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Jin, Ping; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Chi, Jianhua; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Liu, Xingyu; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Yu, Yang; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Ma, Guanchen; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Zhao, Yingjun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Li, Ming; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Peng, Zikun; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Huo, Yabing; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology Gao, Qing-Lei ; Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Medical education and training
Keywords:	MEDICAL EDUCATION & TRAINING, Gynaecological oncology < GYNAECOLOGY, Gynaecological oncology < ONCOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60





I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 **1 Oncologic big data platform for promoting digital competencies and**  
5 **2 professionalism in Chinese medical students: a cross-section study**  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

3

4 Jiahao Liu<sup>1,2#</sup>, Xiaofei Jiao<sup>1,2#</sup>, Shaoqing Zeng<sup>1,2</sup>, Huayi Li<sup>1,2</sup>, Ping Jin<sup>1,2</sup>, Jianhua Chi<sup>1,2</sup>,  
5 Xingyu Liu<sup>1,2</sup>, Yang Yu<sup>1,2</sup>, Guanchen Ma<sup>1,2</sup>, Yingjun Zhao<sup>1,2</sup>, Ming Li<sup>1,2</sup>, Zikun Peng<sup>1,2</sup>,  
6 Yabing Huo<sup>1,2</sup>, Qinglei Gao<sup>1,2\*</sup>  
7  
8

7

8

9 <sup>1</sup> Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College,  
10 Huazhong University of Science and Technology, Wuhan 430030, Hubei, P. R. China.

11 <sup>2</sup> Cancer Biology Research Center (Key Laboratory of the Ministry of Education),  
12 Tongji Hospital, Tongji Medical College, Huazhong University of Science and  
13 Technology, Wuhan 430030, Hubei, P. R. China.  
14

14

15 # These authors contribute equally.  
16

16

17 \* Correspondence to:

18 Prof. Qinglei Gao, Cancer Biology Research Center (Key Laboratory of the Ministry  
19 of Education), Tongji Hospital, Tongji Medical College, Huazhong University of  
20 Science and Technology, 1095 Jiefang Ave, Wuhan 430030, Hubei, P. R. China. Tel:  
21 +86-27-83662681; Fax: +86-27-83662681; E-mail: qingleigao@hotmail.com.  
22

22

23 Word count (excluding title page, abstract, references, figures and tables): 3,202  
24

24

## 25 Abstract

26 **Objectives:** Advancements in big data technology are reshaping the health care system  
27 in China. This study aims to explore the role of medical big data in promoting digital  
28 competencies and professionalism among Chinese medical students.

29 **Design, setting and participants:** This study was conducted among 274 medical  
30 students who attended a workshop on medical big data conducted on July 8, 2021 in  
31 Tongji Hospital. The workshop was based on the first nationwide multifunction  
32 gynecologic oncology medical big data platform in China, at the National Union of  
33 Real-World Gynecologic Oncology Research & Patient Management Platform  
34 (NUWA platform).

35 **Outcome measures:** Data on knowledge, attitudes toward big data technology and  
36 professionalism were collected before and after the workshop. We have measured the  
37 four skill categories: doctor–patient relationship skills, reflective skills, time  
38 management, and interprofessional relationship skills using the Professionalism  
39 Mini–Evaluation Exercise (P-MEX) as a reflection for professionalism.

40 **Results:** A total of 274 students participated in this workshop and completed all the  
41 surveys. Before the workshop, only 27% of them knew the detailed content of medical  
42 big data platforms, and 64% knew the potential application of medical big data. The  
43 majority of the students believed that big data technology is practical in their clinical  
44 practice (77%), medical education (85%), and scientific research (82%). Over 80% of  
45 the participants showed positive attitudes toward big data platforms. They also  
46 exhibited sufficient professionalism before the workshop. Meanwhile, the workshop  
47 significantly promoted students' knowledge of medical big data ( $P < 0.05$ ), and led to  
48 more positive attitudes toward big data platforms and higher levels of professionalism.

49 **Conclusions:** Chinese medical students have primitive acquaintance and positive  
50 attitudes toward big data technologies. The NUWA platform-based workshop may  
51 potentially promote their understanding of big data and enhance professionalism,  
52 according to the self-measured P-MEX scale.

53 **Keywords:** Big data, Chinese medical education, Digital competencies,

1  
2  
3  
4 54 Professionalism, Workshop

5  
6 55

7  
8 56 **Strengths and limitations of this study**

9  
10 57 ● This study depicted the knowledge and professionalism of medical big data among  
11  
12 58 Chinese medical students in the era of big data for the first time.

13  
14 59 ● This study provided a vivid example of how big data can assist medical education,  
15  
16 60 based on the first nationwide gynecologic oncology medical big data platform in  
17  
18 61 China (NUWA platform).

19  
20 62 ● The generalization of the findings is limited by the lack of sample representation  
21  
22 63 and the short follow-up time.

23  
24 64

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

peer review only



## 65 **Background**

66       Electronification of medical records is the signature of the modern health care  
67 system.<sup>[1-3]</sup> Massive clinical and omics data have been produced to enable more detailed  
68 depictions of patients and diseases. Medical big data is thus reshaping our appreciation  
69 of the modern medical system. Technological advancements in data storage, processing,  
70 and analysis accelerated the clinical application of big data-driven products and  
71 contributed to personalized disease management,<sup>[4, 5]</sup> early diagnosis,<sup>[6-8]</sup> and treatment  
72 decision.<sup>[9, 10]</sup> Especially in the COVID-19 pandemic setting, achieving rapid  
73 application of medical big data would meet the pressing clinical need to predict the  
74 progression of diseases using data characteristics.<sup>[11-13]</sup>

75       Big data has brought new challenges for doctors.<sup>[14]</sup> Information overload is a  
76 challenge for every healthcare worker, since they have to acclimate to the nature of big  
77 data, including extraordinary value, volume, velocity, variety, and variability.<sup>[15]</sup>  
78 Challenges can also be expected regarding medical professionalism in the age of big  
79 data. Because big data values realistic health-related information more than ever, the  
80 illusion that digital data outweigh face-to-face physician–patient interactions may  
81 motivate doctors to ignore the importance of professionalism.<sup>[16, 17]</sup> There is no precise  
82 definition of medical professionalism, which is reflected in the attitudes and behaviors  
83 directly related to clinical practice. Epstein RM et al. proposed the definition of  
84 professional competence as wisely using communication, knowledge, technical skills,  
85 clinical reasoning, emotions, values, and reflection in daily practice.<sup>[18]</sup> And Cain J et  
86 al. summarized professionalism as a series of attributes, such as altruism, respect,  
87 honesty, and so on.<sup>[19]</sup> Some studies have shown that digital medical education based  
88 on big data plays a positive effect in promoting professionalism.<sup>[16, 20, 21]</sup> However, there  
89 is a lack of research on the attitude and professionalism of Chinese healthcare workers  
90 regarding medical big data. Therefore, extra lectures and workshops about obtaining  
91 insights into big data and remaining respectful to patients are necessary.

92       In this study, we aimed to depict the knowledge of medical big data and  
93 professionalism in the era of big data among Chinese medical students. What's more,

1  
2  
3  
4 94 we conducted a workshop for Chinese clinical and preclinical students to improve their  
5  
6 95 knowledge of medical big data and their professionalism in the big data era, based on  
7  
8 96 the National Union of Real-World Gynecologic Oncology Research & Patient  
9  
10 97 Management Platform (NUWA platform), which is the first nationwide multifunction  
11  
12 98 gynecologic oncology medical big data platform in China.  
13  
14 99

## 100 **Methods**

### 101 **Study design**

102 A workshop on the introduction of medical big data was conducted among clinical  
103 and preclinical students in Tongji Medical School. We conducted this study on Chinese  
104 medical students' learning and the application of big data in healthcare using pre- and  
105 post-course questionnaires.  
106

### 107 **Workshop design**

108 The workshop was conducted on July 8, 2021 in Tongji Hospital, Tongji Medical  
109 School, Huazhong University of Science and Technology, Wuhan, China. It was part  
110 of the series class of “Medical big data platform learning and applying”.

111 The workshop was delivered by two senior doctors (QLG and SQZ), who have over  
112 five years of experience in medical big data platform development and application. The  
113 workshop consisted of three major parts, (i) a lecture on the content and application of  
114 medical big data; (ii) a lecture on professionalism for doctors in the big data era; and  
115 (iii) learners being allowed to explore the NUWA platform freely according to their  
116 interest. The first and second sections lasted for approximately 45 minutes, and the third  
117 section lasted for 30 minutes.

118 In section one, five major parts were included in lectures given by the senior doctors,  
119 (i) the development of medical big data in China and the world, (ii) the application and  
120 potential of big data in clinical practice, medical education, and scientific research, (iii)  
121 the construction and content of the NUWA platform, (iv) how to use the NUWA

1  
2  
3  
4 122 platform, and (v) plans for the NUWA platform development. For the second section,  
5  
6 123 the following four items were discussed: (i) the attitude health care workers should have  
7  
8 124 when communicating with patients and colleagues, considering that big data have  
9  
10 125 already reshaped our medical system, (ii) the importance of detailed and continuous  
11  
12 126 patient information for the development of big data platforms, (iii) how to protect  
13  
14 127 patient privacy on the databases, and (iv) how to use the NUWA platform to satisfy  
15  
16 128 patients' needs. For Section 3, all students were given a temporary account for the  
17  
18 129 NUWA platform and had access to all deidentified patient information. They were  
19  
20 130 allowed to explore the medical data freely and view structured health care information  
21  
22 131 for half an hour.

23  
24 132

### 25 133 **Data collection**

26  
27 134 Clinical and preclinical students in Tongji Medical School were invited to attend a  
28  
29 135 2-hour class by email or roadshows between June 8, 2021, and July 7, 2021. The  
30  
31 136 contents and speakers for the workshop were presented in the email or during the  
32  
33 137 roadshow. In addition, all 5th- (n=50, preclinical) and 6th-grade (n=49, clinical)  
34  
35 138 students of 8-year undergraduate education in Tongji Medical School took part in this  
36  
37 139 workshop as an additional course.

38  
39 140 The participants completed two surveys, one before the workshop and the other after  
40  
41 141 the workshop. Both questionnaires contain their basic knowledge and attitude toward  
42  
43 142 big data technology and a Professionalism Mini-Evaluation Exercise (P-MEX)  
44  
45 143 instrument. The P-MEX is developed from the mini-Clinical Examination Exercise  
46  
47 144 (mini-CEX) format by Cruess R et al. in 2006 to evaluate professionalism in clinical  
48  
49 145 training.<sup>[22]</sup> It consists of 24 items representing four skill categories: doctor-patient  
50  
51 146 relationship skills, reflective skills, time management, and interprofessional  
52  
53 147 relationship skills.<sup>[23, 24]</sup> And the reliability and validity of P-MEX have been confirmed  
54  
55 148 in both Eastern and Western cultural backgrounds.<sup>[25-27]</sup> The attitude scale is a self-  
56  
57 149 created scale focused on measuring participants' pedagogic evaluation, acceptance and  
58  
59 150 expectation of the big data platform. It consisted of 8 questions: (i) Big data platform

1  
2  
3  
4 151 could assist future medical education, (ii) Big data platform could assist future medical  
5  
6 152 research, (iii) Big data platform could assist future clinical practice, (iv) I am willing to  
7  
8 153 learn how to use big data platform, (v) I am willing to use big data platform in the  
9  
10 154 future, (vi) I am willing to recommend big data platform to my colleagues, (vii) Big  
11  
12 155 data platform could benefit my career, and (viii) Big data platform could benefit all  
13  
14 156 medical careers. For each question, students chose from “totally agree”, “agree”,  
15  
16 157 “disagree”, and “totally disagree”. For the applicability of conclusions, “totally agree”  
17  
18 158 and “agree” were regarded as “positive attitudes”, and “disagree” and “totally disagree”  
19  
20 159 were considered as “negative attitudes”. In addition, the first survey also contained  
21  
22 160 students’ baseline information, and the second survey contained a satisfaction  
23  
24 161 questionnaire. Meanwhile, the browsing histories of participants in Section 3 were also  
25  
26 162 recorded and analyzed to reflect students’ interests.

27 163 The data collection and analysis plan were acknowledged and agreed upon by all  
28  
29 164 participants at the beginning of the workshop, and the study was approved by the  
30  
31 165 Research Ethics Commission of Tongji Hospital of Huazhong University of Science  
32  
33 166 and Technology (2020-S201).

### 167 168 **NUWA platform**

169 The NUWA platform is the first nationwide Gynecological Oncology data-sharing  
170  
171 platform launched by the National Clinical Research Center for Gynecological  
172  
173 Oncology in August 2019. This platform integrated inpatient/outpatient clinical data,  
174  
175 genomic data, and follow-up data to develop a patient-level longitudinal clinico-  
176  
177 genomic database. Information was deidentified and extracted from electronic medical  
178  
179 records. A rigorous data quality check was performed to ensure the accuracy of the data  
180  
181 entries. Since its foundation in 2019, 17 first-class hospitals from different provinces  
182  
183 or cities in China have participated in the NUWA platform until August 2021.

### 178 **Statistical analysis**

179 Descriptive statistics were presented by counts and percentages to describe the

180 demographic information. The Chi-square test was used to compare the changes in  
 181 knowledge and understanding of big data and professionalism before and after the  
 182 workshop. To acquire more practical results, the categories “not at all important” and  
 183 “not important” were combined for the analysis, as well as “important” and “very  
 184 important”. *P values* <0.05 were considered statistically significant. The data were  
 185 analyzed using R Version 4.03.

186

### 187 Patient and Public Involvement

188 No patients or members of the public were involved in this study.

189

## 190 Results

### 191 Characteristics of students

192 A total of 274 students participated in this workshop and completed two surveys. All  
 193 of them were included in the final analysis. Participants were aged between 22 and 28,  
 194 with 148 (54.0%) preclinical students and 126 (46.0%) clinical students. Among them,  
 195 130 (47.4%) were males and 144 (52.4%) were females (**Table 1**).

196 The majority of participants (207, 75.5%) knew of at least one big data platform in  
 197 China or the world. Over two-thirds of them (183, 66.8%) also acknowledged its  
 198 application. However, only 17.2% (47) of them had been involved in any project related  
 199 to medical big data (**Table 1**).

200

201 **Table 1. Baseline characteristics for participants.**

Characteristics	Number of participants, n (%)
Age, years	
22	27 (9.9%)
23	16 (5.8%)
24	67 (24.5%)
25	66 (24.1%)
26	50 (18.2%)

	27	28 (10.2%)
	28	20 (7.3%)
Gender		
	male	130 (47.4%)
	female	144 (52.6%)
Study stage		
	pre-clinical	148 (54.0%)
	clinical	126 (46.0%)
Acknowledgment of any kind of big data platform		
	yes	207 (75.5%)
	no	67 (24.5%)
Know the applications of big data technology		
	yes	183 (66.8%)
	no	91 (33.2%)
Involved in any big data-related projects		
	yes	47 (17.2%)
	no	227 (82.8%)
Total		274 (100%)

202

### 203 Knowledge of big data platforms

204 Before the workshop, approximately a quarter of the students (74, 27%) knew the  
 205 detailed content of at least one medical big data platform, while nearly 64% (174) knew  
 206 the potential application of medical big data platforms. After attending the lessons,  
 207 almost all students could understand the content (253, 92%) and the potential  
 208 application (253, 92%) of medical big data (**Figure 1 and Table S1**).

209 Regardless of the survey completed (pre- or post-course), the majority of students

1  
2  
3  
4 210 believed that big data technology is practical in medical education, clinical practice,  
5  
6 211 and scientific research (85%, 77%, 82% before the workshop; 99%, 87%, 95% after the  
7  
8 212 workshop, respectively) (**Figure 1 and Table S1**).

9  
10 213

#### 11 214 **Students' attitudes toward big data platforms**

12  
13 215 Even before the workshop, most students hold a positive attitude on the potential of  
14  
15 216 the big data platform in promoting medical education (237, 86%), medical research  
16  
17 217 (256, 93%), and clinical practices (247, 90%). They were also willing to learn about  
18  
19 218 and use the big data platform (250, 91%, and 218, 80%, respectively). Of them, 74%  
20  
21 219 (202) were enthusiastic about introducing big data platforms to their colleagues.  
22  
23 220 Interestingly, most students were convinced that the big data platform could benefit  
24  
25 221 their careers (248, 91%), but they were not sure if it would yield the same effect on the  
26  
27 222 others (128, 47%) (**Figure 2 and Table S2**).

28  
29 223 When the workshop was completed, almost all students had positive attitudes toward  
30  
31 224 the big data platform (**Figure 2 and Table S2**). However, there were still 12% (34) and  
32  
33 225 18% (48) of students who were not sure if the big data platform could benefit their and  
34  
35 226 others' medical careers, respectively.

36  
37 227

#### 38 228 **Professionalism**

39  
40 229 Generally, positive attitudes toward all the professionalism items were demonstrated,  
41  
42 230 with more than three-quarters of the students agreeing that all professionalism attributes  
43  
44 231 were "important" or "very important" before the workshop. The three items that most  
45  
46 232 students thought were "not at all important" or "not important" were maintaining  
47  
48 233 patient confidentiality (66, 24% of students chose "not important" or "not at all  
49  
50 234 important"), maintaining appropriate appearance (60, 22% students chose "not  
51  
52 235 important" or "not at all important") and respecting rules and procedures of the system  
53  
54 236 (55, 20% students chose "not important" or "not at all important") (**Figure 3A and**  
55  
56 237 **Table S3**).

57  
58 238 After the lectures, students exhibited a more positive attitude on most  
59  
60

239 professionalism items ( $p < 0.05$ ). The most significant improvements occurred for  
 240 “Maintaining patient confidentiality” (from 76% to 95%), “Listen actively to patient”  
 241 (from 89% to 100%), and “Accepting feedback” (from 89% to 98%) (**Figure 3B and**  
 242 **Table S3**).

#### 244 **Interest in medical data**

245 In the free exploration section, most students (253, 92.3%) viewed the patients’  
 246 hospitalization logs. A total of 76.3% (209) of students were interested in the patients’  
 247 history of illness. Meanwhile, many students were interested in medicine usages  
 248 (56.2%), surgery reports (52.9%), and diseases of rare pathologies (58.6%). Only  
 249 approximately one-third of students viewed image diagnosis (37.6%) and follow-up  
 250 records (32.5%) (**Table 2**).

252 **Table 2. Browser records in the free-exploration section.**

Content	Number of participants, n (%)
History of Illness	
yes	209 (76.3%)
no	65 (23.7%)
Hospitalization logs	
yes	253 (92.3%)
no	21 (7.7%)
Medicine usage	
yes	154 (56.2%)
no	120 (43.8%)
Surgery record	
yes	145 (52.9%)
no	129 (47.1%)
Image diagnosis	
yes	103 (37.6%)



	no	171 (62.4%)
Follow-up records	yes	89 (32.5%)
	no	185 (67.5%)
Rare pathologies	yes	160 (58.4%)
	no	114 (41.6%)
Total		274 (100%)

253

254 **Satisfactory survey**

255 Upon completing the survey, 95.6% (262) of participants were “satisfied” or  
 256 “extremely satisfied” with this workshop, and only 7 out of 274 students were not at all  
 257 satisfied. However, we failed to obtain feedback from them in the next three months.  
 258 Most students thought the workshop was informative (249, 90.9%) and understandable  
 259 (255, 94.1%). The majority of students were also willing to recommend this seminar to  
 260 other students (257, 93.8%) and to participate in similar classes in the future (254,  
 261 92.7%). Regarding the duration of this workshop, only 7 (2.6%) students thought that  
 262 2 hours was too long (**Table 3**).

263

264 **Table 3. Participants ‘answers to the workshop satisfaction survey.**

	Not at all, n (%)	No, n (%)	Yes, n (%)	Yes, extremely, n (%)	Mean	SD
Overall, are you satisfied with this course	7 (2.6%)	5 (1.8%)	57 (20.8%)	205 (74.8%)	3.68	0.640
Did you think the course are informative?	10 (3.6%)	15 (5.5%)	101 (36.9%)	148 (54%)	3.41	0.757
Did you think the duration of this courses is too long	121 (44.2%)	118 (43.1%)	28 (10.2%)	7 (2.6%)	2.07	0.954

Was the course understandable for you	11 (4%)	8 (2.9%)	121 (44.2%)	134 (48.9%)	3.38	0.733
Would you recommend these courses to other students?	6 (2.2%)	11 (4%)	144 (52.6%)	113 (41.2%)	3.33	0.659
Are you willing to take part in similar courses in the future?	6 (2.2%)	14 (5.1%)	35 (12.8%)	219 (79.9%)	3.70	0.666

265 SD, Standard Deviation.

266

## 267 Discussion

268 Our study demonstrated that Chinese medical students have little knowledge of and  
 269 positive attitudes toward big data technologies. They also yield expertise in  
 270 professionalism. Furthermore, workshops based on big data platforms could further  
 271 strengthen their digital competencies and improve doctor–patient communication  
 272 capabilities, which would lead to better fitness during the expansion of medical big  
 273 data.

274 In this workshop, the basic knowledge of big data platforms in China was measured.  
 275 Although a large fraction of students knew about the big data platform and its utility,  
 276 only a few of them could apprehend its composition or how to use it. This situation may  
 277 be caused by the recent rapid development of big data technology in China and  
 278 comparatively lagging relevant education.<sup>[28, 29]</sup> In addition, ethical challenges that  
 279 hinder medical and public health data sharing may have also exacerbated the  
 280 situation.<sup>[30]</sup> Therefore, aside from conducting big data-related workshops, promoting  
 281 medical and public health data sharing policies and evolving relevant legal and ethical  
 282 implications are also of great importance.

283 Students' enthusiasm toward medical big data was also taken into consideration. As  
 284 expected and consistent with previous reports,<sup>[25]</sup> the majority of students exhibited  
 285 positive attitudes toward the big data platform and were willing to acquire further

1  
2  
3  
4 286 proficiency. They believed that big data would play a vital role in future medical  
5  
6 287 education, clinical practice, and scientific research, which are the main tasks for all  
7  
8 288 Chinese doctors. However, not all of them believed that big data could benefit all  
9  
10 289 medical careers. The popularity of information technology in China recently may  
11  
12 290 contribute to the big data-friendly intention of medical students.<sup>[31]</sup> Meanwhile, their  
13  
14 291 limited understanding of related fields may hamper the belief that it could be used  
15  
16 292 productively.<sup>[32-34]</sup> In the meantime, we were delighted to see that almost all students  
17  
18 293 recognized the significance of big data after our workshop. We thus believe that proper  
19  
20 294 education may be an efficient way to diminish misunderstandings and achieve full  
21  
22 295 potential in big data platforms.

23 296 Professionalism was related to the development of big data in our study for the first  
24  
25 297 time in China. Considering that big data can reshape medical activities in all aspects,  
26  
27 298 including doctors' attitudes toward their colleagues and patients,<sup>[35]</sup> a lecture about  
28  
29 299 "professionalism in the age of big data" was presented in the current workshop. Before  
30  
31 300 the seminar, students demonstrated sufficient professionalism, similar to the findings  
32  
33 301 of previous reports, with high scores on the majority of these items. Nevertheless, it is  
34  
35 302 worth noting that nearly 1/4 of the participants did not pay enough attention to  
36  
37 303 maintaining patient confidentiality. Protecting patients' privacy is an essential  
38  
39 304 embodiment of medical ethics and humanities.<sup>[36]</sup> Participants mostly answered these  
40  
41 305 questions from the perspective of big data users before our workshop, ignoring that  
42  
43 306 those data represented thousands of actual patients. After we emphasized the  
44  
45 307 importance of data privacy in the workshop, students realized that respect for patients  
46  
47 308 is the foundation for improving their medical professionalism. Meanwhile, to enhance  
48  
49 309 patient-data privacy protection, the privacy information, such as ID number, was  
50  
51 310 privatized and converted into an alternative ID number using a hashing algorithm  
52  
53 311 before being uploaded to the NUWA platform. Therefore, the personal information of  
54  
55 312 all included patients in the NUWA platform is confidential and unreachable to all users.  
56  
57 313 Several important professionalism levels were also significantly increased in many  
58  
59 314 other elements after the workshop, indicating that this workshop could be a preliminary  
60

1  
2  
3  
4 315 attempt to promote professionalism when the prominence of digital data has changed  
5  
6 316 the way we communicate and when doctors have to spend more time with electronic  
7  
8 317 records than with patients. These changes our workshop brought deserve more attention  
9  
10 318 in China since Chinese doctors are well-known to be overburdened.<sup>[37, 38]</sup> These  
11  
12 319 improvements could also help increase the reliability of medical records and produce  
13  
14 320 convincing and effective medical information.

15  
16 321 Another interesting fact in the results is that approximately 60% of students noticed  
17  
18 322 rare diseases in the free-exploration section, which is hardly involved in routine medical  
19  
20 323 classes.<sup>[39, 40]</sup> We believe this is another strength of big data-based medical education.  
21  
22 324 In traditional medical classes, it is arduous for teachers to grant detailed depictions for  
23  
24 325 every type of rare disease. As a result, students may not have easy access to these  
25  
26 326 exceptional cases and therefore are often not able to diagnose rare diseases when  
27  
28 327 reviewing patients' medical records. The big data platform makes it possible for every  
29  
30 328 user to conveniently browse through cases with rare pathologies, which would benefit  
31  
32 329 patients as well as reduce misdiagnoses.

33  
34 330 Our workshop had some limitations. The most important one is that the NUWA  
35  
36 331 platform is still under development, and the omics data are not currently included.  
37  
38 332 Therefore, another investigation should be conducted when the construction of NUWA  
39  
40 333 has been completed. Meanwhile, we did not calculate the sample size before the  
41  
42 334 conduction of this study, which may lead to potential bias. And the limited sample size  
43  
44 335 may be another reason that restricts the popularization of our conclusion. Furthermore,  
45  
46 336 narrow geographical distributions of study participants interfered with the  
47  
48 337 generalization of results to other populations. Meanwhile, there are concerns that  
49  
50 338 students who volunteered to participate in the workshop may be more enthusiastic about  
51  
52 339 big data technology than those who did not. We thus included all 5th- (preclinical) and  
53  
54 340 6th-grade (clinical) students of 8-year undergraduate education in Tongji Medical  
55  
56 341 School to make the study sample more representative. As for the measurement of  
57  
58 342 attitudes and professionalism, a qualitative approach would have been more appropriate  
59  
60 343 and should be considered in the future. Meanwhile, this workshop is too short to

1  
2  
3  
4 344 produce fundamental improvements in students' attitudes and professionalism. In the  
5 345 future, more long-term studies are wanted to draw a clear conclusion.  
6  
7  
8 346

## 9 347 **Conclusion**

10 348 This study depicted Chinese students' knowledge of medical big data for the first  
11  
12 349 time and the NUWA platform-based workshop had the potential to improve their  
13  
14 350 understanding of big data and enhance professionalism.  
15  
16  
17

18 351

## 19 352 **List of abbreviations**

20 353 P-MEX, Professionalism Mini-Evaluation Exercise instrument. NUWA platform, the  
21  
22 354 National Union of Real-World Gynecologic Oncology Research & Patient  
23  
24 355 Management Platform.  
25  
26  
27

28 356

## 29 357 **Contributorship**

30 358 QLG, JHL and XFJ designed the study and survey. QLG, JHL, XFJ, ML, ZKP, YBH  
31  
32 359 and SQZ conducted the workshop. PJ, HYL, JHC, XYL, YY, and SQZ analyzed and  
33  
34 360 interpreted the data. YJZ and GCM performed interpretation of data and discussion of  
35  
36 361 findings. QLG conceptualized and designed the study, supervised the project, analyzed  
37  
38 362 and interpreted the data, and wrote the paper. All authors approved the final version of  
39  
40 363 the manuscript and agreed to the submission of this manuscript.  
41  
42  
43

44 364

## 45 365 **Funding statement**

46 366 This work was supported by the Teaching Reform Project for Postgraduates of  
47  
48 367 Huazhong University of Science and Technology (YY202006), the Major Project of  
49  
50 368 Medical Education (20A0769), and Teaching Research Project of The Second Clinical  
51  
52 369 School, Tongji Medical College, Huazhong University of Science and Technology  
53  
54 370 (2021022).  
55  
56

57 371

## 58 372 **Competing of interests**

1  
2  
3  
4 373 The authors declare that they have no competing interests.

5  
6 374

7  
8 375 **Ethics approval**

9 376 The data collection and analysis plan were acknowledged and agreed by all participants  
10  
11 377 and the study has been approved by the Research Ethics Commission of Tongji Hospital  
12  
13 378 of Huazhong University of Science and Technology (2020-S201).

14  
15 379

16  
17 380 **Data sharing**

18  
19 381 The data used in the current study are available from the corresponding author on  
20  
21 382 reasonable request via e-mail ([qingleigao@hotmail.com](mailto:qingleigao@hotmail.com)).

22  
23 383

24  
25 384 **Consent for publication**

26  
27 385 Not applicable.

28  
29 386

30  
31 387 **Acknowledgements**

32  
33 388 Not applicable.

34  
35 389

36  
37 390 **Reference**

38 391 [1] Bezemer T, de Groot MC, Blasse E, et al. A Human(e) Factor in Clinical Decision Support  
39 392 Systems [J]. *J Med Internet Res*. 2019;21(3):e11732-e.

40 393 [2] Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine [J]. *Ann Intern Med*.  
41 394 2016;164(8):562-3.

42 395 [3] Wu WT, Li YJ, Feng AZ, et al. Data mining in clinical big data: the frequently used databases,  
43 396 steps, and methodological models [J]. *Military Medical Research*. 2021;8(1):44.

44 397 [4] Olatosi B, Zhang J, Weissman S, et al. Using big data analytics to improve HIV medical care  
45 398 utilisation in South Carolina: A study protocol [J]. #N/A. 2019;9(7):e027688.

46 399 [5] Carra G, Salluh JIF, da Silva Ramos FJ, et al. Data-driven ICU management: Using Big Data  
47 400 and algorithms to improve outcomes [J]. *J Crit Care*. 2020;60:300-4.

48 401 [6] Kumar NMS, Eswari T, Sampath P, et al. Predictive Methodology for Diabetic Data Analysis  
49 402 in Big Data In: Vijayakumar V, Neelanarayanan V, editors. *Big Data, Cloud and Computing*  
50 403 *Challenges* [M]. *Procedia Computer Science*. 502015. p. 203-8.

51 404 [7] Razavian N, Blecker S, Schmidt AM, et al. Population-Level Prediction of Type 2 Diabetes  
52 405 From Claims Data and Analysis of Risk Factors [J]. #N/A. 2015;3(4):277-87.

53 406 [8] Fitzgerald RC. Big data is crucial to the early detection of cancer [J]. *Nat Med*. 2020;26(1):19-  
54 407 20.

- 1  
2  
3 408 [9] Fridley BL, Koestler DC, Godwin AK. Individualizing care for ovarian cancer patients using  
4 409 big data [J]. *J Natl Cancer Inst.* 2014;106(5).
- 6 410 [10] Cammarota G, Ianiro G, Ahern A, et al. Gut microbiome, big data and machine learning to  
7 411 promote precision medicine for cancer [J]. *Nat Rev Gastroenterol Hepatol.* 2020;17(10):635-  
8 412 48.
- 10 413 [11] Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19  
11 414 pneumonia on chest CT using multinational datasets [J]. *Nat Commun.* 2020;11(1):4080.
- 12 415 [12] Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with  
13 416 COVID-19 [J]. *Nat Med.* 2020;26(8):1224-8.
- 15 417 [13] Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New  
16 418 Technology, and Proactive Testing [J]. *Jama.* 2020;323(14):1341-2.
- 18 419 [14] Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research  
19 420 applications and clinical care [J]. *Nature reviews Genetics.* 2012;13(6):395-405.
- 20 421 [15] Andreu-Perez J, Poon CC, Merrifield RD, et al. Big data for health [J]. *IEEE J Biomed Health*  
21 422 *Inform.* 2015;19(4):1193-208.
- 23 423 [16] O'Connor S, Zhang M, Honey M, et al. Digital professionalism on social media: A narrative  
24 424 review of the medical, nursing, and allied health education literature [J]. *Int J Med Inf.*  
25 425 2021;153:104514.
- 27 426 [17] Attai DJ, Anderson PF, Fisch MJ, et al. Risks and benefits of Twitter use by  
28 427 hematologists/oncologists in the era of digital medicine [J]. *Seminars in hematology.*  
29 428 2017;54(4):198-204.
- 31 429 [18] Epstein RM, Hundert EM. Defining and assessing professional competence [J]. *Jama.*  
32 430 2002;287(2):226-35.
- 33 431 [19] Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age [J]. *Currents in*  
34 432 *Pharmacy Teaching and Learning.* 2009;1(2):66-70.
- 36 433 [20] Hao X, Peng X, Ding X, et al. Application of digital education in undergraduate nursing and  
37 434 medical interns during the COVID-19 pandemic: A systematic review [J]. *Nurse Educ Today.*  
38 435 2022;108:105183.
- 40 436 [21] Kailey BS, Dev D, Sado DM, et al. CardioWebinar: the evolution of digital education during  
41 437 the COVID-19 pandemic [J]. *Heart.* 2021;107(24):2004-5.
- 43 438 [22] Cruess R, McIlroy JH, Cruess S, et al. The Professionalism Mini-Evaluation Exercise: A  
44 439 Preliminary Investigation [J]. 2006;81(10):S74-S8.
- 45 440 [23] Cruess R, McIlroy JH, Cruess S, et al. The Professionalism Mini-Evaluation Exercise: A  
46 441 Preliminary Investigation [J]. *Academic Medicine.* 2006;81(10).
- 47 442 [24] Norcini JJ, Blank LL, Arnold GK, et al. The mini-CEX (clinical evaluation exercise): a  
48 443 preliminary investigation [J]. *Ann Intern Med.* 1995;123(10):795-9.
- 50 444 [25] Parthiban N, Boland F, Fadil Azim DH, et al. Asian medical students' attitudes towards  
51 445 professionalism [J]. *Med Educ Online.* 2021;26(1):1927466.
- 53 446 [26] Tsugawa Y, Ohbu S, Cruess R, et al. Introducing the Professionalism Mini-Evaluation Exercise  
54 447 (P-MEX) in Japan: results from a multicenter, cross-sectional study [J]. *Acad Med.*  
55 448 2011;86(8):1026-31.
- 57 449 [27] Bajwa NM, Nendaz MR, Posfay-Barbe KM, et al. A Meaningful and Actionable  
58 450 Professionalism Assessment: Validity Evidence for the Professionalism Mini-Evaluation  
59 451 Exercise (P-MEX) Across 8 Years [J]. *Acad Med.* 2021;96(11s):S151-s7.



- 1  
2  
3 452 [28] Wu J, Wang J, Nicholas S, et al. Application of Big Data Technology for COVID-19 Prevention  
4 453 and Control in China: Lessons and Recommendations [J]. *J Med Internet Res.*  
5 454 2020;22(10):e21980.  
6  
7 455 [29] Shen T, Wang C. Big Data Technology Applications and the Right to Health in China during  
8 456 the COVID-19 Pandemic [J]. *International journal of environmental research and public health.*  
9 457 2021;18(14).  
10  
11 458 [30] Li X, Cong Y. A Systematic Literature Review of Ethical Challenges Related to Medical and  
12 459 Public Health Data Sharing in China [J]. *Journal of Empirical Research on Human Research*  
13 460 *Ethics.* 2021:15562646211040299.  
14  
15 461 [31] Liang W, Yao J, Chen A, et al. Early triage of critically ill COVID-19 patients using deep  
16 462 learning [J]. *Nature communications.* 2020;11(1):3543-.  
17  
18 463 [32] Meo SA, Sattar K, Ullah CH, et al. Progress and prospects of medical education research in  
19 464 Asian Countries [J]. *Pakistan journal of medical sciences.* 2019;35(6):1475-81.  
20  
21 465 [33] Schmid RD, Xiong X. Biotech in China 2021, at the beginning of the 14th five-year period  
22 466 ("145") [J]. *Applied microbiology and biotechnology.* 2021;105(10):3971-85.  
23  
24 467 [34] Wu H, Li W, Xie A, et al. Funding of health professional education: China's 20-year process  
25 468 and a global comparison [J]. *Medical education.* 2021.  
26  
27 469 [35] Silva Junior G, Askari M, Oliveira J. Digital Health and Possible Solutions to Improve the Care  
28 470 in the Field of Nephrology [J]. *Contributions to nephrology.* 2021;199:1-15.  
29  
30 471 [36] Doukas DJ, McCullough LB, Wear S. Perspective: Medical education in medical ethics and  
31 472 humanities as the foundation for developing medical professionalism [J]. *Acad Med.*  
32 473 2012;87(3):334-41.  
33  
34 474 [37] Liu S, Chen Y, Li S, et al. What Are the Important Factors Influencing the Recruitment and  
35 475 Retention of Doctoral Students in a Public Health Setting? A Discrete Choice Experiment  
36 476 Survey in China [J]. *International journal of environmental research and public health.*  
37 477 2021;18(18).  
38  
39 478 [38] Jingang A. Which future for doctors in China? [J]. *Lancet.* 2013;382(9896):936-7.  
40  
41 479 [39] Walkowiak D, Domaradzki J. Are rare diseases overlooked by medical education? Awareness  
42 480 of rare diseases among physicians in Poland: an explanatory study [J]. *Orphanet journal of rare*  
43 481 *diseases.* 2021;16(1):400.  
44  
45 482 [40] Domaradzki J, Walkowiak D. Knowledge and Attitudes of Future Healthcare Professionals  
46 483 Toward Rare Diseases [J]. *Front Genet.* 2021;12:639610.  
47  
48 484  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4 486 **Figure Legends**

5 487 **Figure 1. Basic knowledge for big data platform (A) before and (B) after the**  
6  
7  
8 488 **workshop.**

9 489 \* means that there is a significant difference before and after the workshop.

10  
11 490 **Figure 2. Students' attitude towards big data platform (A) before and (B) after the**  
12  
13 491 **workshop.**

14  
15 492 \* means that there is a significant difference before and after the workshop.

16  
17 493 **Figure 3. The professionalism for students (A) before and (B) after the workshop.**

18  
19 494 \* means that there is a significant difference before and after the workshop.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

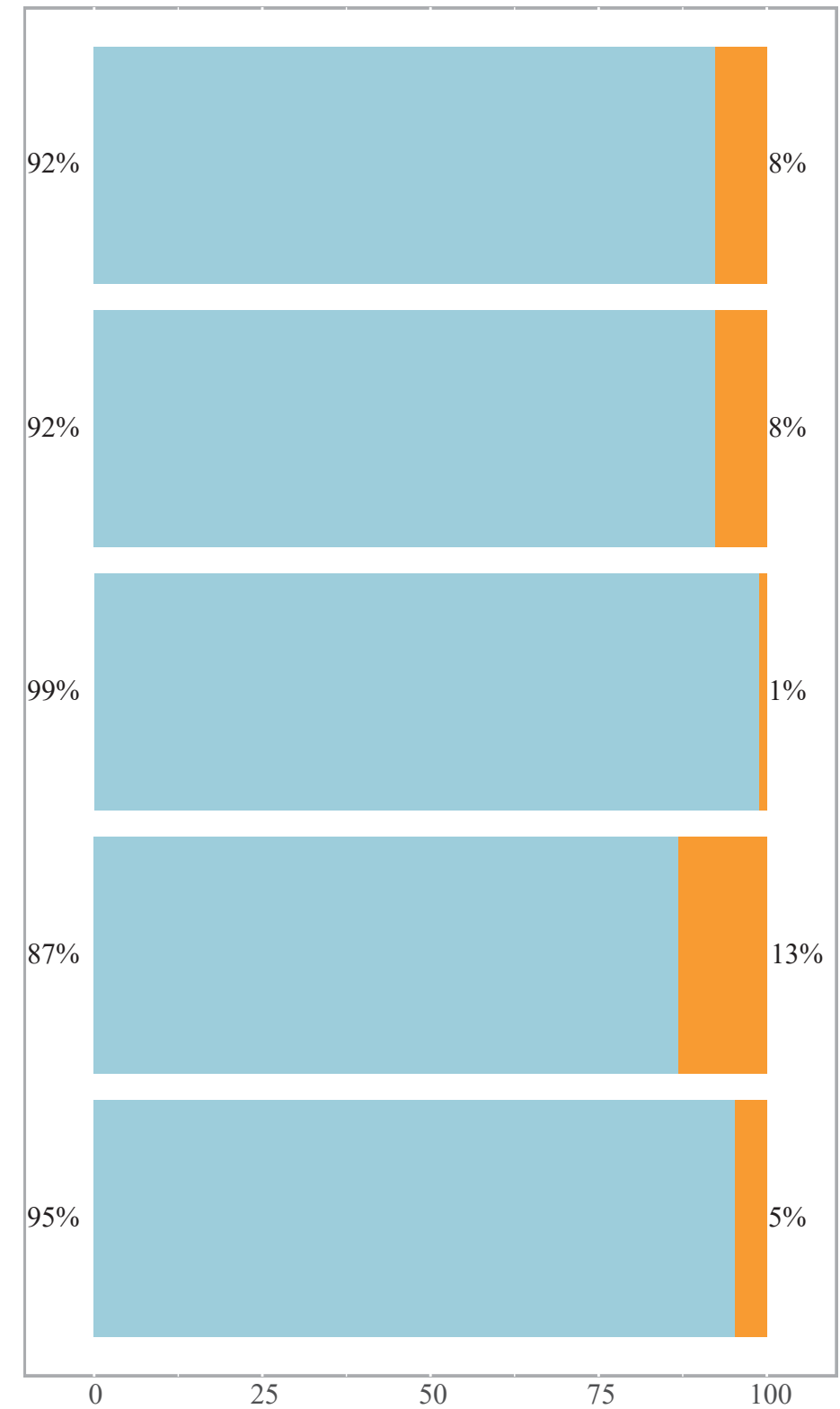
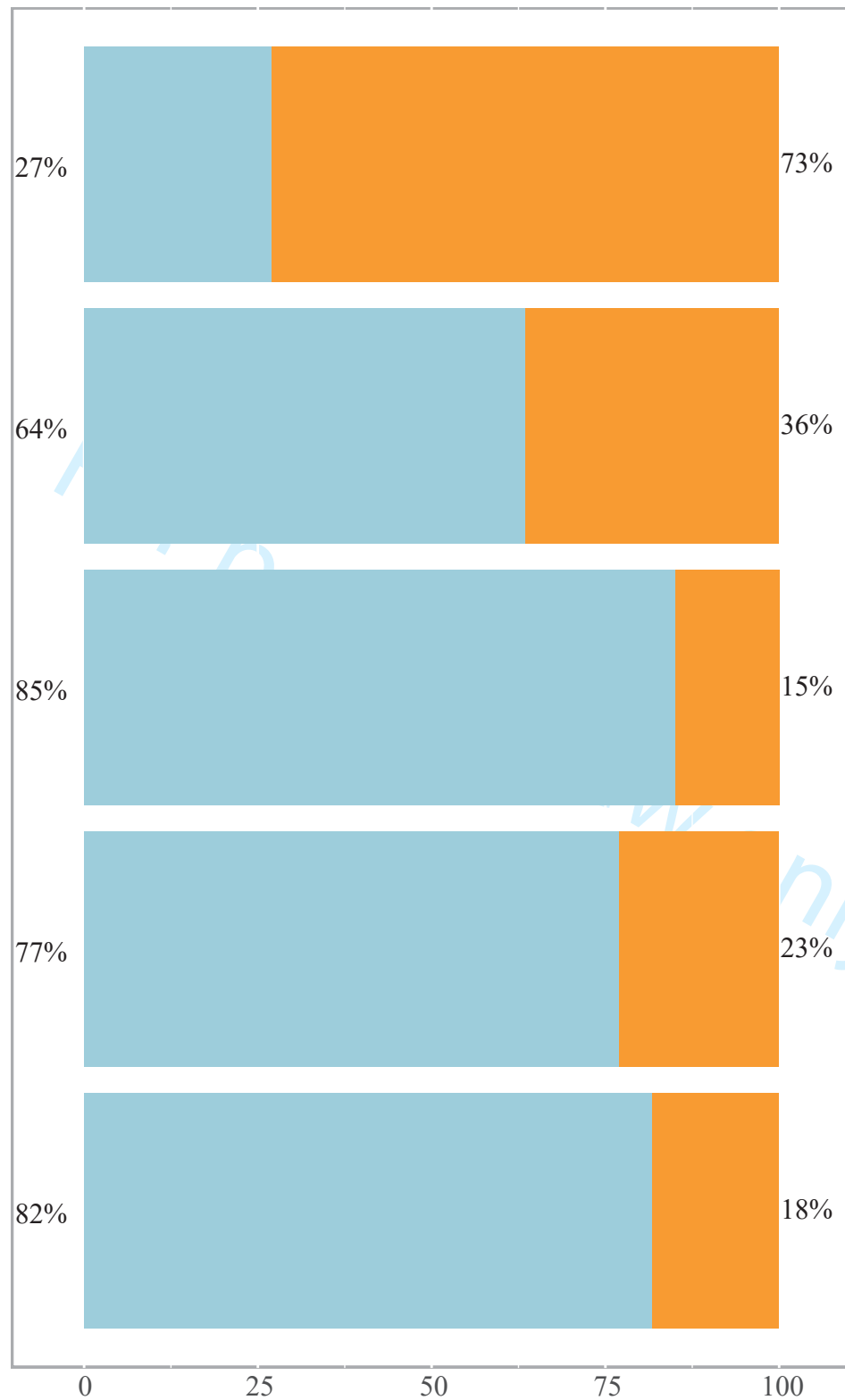
\*I know the content of big data platform

\*I know the potential applications of big data technology

\*Big data technology is practical in medical education

\*Big data technology is practical in clinical practice

\*Big data technology is practical in scientific research



Percentage

Percentage

Yes No

A

BMJ Open

B

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

\*Big data platform could assist future medical education

Big data platform could assist future medical reserch

\*Big data platform could assist future clinical practice

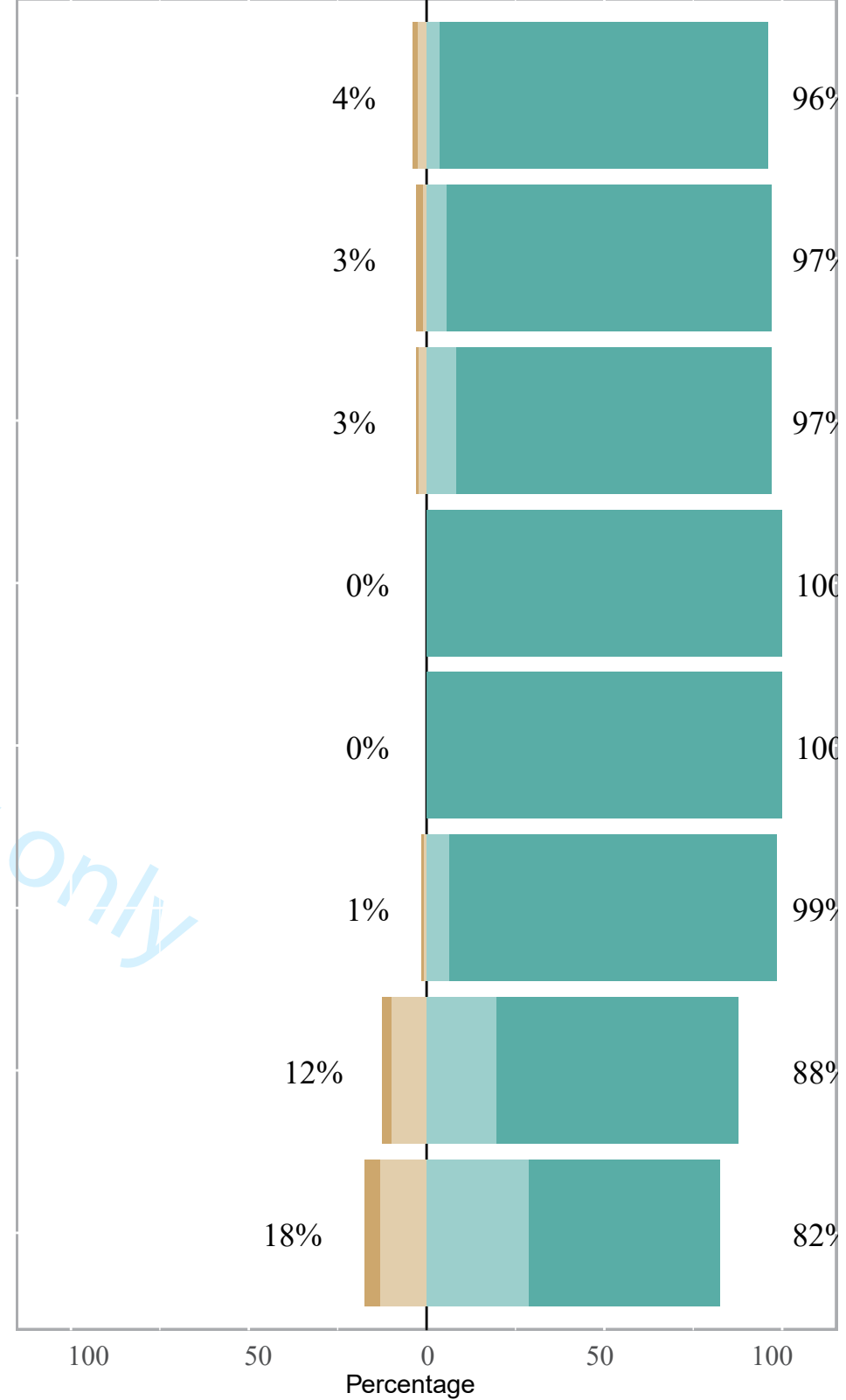
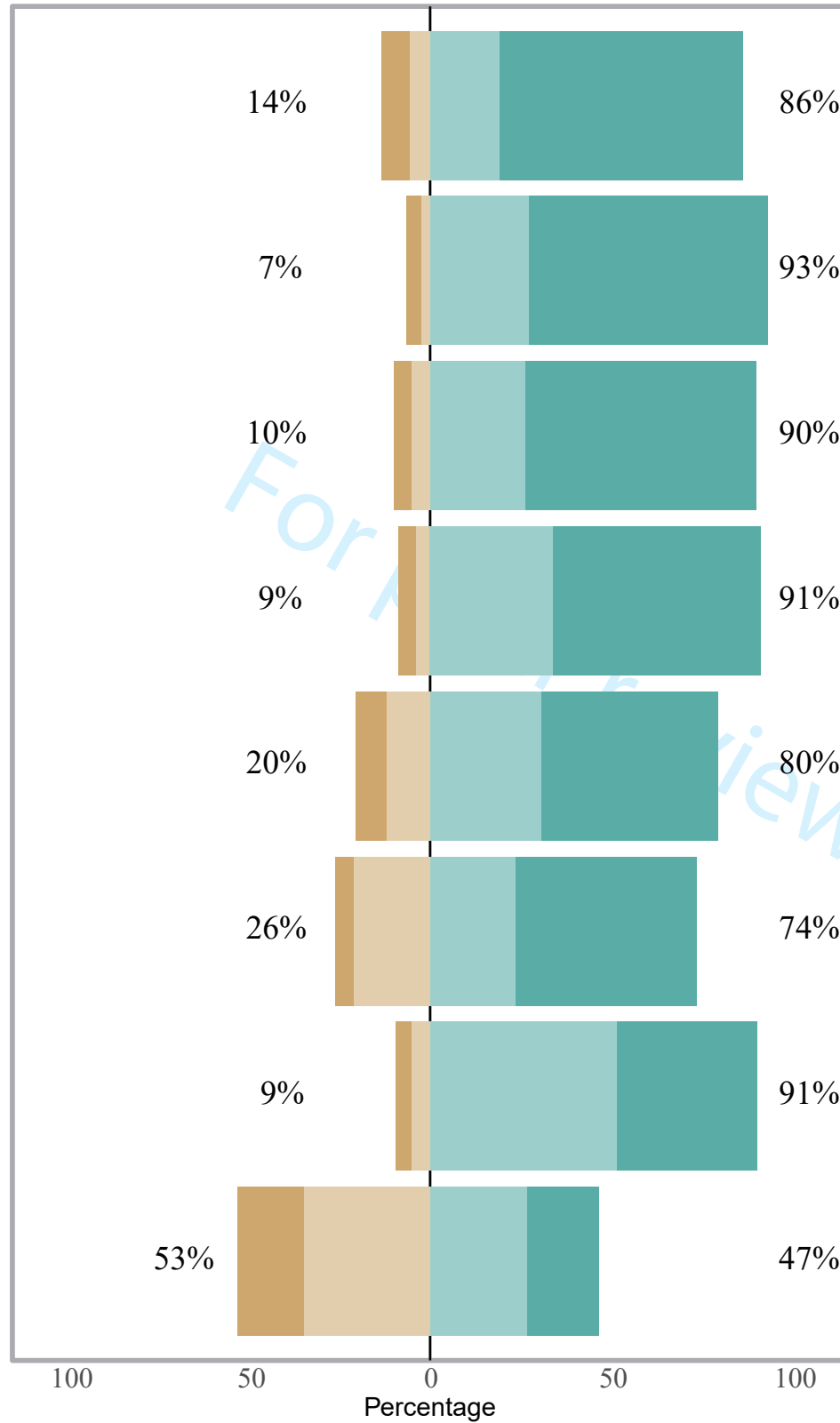
\*I am willing to learn how to use big data platform

\*I am willing to use big data platform in the future

\*I am willing to recommanded big data platform to my colleges

Big data platform could benefit my career

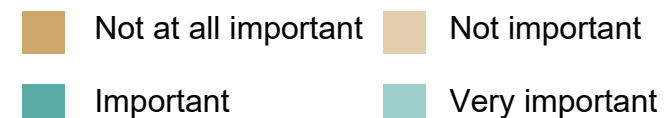
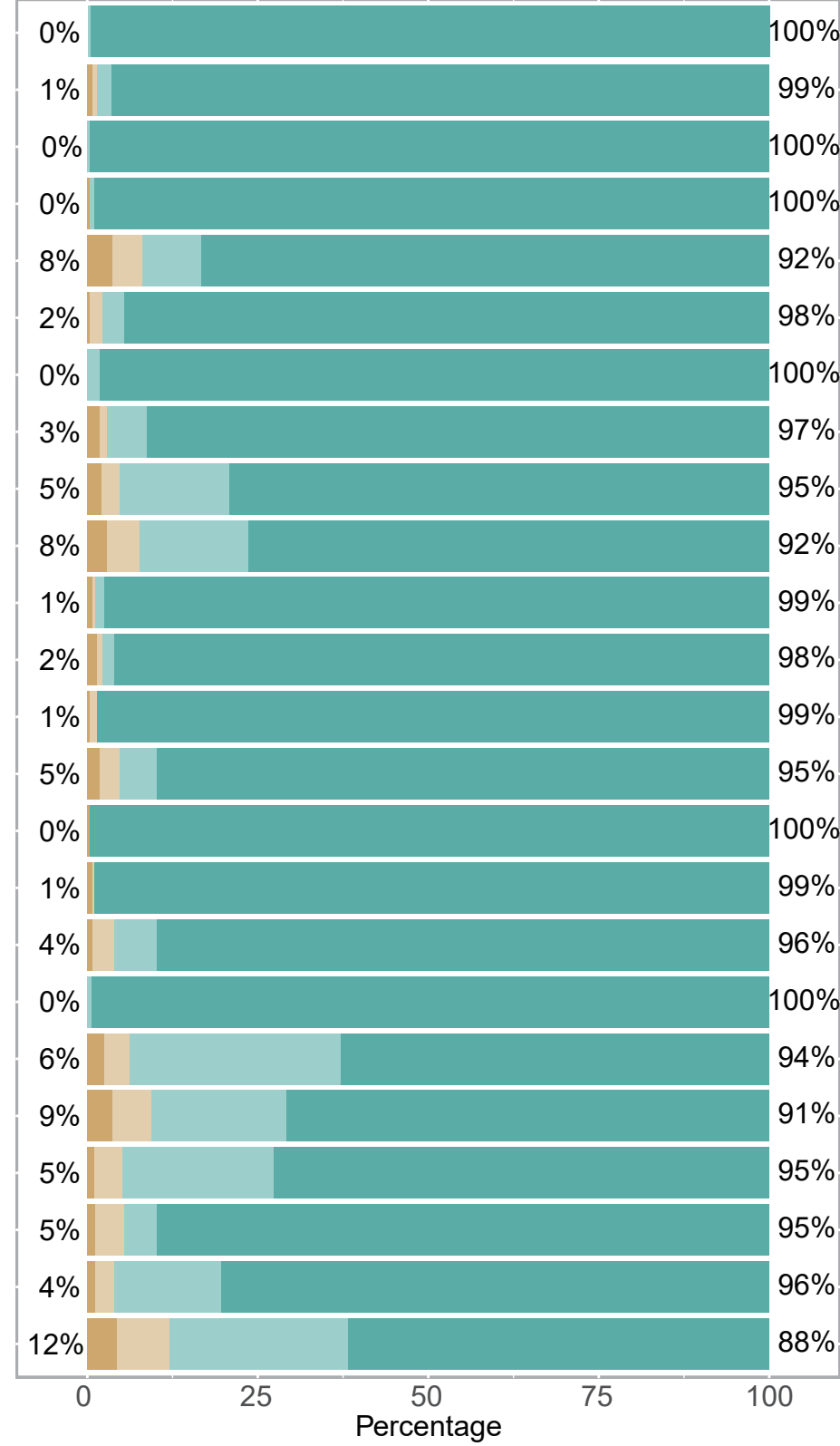
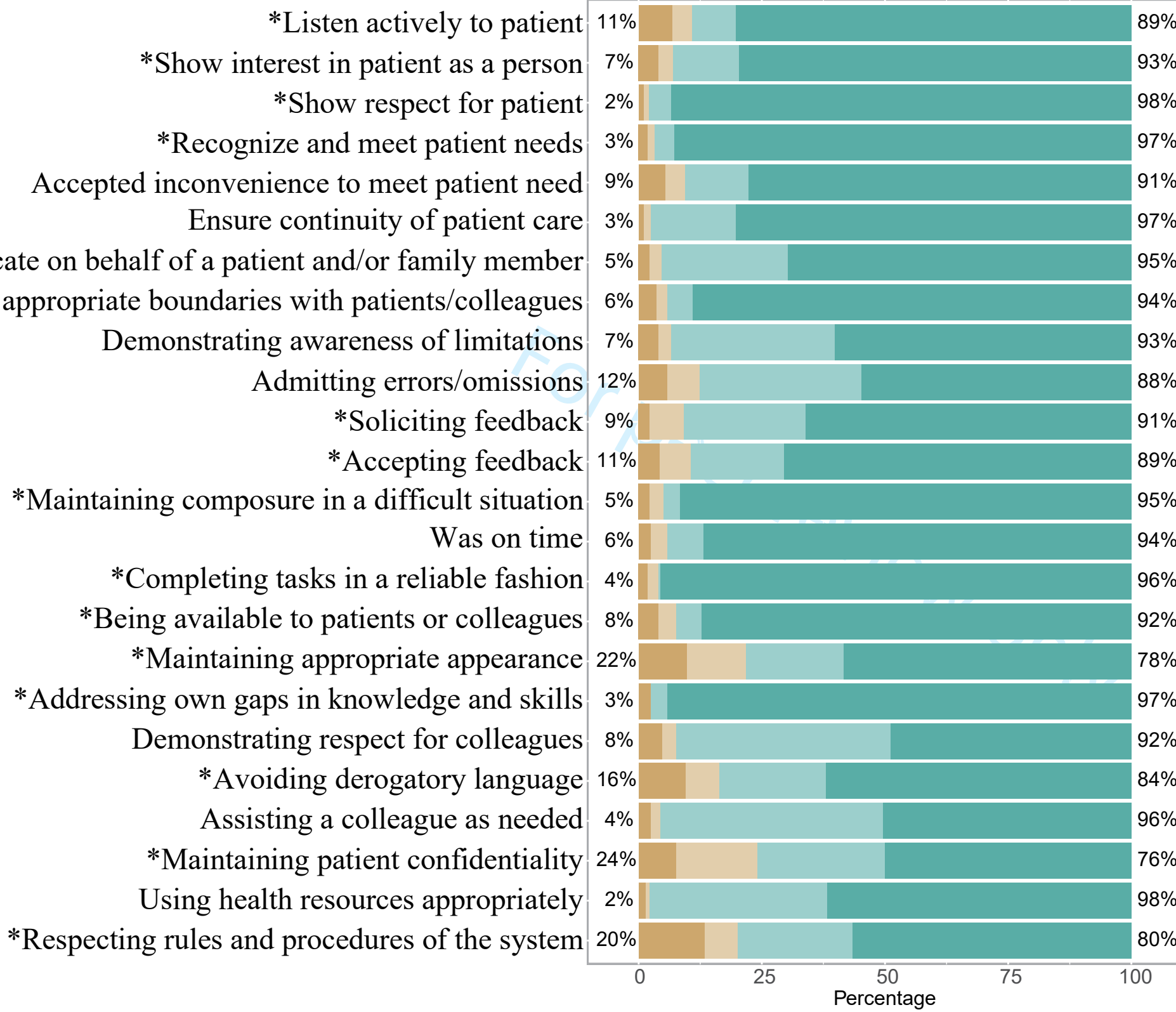
\*Big data platform could benefit all medical careers



Totally disagree Disagree

Totally agree Agree

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53



## Supplemental tables

Table S1, basic knowledge for big data platform before and after the workshop.

Items	Before Workshop		After Workshop		<i>P</i>
	YES, n	NO, n	YES, n	NO, n	
I know the content of big data platform	74	200	253	21	<0.0001
I know the potential applications of big data technology	174	100	253	21	<0.0001
Big data technology is practical in medical education	233	41	271	3	<0.0001
Big data technology is practical in clinical practice	211	63	238	36	0.0039
Big data technology is practical in scientific research	224	50	261	13	<0.0001

Table S2, students' attitude towards big data platform before and after the workshop.

Items	Before Workshop						After Workshop						<i>P</i>
	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	Mean	SD	Totally disagree, n	Disagree, n	Agree, n	Totally agree, n	Mean	SD	
Big data platform could assist future medical education	21	16	53	184	3.50	0.910	4	7	10	253	3.87	0.504	0.0002
Big data platform could assist future scientific research	11	7	75	181	3.55	0.735	5	3	16	250	3.86	0.499	0.0705
Big data platform could assist future clinical practice	13	14	72	175	3.49	0.799	2	6	23	243	3.85	0.464	0.0017
I am willing to learn how to use big data platform	13	11	93	157	3.43	0.783	0	0	0	274	4	0	<0.0001

I am willing to use big data platform in the future	23	33	84	134	3.20	0.953	0	0	0	274	4	0	<0.0001
I am willing to recommended big data platform to my colleagues	14	58	65	137	3.19	0.941	2	2	18	252	3.90	0.388	<0.0001
Big data platform could benefit my career	12	14	142	106	3.25	0.744	7	27	54	186	3.53	0.776	0.3382
Big data platform could benefit all medical careers	50	96	74	54	2.48	1.006	12	36	79	147	3.32	0.863	<0.0001

SD, Standard Deviation.

Table S3, the professionalism for students before and after the workshop.

Items	Before workshop				After workshop				P
	Not at all important, n	Not important, n	Important, n	Very important, n	Not at all important, n	Not important, n	Important, n	Very important, n	
Doctor-patient relationship skills									
Listen actively to patient	19	11	24	220	0	0	1	273	<0.0001
Show interest in patient as a person	11	8	37	218	2	2	6	264	0.0029
Show respect for patient	3	3	12	256	0	0	1	273	0.0401
Recognize and meet patient needs	5	4	11	254	1	0	2	271	0.0255
Accepted inconvenience to meet patient need	15	11	35	213	10	12	24	228	0.6503
Ensure continuity of patient care	3	4	47	220	1	5	9	259	1

1	Advocate on behalf of a patient and/or family member	6	7	70	191	0	0	5	269	0.0008
2	Maintained appropriate boundaries with									
3	patients/colleagues	10	6	14	244	5	3	16	250	0.1440
4										
5										
6	Reflective skills									
7										
8	Demonstrating awareness of limitations	11	7	91	165	6	7	44	217	0.4595
9										
10	Admitting errors/omissions	16	18	90	150	8	13	44	209	0.0880
11										
12	Soliciting feedback	6	19	68	181	2	1	4	267	<0.000
13										<i>1</i>
14	Accepting feedback	12	17	52	193	4	2	5	263	0.0001
15										
16	Maintaining composure in a difficult situation	6	8	9	251	1	3	0	270	0.0310
17										
18	Time management									
19										
20	Was on time	7	9	20	238	5	8	15	246	0.7027
21										
22	Completing tasks in a reliable fashion	5	6	1	262	1	0	0	273	0.0086
23										
24	Being available to patients or colleagues	11	10	14	239	2	1	0	271	0.0004
25										
26	Interprofessional relationship skills									
27										
28	Maintaining appropriate appearance	27	33	54	160	2	9	17	246	<0.000
29										<i>1</i>
30	Addressing own gaps in knowledge and skills	7	0	9	258	0	0	2	272	0.0225
31										
32	Demonstrating respect for colleagues	13	8	119	134	7	10	85	172	0.6139
33										
34	Avoiding derogatory language	26	19	59	170	10	16	54	194	0.0220
35										
36	Assisting a colleague as needed	7	5	124	138	3	11	61	199	0.8407
37										
38										
39										
40										
41										
42										
43										
44										
45										
46										

1	Maintaining patient confidentiality	21	45	71	137	3	12	13	246	<i>&lt;0.000</i>
2										<i>1</i>
3										
4	Using health resources appropriately	4	2	99	169	3	8	43	220	<i>0.3244</i>
5										
6	Respecting rules and procedures of the system	37	18	64	155	12	21	72	169	<i>0.0146</i>
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										
39										
40										
41										
42										
43										
44										
45										
46										

For peer review only



## STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	5
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	7-8
		(b) Describe any methods used to examine subgroups and interactions	NA
		(c) Explain how missing data were addressed	7-8
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	7-8
		(e) Describe any sensitivity analyses	NA

Continued on next page

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

<b>Results</b>			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	All participants were eligible and included in this study.
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8-9
		(b) Indicate number of participants with missing data for each variable of interest	Very high response rates with no missing data, as shown in supplementary data.
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	NA
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	NA
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	8
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	In all tables and figures.
		(b) Report category boundaries when continuous variables were categorized	In all tables and figures.
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16
Generalisability	21	Discuss the generalisability (external validity) of the study results	15
<b>Other information</b>			

1  
2 Funding 22 Give the source of funding and the role of the funders for the present study 16  
3 and, if applicable, for the original study on which the present article is  
4 based  
5

6  
7 \*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and  
8 unexposed groups in cohort and cross-sectional studies.  
9

10  
11 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and  
12 published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely  
13 available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at  
14 <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is  
15 available at [www.strobe-statement.org](http://www.strobe-statement.org).  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60