Journal of Prosthodontic Research

Official Journal of Japan Prosthodontic Society

Original Article

Accuracy of online symptom checkers for diagnosis of orofacial pain and oral medicine disease

Yuya Yoshida^{a,b}, Glenn Thomas Clark^{b*}

^aDepartment of Prosthodontics, School of Dentistry, Showa University, Tokyo

^bOrofacial Pain and Oral Medicine Center and Distance Learning Office, Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, USA

Abstract

Purpose: The aim of this study was to compare and contrast the diagnostic accuracy of multiple online symptom checkers when used for the diagnosis of orofacial pain and oral medicine related disease vignettes. The comparison condition used in this study was the diagnostic accuracy achieved by advanced specialty residents on these same vignettes using a virtual patient system.

Methods: 27 individual disease vignettes were utilized. These vignettes had a variety of orofacial pain and oral medicine diseases. Post graduate orofacial pain and oral medicine residents at our University of Southern California interacted with their randomly assigned virtual patients were analyzed [n=574]. Virtual patient accuracy was based on whether the user selected the primary diagnosis as one of their top four choices after interviewing. Eleven English-language symptom checkers accuracy was based on whether the vignettes produced the primary diagnosis as one of their top four choices. Using these data, symptom checker and virtual patient accuracy rates were calculated.

Results: The primary diagnosis on virtual patient encounters was found within the top four choices a mean of 67.2% of the time. The primary diagnosis for the same vignettes entered into the 11 symptom checkers was found within the top four choices a mean of 5.9% of the time.

Conclusions: The accuracy of currently available symptom checkers that patient might use for self-diagnosis of common orofacial pain and oral medicine diseases was low, this result suggest that the improved diagnostic algorithms are needed.

Keywords: Symptom checker, Virtual patients, Diagnosis

Received date: 25 September 2019 , Accepted date: 29 April 2020, J-STAGE Advance published date: 9 September 2020

1. Introduction

In the United States, pain in the orofacial regions affects 21.7% of the population and costs over \$32 billion each year [1]. Orofacial pain patients often present having seen a number of previous clinicians (a mean of 5.3) and report having the pain for many years (mean of 4.2 years) prior to seeing an orofacial pain specialist [2]. Recently, more and more people use the internet to research their health concerns and performing “online self-diagnosis”, is increasingly prevalent, with 35% of US adults having attempted to diagnose their own symptoms online [3, 4]. There are a variety of web-based symptom checker programs available to patients. Symptom checkers are software tools that allow users to submit a set of symptoms and receive advice related to them in the form of a diagnosis list, health information or triage [5]. When they

work well, symptom checkers help educate patients on a range of diagnoses that might fit their symptoms.

Unfortunately, there is little evidence of their effectiveness of these programs at helping a patient achieve a reasonable tentative diagnosis in the orofacial pain and oral medicine disease arenas. The aim of this study was to determine the diagnostic accuracy of 11 online symptom checkers using a set of 14 orofacial pain patient vignettes and 12 oral medicine disease related patient vignettes. The basic method we used involved entering relevant symptom data derived from our patient vignettes into the symptom checkers and then assessing the diagnostic accuracy achieved. We compared the symptom checker derived diagnoses with the diagnoses achieved on these same 26 patient vignettes by residents users of a virtual patient diagnostic simulation system [6]. The residents were all enrolled in an orofacial pain and oral medicine program and each resident encountered 4 test patients that were randomly assigned. The gold standard or true diagnosis was the diagnoses assigned apriori to each patient vignette by an expert faculty panel.

* Corresponding author at: Orofacial Pain and Oral Medicine Center and Distance Learning Office, Herman Ostrow School of Dentistry, University of Southern California, Los Angeles. 925 West 34th Street Los Angeles, CA 90089-0641, USA.

E-mail address: gtc@usc.edu (G.T. Clark)

2. Material and methods

2.1. Subjects

The two subject groups in this study were a group of postgraduate orofacial pain and oral medicine residents (group 1) and a group of eleven online English-language symptom checkers (group 2). In group 1 there were 93 residents (61 males and 32 females; mean age 46.00 ± 8.91 ; range 32–67) were all enrolled and had completed at least 1 year of study in the area of orofacial pain and oral medicine at the Herman Ostrow School of Dentistry of USC. In group 2 we identified a selection of 11 online publicly available, English-language symptoms checkers (Table 1). The search began by identifying symptom checkers that were available as apps in the Apple app store and Google Play using two search phrases (“symptom checker”, “medical diagnosis”). We examined the first 240 search results by hand to find eligible online symptom checkers. We then entered the same two search phrases in Google and examined the first 300 results. These cut-off points for our search was used because the probability of relevant search results identified using Google declines substantially after the first 300 results [7]. We supplemented our searches by asking the developers of two symptom checkers if they knew of other competing products. In total we identified 143 symptom checkers and we then excluded 132 that used the same medical content and logic as another tool (and therefore would have identical performance). We excluded symptom checkers if they only focused on a single class of illness (for example, orthopedic problems), only provided medical advice (for example, what symptoms are typically associated with a certain condition) or were not working when we accessed them. After these exclusions, we evaluated 11 symptom checkers. This research met the requirements outlined in 45 CFR 46.101(b)(4) and qualified for exemption from Institutional Review Board review. This exemption was approved by the University Park Institutional Review Board of the University of Southern California (USC UPIRB # UP-11-00292).

2.2. Virtual patients

We selected 26 patients vignettes that we had previously entered into our proprietary a single player networked “serious game” simulation system. This virtual patient system that allows students (users) to interact with a set of standardized virtual patients and improve their skills in interviewing patients and making clinical decisions using patient cases with a variety of orofacial pain and oral medicine problems. The manufacture of the virtual patient system of the diagnosis part is done by experts in the field of orofacial pain and intraoral medicine or experts qualified to the board.

2.3. Patient cases

We had 31 patient vignettes that were present on our virtual patient system (Fig. 1) and we selected cases only cases that our residents interacted with. We excluded 5 of these virtual patient case vignettes from our data because there were too few resident-virtual patient encounters to analyze. The clinical data extracted from the remaining 26 cases were input into the various symptom checkers described previously.

2.4. Practice on virtual patient system

Before a resident interacted with a virtual patient, they were instructed about how to use the virtual patient system and practiced on six training cases. The encounter had the residents ask medical interview questions as needed, then to ask physical examination questions as needed, and finally, users selected their final choices for the case. There was a countdown clock and all encounters had to be completed within thirty minutes. All participants were provided with a one-page guide sheet that outlined the

common elements of a medical interview and physical examination of the head and neck region. Following training, the residents were given at least 4 test cases that were presented in a random order. Only those cases that residents interacted with in a test mode were analyzed for our study. The total number of resident-virtual patient encounters was 574 and the mean resident-encounters per virtual patient case was 6.17 ± 2.48 (Fig. 2).

2.5. Outcome

Depending on the virtual patient vignette, the patient case had between 1 and 4 diagnoses. When multiple diagnoses were present, our expert faculty reviewed all diagnoses and all answers in the patient script and selected one of the diagnoses as the primary or most important diagnosis for our analysis. If either the resident or the symptom checkers achieved the primary diagnosis as one of the top four choices and was counted as a correct diagnosis result. Correct diagnosis answer rates were then calculated for residents (Group 1) and for the symptoms checkers (Group 2). We analyzed the results from 14 orofacial pain cases from 12 oral medicine cases, separately. Moreover, because the diagnosis choices within the virtual patient system were relatively specific (see Table 2) we calculated the accuracy of the symptom checker in two ways. One accuracy rate was determined when we accepted the terms “TMJD” (temporomandibular joint dysfunction) or “Oral Cancer” as correct. A second accuracy rate was calculated if considered the broad and non-specific diagnosis of “TMJD” or “Oral Cancer” as incorrect.

2.6. Statistics

Diagnosis accuracy scores were calculated and groups were compared for differences using Kruskal-Wallis / Steel-Dwass tests, with 5% set as the level of significance.

3. Results

3.1. Question number

The mean number of interview questions asked by the residents on their 4 test cases was 20.2 ± 5.1 . The mean number of physical examination questions asked by the residents was 14.3 ± 3.8 . In general, the questions asked by the symptom checkers were 14.6 ± 5.7 (Table-3) and did not include physical examination findings. One symptom checker (MEDoctor) asked over 50 questions, but in contrast some symptom checkers (Esagil, Helthline) asked only 1 or 2 questions of our patients vignettes.

3.2. Each symptom checkers accuracy

Table 4 shows the accuracy of each symptom checkers.

3.3. First choice accuracy of diagnosis in orofacial pain and oral medicine cases

A correct diagnostic answer was considered present, if the primary diagnosis was the top choice. Calculations were produced for orofacial pain and for oral medicine cases separately. The residents provided the correct diagnosis as the top choice in 65.1% (95% confidence interval 28.6% to 100.0%) and 69.5% (95% confidence interval 44.4 % to 88.9%) respectively. For the symptom checker, if we accepted the terms “TMJD” or “Oral Cancer” as correct, then the correct diagnosis for orofacial pain and oral medicine cases was found first in 34.4% (95% confidence interval 0.0% to 72.7%) and 6.8% (95% confidence interval 0.0% to 72.7%) respectively. If we did not accept the terms “TMJD” or “Oral Cancer” as correct, the 10 symptom checkers provided the correct diagnosis first in 9.1% (95% confidence interval 0.0% to 45.5%) and 1.5% (95% confidence

Table 1. Symptom checkers included in this study.

	Symptom checkers	Country
1	Esagil	USA
2	FreeMD	USA
3	Helthline	USA
4	Isabel	UK
5	Mayo clinic	USA
6	MEDdoctor	USA
7	Symcat	USA
8	Symptify	USA
9	Symptomate	Poland
10	WebMD	USA
11	ADA	USA

Fig. 1. Example of Virtual Patient. **Fig. 2.** Timeline in this study. VP: Virtual patient, TMJD: temporomandibular joint dysfunction.

interval 0.0% to 9.1%) respectively.

3.4. Top-four accuracy of diagnoses in all cases combined

A correct answer was considered present if the primary diagnosis on virtual patient encounters was found within the top four choices. For the resident encounter with virtual patients, this happened a mean of 67.2% of the time (95% confidence interval 28.5% to 100.0%). The primary diagnosis for the same vignettes entered into the 11 SC was found within the top four choices a mean of 5.9% of the time (95% confidence interval 0.0% to 45.5%).

3.5. Top-four accuracy of diagnosis in OFP and OM cases

A correct answer was considered present if the primary diagnosis with the top four choices and calculations were produced for orofacial pain and for oral medicine cases separately. Residents listed the correct diagnosis

Table 2. Virtual patient cases in this study.

Orofacial pain cases	Oral Medicine cases
Acute trismus	Benign mucous membrane pemphigoid
Burning mouth syndrome	BRONJ
Capsulitis of TMJ	Candidiasis
Chronic daily headache (Migraine)	Cement-osseous dysplasia
Disc displacement no reduction (Closed Locking)	Epithelial dysplasia
Disc displacement with reduction (Clicking)	Frictional keratosis
Disc displacement with reduction (Episodic Locking)	Lichen planus
Dental infection	Lymphoma
Dysguesia (taste distortion)	Nicotinic stomatitis
Episodic cluster headache	Oral lichenoid reaction
Episodic migraine	Oral Squamous cell carcinoma
Generalized osteoarthritis	Paraneoplastic pemphigus
Localized myalgia (Jaw)	Proliferative verrucous leukoplakia
Localized osteoarthritis	Recurrent/Metastatic carcinoma
Medication overuse headache	
Muscle contracture	
Regional myofascial pain	
Trigeminal neuropathy	

Table 3. Total number of questions. The mean number of questions for all cases combined and for orofacial pain and oral medicine cases separately are shown.

	All cases (n=26)	OFP cases (n=14)	OMed cases (n=12)
Esagil	1.30	1.00	1.43
FreeMD	15.00	15.00	15.00
Helthline	2.00	2.00	2.00
Isabel	6.88	6.86	6.92
Mayo clinic	11.67	11.67	-----
MEDdoctor	61.09	67.31	53.00
Symcat	22.48	23.21	21.33
Symptify	22.60	25.07	19.45
Symptomate	11.00	10.64	11.42
WebMD	3.32	3.46	3.11
ADA	3.55	3.25	3.90
Average	14.63	15.41	13.76

within the top four diagnosis in 86.5% (95% confidence interval 55.6% to 100.0%) and 75.6% (95% confidence interval 44.4% to 100.0 %) of virtual patient evaluations. For the symptom checker, if we accepted the terms “TMJD” or “Oral Cancer” as correct, then the correct diagnosis for orofacial pain and oral medicine cases was found within the top 4 choice in 39.0% (95% confidence interval 0.0% to 72.7%) and 10.6% (95% confidence interval 0.0% to 63.6 %) respectively. If we did not accept the terms “TMJD” or “Oral Cancer” as correct, the 11 symptom checkers provided the correct diagnosis within the top four diagnosis 9.1% (95% confidence interval 0.0% to 45.5%) and a 5.3% (95% confidence interval 0.0% to 54.5 %) respectively.

4. Discussion

The data in this study shows that the residents correctly identified the diagnosis at a significantly higher rate than the symptoms checkers. The higher rate achieved by the residents is most likely because they asked more questions and could “mock-examine” the patient to confirm their diagnostic hypotheses generated by the interview process. The virtual patient system also allowed residents to order diagnostic tests (biopsies and

Table 4. Diagnostic accuracy for all symptom checkers. The accuracy rates were calculated for all cases combined and for orofacial pain and oral medicine cases separately in two ways: (#1) when the correct diagnosis was listed as the first or “primary” diagnosis and (#2) when it was listed anywhere among the first four diagnoses. A: Accuracy rate data when we accepted “TMJD” and “Oral Cancer” as a substitute diagnostic term for a more specific diagnosis on the orofacial pain patient vignettes and oral medicine patient vignettes. B: Accuracy rate data when we didn’t accept “TMJD” and “Oral Cancer” as a substitute diagnostic term is listed.

% Diagnosis	A. Accepted “TMJD” as a substitute diagnosis						B. Didn’t accept “TMJD” as a substitute diagnosis					
	All cases (n=26)		OFP cases (n=14)		OMed cases (n=12)		All cases (n=26)		OFP cases (n=14)		OMed cases (n=12)	
	Primary Diagnosis	Listed in Top 4	Primary Diagnosis	Listed in Top 4	Primary Diagnosis	Listed in Top 4	Primary Diagnosis	Listed in Top 4	Primary Diagnosis	Listed in Top 4	Primary Diagnosis	Listed in Top 4
Esagil	0.00	3.85	0.00	0.00	0.00	8.33	0.00	0.00	0.00	0.00	0.00	0.00
FreeMD	23.08	26.92	42.86	42.86	0.00	8.33	3.85	7.69	7.14	7.14	0.00	8.33
Helthline	38.46	42.31	64.29	71.43	8.33	8.33	7.69	7.69	7.14	7.14	8.33	8.33
Isabel	23.08	30.77	35.71	42.86	8.33	16.67	11.54	19.23	14.29	21.43	8.33	16.67
Mayo clinic	7.69	7.69	14.29	14.29	0.00	0.00	7.69	7.69	14.29	14.29	0.00	0.00
MEDoctor	19.23	19.23	35.71	35.71	0.00	0.00	7.69	7.69	14.29	14.29	0.00	0.00
Symcat	30.77	38.46	57.14	64.29	0.00	8.33	0.00	3.85	0.00	0.00	0.00	8.33
Symptify	26.92	30.77	50.00	57.14	0.00	0.00	7.69	7.69	14.29	14.29	0.00	0.00
Symptomate	11.54	11.54	21.43	21.43	0.00	0.00	3.85	3.85	7.14	7.14	0.00	0.00
WebMD	30.77	38.46	57.14	64.29	0.00	8.33	11.54	11.54	21.43	21.43	0.00	0.00
ADA	26.92	34.62	50.00	64.29	0.00	0.00	3.85	3.85	7.14	7.14	0.00	0.00
Average	21.68	25.87	38.96	43.51	1.52	5.30	5.94	7.34	9.74	10.39	1.52	3.79

Table 5. Accuracy of diagnosis decisions. This analysis showed that the residents correctly identified the diagnosis at a significantly higher rate than the symptoms checkers for both the primary diagnosis analysis and the “top 4” analysis. This discrepancy was more evident if the non-specific diagnosis of “TMJD” as a substitute term for a more specific diagnosis was rejected. The above finding did not change when we looked at the orofacial pain cases and the oral medicine cases separately.

% Diagnosis	A. All cases(26 cases)		B. Orofacial pain cases (14 cases)		C. Oral medicine cases (12 cases)	
	Primary diagnosis	Listed in Top 4	Primary diagnosis	Listed in Top 4	Primary diagnosis	Listed in Top 4
Resident in Orofacial pain and oral medicine	70.00 (28.57-100.00)	85.71 (44.44-100.00)	65.53 (28.57-100.00)	91.42 (55.56-100.00)	71.42 (44.44-88.89)	80.85 (0.00-63.63)
Symptom Checker TMJD/Oral Cancer accepted	9.09 (0.00-72.72)	27.27 (0.00-72.72)	34.41 (0.00-72.72)	45.45 (0.00-72.72)	0.00 (0.00-63.63)	0.00 (0.00-63.63)
Symptom Checker TMJD/Oral Cancer not accepted	0.00 (0.00-45.45)	0.00 (0.00-54.54)	0.00 (0.00-45.45)	0.00 (0.00-45.45)	0.00 (0.00-9.09)	0.00 (0.00-54.54)
p-value	<.0001		<.0001		<.0001

Fig. 3. Accuracy of diagnosis decisions. Comparative statistical analysis for all 26 cases (A) and for the subset of orofacial pain (B) and oral medicine (C) cases. *p < 0.05. Data are expressed as median (min-max). Kruskal-Wallis/Steel-Dwass tests were used.

CBCT images) and get report results from these tests before rendering a diagnosis. Because symptom checkers did not usually include and physical examination findings or diagnostic test results a lower level of accuracy would be expected. Moreover, the average number of questions asked by the symptom checkers were lower by almost half. The diagnoses

rendered by the various symptom checkers for both orofacial pain and oral medicine were also not very specific in that they used broad categories such as oral cancer or temporomandibular joint dysfunction.

4.1. Comparisons with other studies

It is interesting to note that a disease-focused symptom checker listed the correct diagnosis 70% of the time for ear, nose, and throat symptoms [8]. Another study focused symptom checker accuracy on older adult subjects and used only two diseases for the patient vignettes (mononucleosis and scarlet fever) to evaluate WebMed reported a diagnostic accuracy rate of 50% [3]. In our study which included multiple aged patient with a wider range of diseases, WebMed accuracy was 30.77% and only when we accepted the non-specific term TMJD, as a correct diagnosis.

4.2. Limitation

The low diagnostic accuracy results, even when we accepted the generic term oral cancer, for our vignettes dealing with oral mucosal and osseous diseases is not surprising. The majority of the symptom checkers only asked a small number of questions and often patients with an oral lesion are not cognizant of their symptoms as they are often painless tissue changes. A good portion of the diagnostic work up in oral mucosal and osseous disease requires that the clinician see a radiograph or inspect the lesion visually and then get biopsy results to establish the diagnosis. Clearly, this is a limitation of patient reported symptoms and symptom checkers. Some symptom checkers asked a larger number of questions. Other checkers asked only one or two questions and ask no questions

about the patient's history of present illness or current and past illnesses.

5. Conclusion

Currently available symptom checkers do not deal very well with TMD, Orofacial Pain problems or mucosal diseases and need improvement.

Conflict of interest statement

The authors declare no conflicts of interest associated with this manuscript.

Acknowledgements

The authors are grateful to A. Vistoso, L. Nocera for technical assistance and helpful advice. We also thank all members of Department of Prosthodontics, Showa University for encouragement and kind support. This study was supported by Showa University Research Grant for Young Researchers.

References

- [1] Horst OV, Cunha-Cruz J, Zhou L, Manning W, Mancini L, DeRouen TA. Prevalence of pain in the orofacial regions in patients visiting general dentists in the Northwest Practice-based REsearch Collaborative in Evidence-based DENTistry research network. *J Am Dent Assoc* 2015;146(1943-4723 (Electronic)).
- [2] Rosenbaum RS, Fau FJ, Okeson JP. Orofacial pain emerging as a dental specialty. *J Mass Dent Soc* 2001;49(0025-4800 (Print)).
- [3] Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: Results from a scenario-based think-aloud protocol. *J Med Internet Res* 2014;16(1).
- [4] Fox S, Duggan M. Internet and American life project. Pew Research Center and California Health Care Foundation. *Health Online* 2013 2013;4.
- [5] Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015(1756-1833 (Electronic)).
- [6] Clark GT, Suri A, Fau - Enciso R, Enciso R. Autonomous virtual patients in dentistry: system accuracy and expert versus novice comparison. *J Dent Educ* 2012;76(1930-7837 (Electronic)).
- [7] Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Soc Sci Med* 2015(1873-5347 (Electronic)).
- [8] Farmer SEJ, Bernardotto M, Singh V. How good is Internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? *Clin Otolaryngol* 2011;36(5):517-8.