Original Article

Hong Kong Journal of Emergency Medicine
2020, Vol. 27(4) 217–222

DOI: 10.1177/1024907919842486
journals.sagepub.com/home/hkj
SAGE

Triage accuracy of online symptom checkers for Accident and Emergency Department patients

Stephanie Wing Yin Yu[1], Andre Ma[1], Vivian Hiu Man Tsang[1], Lulu Suet Wing Chung[1], Siu-Chung Leung[2] and Ling-Pong Leung[3]

Abstract

Background: Overutilisation of the Accident and Emergency Department is an increasingly serious healthcare challenge. Online symptom checkers could help alleviate this challenge by allowing patients to self-triage before visiting the Accident and Emergency Department.

Objectives: This study aimed to assess the triage accuracy of online symptom checkers, which would help determine the potential roles of symptom checkers in an Accident and Emergency Department setting.

Methods: A total of 100 random Accident and Emergency Department records were sampled from the Queen Mary Hospital in Hong Kong. The inclusion criteria were patients over the age of 18 attending the Queen Mary Hospital Accident and Emergency Department in 2016. Symptom checkers by Drugs.com and FamilyDoctor were selected as representative tools. One triage recommendation was generated by each symptom checker for each case record. Each symptom checker's triage accuracy was then evaluated using a few outcome measures: overall sensitivity, sensitivity for emergency cases and specificity for non-emergency cases, when compared with the triage categories assigned by the triage nurses.

Results: The results showed that Drugs.com had a higher overall triage accuracy than FamilyDoctor (74% and 50%, respectively), but both checkers are inadequately sensitive to emergency cases (70% and 45%, respectively) with low negative predictive values (43% and 24%, respectively).

Conclusion: In their current states, symptom checkers are not yet suitable as alternatives to Accident and Emergency Department triage protocols due to their low overall sensitivities and negative predictive values. However, symptom checkers might serve as useful Accident and Emergency Department adjuncts in other ways, such as to provide more information prior to a patient's arrival to streamline the triage and preparation process at the Accident and Emergency Department.

Keywords

Emergency service, hospital, medical overuse, online symptom checkers, triage

Introduction

The Internet is an easy source of health information, and many patients have been found to research their own discomforts online before seeking professional advice. Data have shown that almost three quarters of adults in the United States had searched for health information on the Internet over the past 12 months,[1] while more than one-fifth of adults in the United Kingdom have self-diagnosed through the Internet instead of visiting a healthcare professional.[2]

[1]Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR
[2]Accident and Emergency Department, Queen Mary Hospital, Pokfulam, Hong Kong SAR
[3]Emergency Medicine Unit, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR

Corresponding author:
Stephanie Wing Yin Yu, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 2/F, William MW Mong Block, 21 Sassoon Road, Pokfulam, Hong Kong SAR.
Email: yusteph@hku.hk

Generic keyword online searches yield too much information for digestible use.[1,3] More constructive and clinically relevant programs, known as symptom checkers, have been created with the aim of providing patients at home with differential diagnoses and triage advice based on self-reported symptoms. Various organisations, such as the National Health Service (NHS) and the Mayo Clinic, have launched their own symptom checkers, with the NHS symptom checker reporting up to 15 million visits per month.[4,5] Symptom checkers' advice has the power to influence patients' health-seeking behaviours and health outcomes, and these checkers' technology can be leveraged to alleviate various challenges in healthcare. One such healthcare challenge is the growing rate of inappropriate utilisation of Accident and Emergency Department (A&E) services. A study in the United States estimated that USD 38 billion is being wasted every year due to inappropriate A&E usage.[6] This overutilisation of emergency services leads to decreased overall quality of care for both urgent and non-urgent users of the A&E.[7]

Symptom checkers could be a method to alleviate the overcrowding problem at A&Es, by allowing patients unsure about the urgency of their condition to self-triage and instructing patients with non-urgent conditions to refrain from visiting the A&E. The idea of triaging ambivalent patients before they present at A&Es is not a novel concept – the NHS developed a telephone triage line to ensure that patients are receiving medical care in appropriate settings.[8] With continual technological development, symptom checkers could serve the same purpose at much lower financial and manpower requirements.

Before recommendations for the widespread adoption of symptom checkers can be made, it is imperative to thoroughly assess the reliability of these programs' triage advice. A number of studies have been conducted in recent years on the accuracy of online symptom checkers,[4,9–12] but to date no study has evaluated the accuracy of these symptom checkers with real-life cases and in an A&E setting, which was the literature gap that this study aimed to fill. The study's primary objective was to determine the accuracy of online symptom checkers in the triage of real emergency and non-emergency cases. Our results can contribute to the analysis of whether symptom checkers could serve as a major adjunct in triage protocols for patients unsure of their need for A&E care.

Methods

The study protocol was approved by the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster. Written informed consent was not necessary because no patient data have been included in the article.

Selection of symptom checkers

The inclusion criteria for symptom checker selection were free and publicly available programs that provide triage recommendations across all specialties for adult patients. A previous study found 15 such symptom checkers and ranked the checkers by their triage accuracies based on standardised clinical vignettes.[4] Among the top 10 most accurate checkers, HMS Family Health Guide (Harvard Medical School, USA), Steps2Care (Paramount Care Inc., USA), FreeMD (DSHI Systems, Inc., USA) and EarlyDoc (EarlyDoc, Netherlands) have since been discontinued, and Healthy Children (American Academy of Pediatrics, USA) was excluded due to its focus on paediatric cases. The remaining symptom checkers were Symptify (LLC 2013, USA), Symptomate (Infermedica, Inc., Poland), Doctor Diagnose (AppColliders, USA), Drugs.com (The Drugsite Trust, New Zealand) and FamilyDoctor (American Academy of Family Physicians, USA). The first three are mobile apps with 5000–10,000, 100,000–500,000 and 10,000–50,000 downloads, respectively, while Drugs.com and FamilyDoctor are websites that receive over 6 million and 330,000 unique visitors per month, respectively.[13,14] These five checkers were independently tested by three investigators on five sampled A&E cases. Drugs.com and FamilyDoctor were ultimately chosen as the representations of online symptom checkers because of their accuracy, popularity and suitability for our study design given the limited information documented on A&E charts.

Selection of patient records

Patient records were obtained from the Queen Mary Hospital of Hong Kong, whose 24-h A&E serves an average of 300–400 patients per day. The inclusion criteria were patients over 18 years old attending the A&E between 1 January and 31 December 2016. The minimum sample size required for the analysis, based on a stringent prevalence of emergency cases set at 50%, power of 80% and statistical significance of 0.05, was estimated to be 100 cases.[15] Figure 1 displays the flow of record sampling in this study.

Upon arrival at the A&E, every patient is assigned one of the five triage categories outlined by the Hong Kong triage guidelines: Category 1 (Cat 1) for 'critical cases', Category 2 (Cat 2) for 'emergency cases', Category 3 (Cat 3) for 'urgent cases', Category 4 (Cat 4) for 'semi-urgent cases' and Category 5 (Cat 5) for 'non-urgent cases'. In this study, for each of the 20 randomly selected days, one eligible patient chart from each triage category was randomly sampled for both checkers. For each A&E patient chart that was incomplete, illegible or whose chief complaint was unavailable on a specific checker, a replacement case was sampled from the same triage category on the same day for that checker only, or from a different randomly selected day

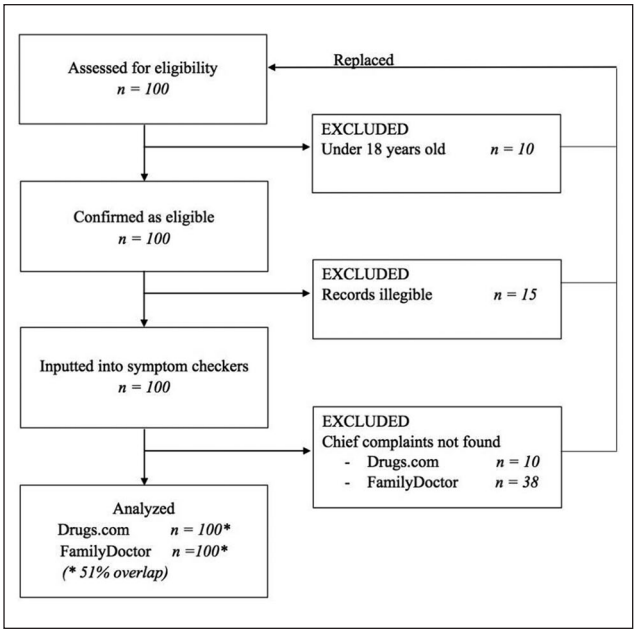

Figure 1. Flowchart of the study procedure.

if no more cases are eligible. Hence, in total, 100 cases from at least 20 different days in 2016, with 20 cases from each of the five triage categories, were inputted for analysis into each symptom checker, with some overlap in cases between the two checkers.

Assessment of symptom checkers

Each medical record was transcribed by one investigator into a standardised format containing the patient's age, gender, presenting complaints, vital signs, past medical history and physical examination findings, with the triage category blinded and recorded separately. Two investigators independently inputted information from the transcribed records into both symptom checkers and recorded the checkers' triage results. A third investigator independently resolved any disagreements between the original two investigators' results.

Outcome measures

For this study, emergency levels were defined as such: Cat 1–4 cases were classified as 'emergency' patients who appropriately visited the A&E, while Cat 5 cases were classified as 'non-emergency' patients who did not need A&E services. Triage category assigned by the A&E nurse was considered accurate in all cases.

Symptom checkers' triage advices are provided through direct instructions addressing the user. For this study, instructions that included one or more of the phrases 'visit the ER now', 'seek medical help immediately' or 'you likely have a life-threatening condition' were classified as 'emergency' advice, while instructions that included one or more of the phrases 'seek medical help today', 'make an appointment with a specialist', 'call your doctor to make an appointment' or 'self-care' were classified as 'non-emergency' advice. Two investigators independently classified each symptom checker's triage advice into 'emergency' and 'non-emergency', and a third investigator independently resolved any disagreements in classifications. Checkers' instructions that were similar but not exactly equal to the ones listed above were classified at the investigators' own discretion. Triage recommendations that required users to apply their own judgement to assess urgency, such as 'if you think the problem is serious, call your doctor right away', were universally considered incorrect, because these recommendations are not useful for self-triage purposes.

Data analysis

All analyses were performed using SPSS Statistics v22.0 (IBM Corp., USA). Descriptive statistics was used for demographic data, and quantitative data were reported as mean ± standard deviation (SD), or percentages with 95% confidence intervals (CIs), as appropriate. The percent agreement of each symptom checker with A&E triage nurse was calculated for all cases, emergency and non-emergency cases and cases of each triage category. Overall sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were also calculated. The intraclass correlation coefficients (ICCs) were computed for level of agreement between independent investigators, with ICC > 0.7 considered adequate by convention.

Results

Each symptom checker was tested with 100 A&E charts with an equal distribution between the five triage categories. A total of 51 charts were used to test both checkers, with two different sets of 49 charts used to test only one checker, making up a total of 149 charts sampled. Among the cases sampled for Drugs.com, the mean age was 56.6 years, with 42% being male. Among the cases sampled for FamilyDoctor, the mean age was 55.4 years, with 45% being male. There was no significant difference in the average age of sampled cases between the two checkers ($t = 0.57$, $p = 0.57$).

Table 1 displays the triage accuracies of the symptom checkers. Drugs.com was more accurate than FamilyDoctor (74% vs 50%) and had a lower under-triage rate. All ICCs for the independent assignments of emergency statuses exhibited adequate agreements (>0.7).

Table 2 displays the overall sensitivities, specificities, PPVs and NPVs of the two checkers in identifying urgent cases. Drugs.com performed better than FamilyDoctor in all four parameters.

Table 1. Triage accuracies of Drugs.com and FamilyDoctor.

	Drugs.com		FamilyDoctor	
	% (n = 100)	95% CI (%)	% (n = 100)	95% CI (%)
Accurately triaged[a]	74	64–82	50	40–60
Under-triaged[a]	24	16–34	45	35–55
Over-triaged[a]	2	0.24–7.0	5	1.6–11
Total (%)	100		100	
Cases by emergency status	Drugs.com results[b]		FamilyDoctor results[c]	
	Emergency	Non-emergency	Emergency	Non-emergency
Emergency (total n = 80)	56	24	35	45
Non-emergency (total n = 20)	2	18	5	15

CI: confidence interval; ICC: intraclass correlation coefficient.
[a]Emergency level of checker's result when compared to that of triage category documented on A&E chart.
[b]ICC (emergency status assignment) = 0.85 (95% CI: 0.80–0.89); ICC (recommendation assignment) = 0.73 (95% CI: 0.60–0.82).
[c]ICC (emergency status assignment) = 0.86 (95% CI: 0.82–0.90); ICC (recommendation assignment) = 0.75 (95% CI: 0.63–0.83).

Table 2. The overall sensitivities, specificities, PPVs and NPVs of Drugs.com and FamilyDoctor.

	Drugs.com		FamilyDoctor	
	%	95% CI	%	95% CI
Sensitivity	70	59–80	44	33–55
Specificity	90	68–99	75	51–91
PPV	97	88–100	88	73–96
NPV	43	28–59	25	15–38

PPV: positive predictive value; NPV: negative predictive value; CI: confidence interval.

Table 3 displays the triage accuracies of the two symptom checkers according to each triage category, with accurate being defined as assigning Cat 1–4 cases to 'emergency' or assigning Cat 5 cases to 'non-emergency'. Drugs.com was more accurate than FamilyDoctor in every category, and both symptom checkers performed better for non-emergency cases than the emergency ones.

Discussion

Both Drugs.com and FamilyDoctor performed suboptimally in overall triage accuracies (74% and 50%, respectively). According to the audit study performed by Semigran et al.[4] in 2015, the overall triage accuracies of Drugs.com and FamilyDoctor were 60% and 54%, respectively. The discrepancy between our data and Semigran's is likely due to our use of real patient records compared to their expert-written vignettes, since the two methodologies are otherwise largely similar. As opposed to hypothetical patient scenarios, our test samples represent a more diverse and realistic range of patient presentations and offer more insight into how symptom checkers perform in real clinical situations.

When considering whether or not symptom checkers' level of sensitivity and specificity suffices to justify incorporation into an A&E's triage protocol, one must consider the accuracy of the A&E's existing triage procedure. A review of the literature found that triage accuracies vary across centres. Studies have estimated an overall triage accuracy of 59.2% among three community hospitals in United Arab Emirates, Brazil and the United States,[16] 59.6% among four Swiss hospitals,[17] 82.9% in a single Brazilian centre[18] and 62.2% among the paediatric units for four Australian hospitals.[19] Even though online symptom checkers are theoretically less costly to implement than triage nurses, checkers must perform at least as well as the triage protocol in place before they can be considered as replacements to A&E nurses. Applying our data locally, where the A&E triage accuracy has been estimated to be 78%,[20] we conclude that online symptom checkers are not yet of sufficient quality to replace triage nurses in our locale.

On top of considering the checkers' overall accuracy, it is also important to analyse how the checkers fare for each triage category. In this study, both checkers were predictably most sensitive for cases in Cat 1 (critical cases) and Cat 5 (non-urgent cases), which represent two extremes of the urgency spectrum where the appropriate course of action is more obvious (Table 3). However, all cases in Cat 1–4 require A&E attention, and both Drugs.com and FamilyDoctor displayed generally low sensitivities for emergency cases (70% and 44%, respectively), especially when compared to their specificities for non-emergency cases (90% and 75%, respectively; Tables 1 and 2). This agrees with the observation that both checkers tended to under-triage rather than over-triage (Table 1), leading to a much higher PPV (97% for Drugs.com and 88% for FamilyDoctor) than NPV (43% for Drugs.com and 25% for FamilyDoctor; Table 2). One

Table 3. Triage accuracies of Drugs.com and FamilyDoctor according to triage category.

A&E triage category	Drugs.com		FamilyDoctor	
	Accuracy (%)	95% CI	Accuracy (%)	95% CI
Category 1	95	75–100	65	41–85
Category 2	65	41–85	35	15–59
Category 3	65	61–85	30	12–54
Category 4	55	32–77	50	27–73
Category 5	90	68–99	70	46–88

CI: confidence interval.

could argue that, for a symptom checker to safely serve as a major adjunct in the A&E, sensitivity to emergency cases is more important than to non-emergency cases, because under-triaging the former could lead to preventable deaths. In this study, Drugs.com under-triaged 30% and FamilyDoctor under-triaged 56% of the emergency cases (Table 1). These accuracies are sub-par compared both to local A&E nurses, who have been estimated to under-triage about 15% of A&E cases,[20] and to some telephone triage services, which have been found to under-triage about 19% of urgent cases.[21] Urgent patients mistriaged by symptom checkers would either not seek medical care or need to exercise their own layperson judgement, both of which leave room for otherwise avoidable morbidity and mortality. Therefore, we conclude that online symptom checkers, at least in their current states, are unsuitable to serve as the sole triage tool for potential A&E cases.

Since the replacement of triage nurses by automated programs was likely an overly aggressive goal that involves both practical and ethical issues, there are other, perhaps less grandiose, niches that symptom checkers could fill in our current A&E system. For example, symptom checkers could serve as an adjunct in prehospital care to optimise A&E resources. Patients, family members or ambulance staff can fill in the symptom checker questionnaire en route to the hospital – if this information can be connected to A&E staff, it could help cut down on the time needed to take history and make treatment-related preparations after the patient arrives. A rough guide into the urgencies of incoming cases can also allow A&E staff to better streamline their triage process and ensure that care is provided in the appropriate sequence.

Improvements must also be made to online symptom checkers before they can be widely applied to A&E settings. The low NPV of both checkers suggest that triage advice should be more aggressive to capture all potentially life-threatening conditions. In addition, incorporating regional and seasonal epidemiology and the past medical history of patients would allow a clearer analysis of the presenting symptoms. In the context of A&E cases, having offline programs and a search option for chief complaints would help the checkers be more user-friendly and provide timelier triage advice. Regardless of whether or not symptom checkers can aid or replace A&E triage nurses, given the increasing trend of laymen utilising the Internet for health-related information, more research should be conducted to improve symptom checkers' triage algorithms to generate more accurate results.

Limitations

One major weakness of this study design is that the reference standard was taken as the triage nurse's triage category, but different A&E nurses come with different backgrounds, experiences and, thus, triage abilities. Hence, A&E nurses' triage decision may not be the best surrogate for appropriateness of A&E use in some cases, and future studies should consider using a stricter standard, such as a triage consensus among several senior nurses, to compare symptom checkers against. Moreover, the symptom checkers' source of information is solely based on the A&E staff's assessment and documentation, which, if inadequate, could have led to an underestimation of the checkers' accuracies. On the other hand, cases with chief complaints not available on the symptom checkers were replaced by more compatible cases, which likely resulted in an overestimation of the checkers' accuracies. Since only 51% of the tested A&E cases overlapped between the two checkers, comparing the checkers' accuracies with our study design may not have been ideal, but the conclusions drawn from each of the checker's results and accuracy are still valid.

Conclusion

Online symptom checkers are currently inappropriate to serve as a primary triage tool in the A&E, due to their low overall accuracies and NPVs. More aggressive triaging guidelines and additional improvements in function are necessary for symptom checkers to achieve a higher level of sensitivity and specificity. Potential applications of current checkers in A&E settings include providing a tentative triage level prior to the patient's arrival to the A&E to help with resource preparation. Further research must be undertaken to improve symptom checkers' triage algorithms to generate more accurate results.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Availability of data and materials

Data and materials can be made available to the public through direct requests to the corresponding author.

Ethical approval

The study protocol was approved by the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster.

Human rights

No human subjects were included in the study.

Informed consent

No human subjects were included in the study.

ORCID iDs

Stephanie Wing Yin Yu https://orcid.org/0000-0003-0674-1515

Siu-Chung Leung https://orcid.org/0000-0003-0169-9899

References

1. Fox S and Duggan M. One in three American adults have gone online to figure out a medical condition. Pew Research Center, http://www.pewinternet.org/2013/01/15/health-online-2013/ (2013, accessed 26 October 2018).
2. Push Doctor. One in four people in the UK admit to self-diagnosis of an illnesses rather than making time for a doctor's appointment, https://www.pushdoctor.co.uk/digital-health-report (2015, accessed 26 October 2018).
3. North F, Ward WJ, Varkey P, et al. Should you search the Internet for information about your acute symptom? *Telemed J E Health* 2012; 18(3): 213–218.
4. Semigran HL, Linder JA, Gidengil C, et al. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015; 351: h3480.
5. Gann B. Giving patients choice and control: health informatics on the patient journey. *Yearb Med Inform* 2012; 7: 70–73.
6. New England Healthcare Institute (NEHI). A matter of urgency: reducing emergency department overuse. NEHI Research Brief, https://www.nehi.net/publications/6-a-matter-of-urgency-reducing-emergency-department-overuse/view (2010, accessed 26 October 2018).
7. Nonurgent Use of Hospital Emergency Departments. United States Senate, Session 112-789. Printer's No. 81-788.
8. Turner J, Coster J, Chambers D, et al. *What evidence is there on the effectiveness of different models of delivering urgent care? A rapid review*. Southampton: Health Services and Delivery Research, 2015.
9. Bisson LJ, Komm JT, Bernas GA, et al. Accuracy of a computer-based diagnostic program for ambulatory patients with knee pain. *Am J Sports Med* 2014; 42(10): 2371–2376.
10. Farmer SE, Bernardotto M and Singh V. How good is Internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? *Clin Otolaryngol* 2011; 36(5): 517–518.
11. Poote AE, French DP, Dale J, et al. A study of automated self-assessment in a primary care student health centre setting. *J Telemed Telecare* 2014; 20(3): 123–127.
12. Wolf JA, Moreau JF, Akilov O, et al. Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatol* 2013; 149(4): 422–426.
13. Alexa Internet, Inc. drugs.com Traffic Statistics, https://www.alexa.com/siteinfo/drugs.com (2018, accessed 26 October 2018).
14. Alexa Internet, Inc. familydoctor.org Traffic Statistics, https://www.alexa.com/siteinfo/familydoctor.org (2018, accessed 26 October 2018).
15. Bujang MA and Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. *J Clin Diagn Res* 2016; 10(10): YE01–YE06.
16. Mistry B, Stewart De Ramirez S, Kelen G, et al. Accuracy and reliability of emergency department triage using the emergency severity index: an international multicenter assessment. *Ann Emerg Med* 2018; 71(5): 581.e3–587.e3.
17. Jordi K, Grossmann F, Gaddis GM, et al. Nurses' accuracy and self-perceived ability using the Emergency Severity Index triage tool: a cross-sectional study in four Swiss hospitals. *Scand J Trauma Resusc Emerg Med* 2015; 23: 62.
18. Hinson JS, Martinez DA, Schmitz PSK, et al. Accuracy of emergency department triage using the Emergency Severity Index and independent predictors of under-triage and over-triage in Brazil: a retrospective cohort analysis. *Int J Emerg Med* 2018; 11(1): 3.
19. Allen AR, Spittal MJ, Nicolas C, et al. Accuracy and inter-rater reliability of paediatric emergency department triage. *Emerg Med Australas* 2015; 27(5): 447–452.
20. Fan MM and Leung LP. Validation of the Hong Kong accident and emergency triage guidelines. *Hong Kong Med J* 2013; 19(3): 198–202.
21. Giesen P, Ferwerda R, Tijssen R, et al. Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency. *Qual Saf Health Care* 2007; 16(3): 181–184.