A type of field experiment used in economics, sociology, political science, and psychology, an audit study is one in which trained employees of the researcher ("auditors") are matched on all characteristics except the one being tested for discrimination. These auditors then apply for a service, be it a job, financial advice regarding their stock portfolio,[1] housing, or a credit card, to test for discrimination.

Evaluation of symptom checkers for self diagnosis and triage: audit study

Hannah L Semigran,[1] Jeffrey A Linder,[2] Courtney Gidengil,[3, 4] Ateev Mehrotra[1, 5]

[1]Department of Health Care Policy, Harvard Medical School, Boston, MA 02115, USA

[2]Division of General Medicine and Primary Care, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA

[3] Division of Infectious Diseases, Boston Children's Hospital, Boston, MA, USA

[4]RAND Corporation, Boston, MA, USA

[5]Division of General Internal Medicine and Primary Care, Beth Israel Deaconess Medical Center, Boston, MA, USA

Correspondence to: A Mehrotra
mehrotra@hcp.med.harvard.edu

Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmj.h3480)

Cite this as: *BMJ* 2015;351:h3480
doi: 10.1136/bmj.h3480

Accepted: 15 June 2015

ABSTRACT

OBJECTIVE

To determine the diagnostic and triage accuracy of online symptom checkers (tools that use computer algorithms to help patients with self diagnosis or self triage).

DESIGN

Audit study.

SETTING

Publicly available, free symptom checkers.

PARTICIPANTS

23 symptom checkers that were in English and provided advice across a range of conditions. 45 standardized patient vignettes were compiled and equally divided into three categories of triage urgency: emergent care required (for example, pulmonary embolism), non-emergent care reasonable (for example, otitis media), and self care reasonable (for example, viral upper respiratory tract infection).

MAIN OUTCOME MEASURES

For symptom checkers that provided a diagnosis, our main outcomes were whether the symptom checker listed the correct diagnosis first or within the first 20 potential diagnoses (n=770 standardized patient evaluations). For symptom checkers that provided a triage recommendation, our main outcomes were whether the symptom checker correctly recommended emergent care, non-emergent care, or self care (n=532 standardized patient evaluations).

RESULTS

The 23 symptom checkers provided the correct diagnosis first in 34% (95% confidence interval 31% to 37%) of standardized patient evaluations, listed the correct diagnosis within the top 20 diagnoses given in 58% (55% to 62%) of standardized patient evaluations, and provided the appropriate triage advice in 57% (52% to 61%) of standardized patient evaluations. Triage performance varied by urgency of condition, with appropriate triage advice provided in 80% (95% confidence interval 75% to 86%) of emergent cases, 55% (47% to 63%) of non-emergent cases, and 33% (26% to 40%) of self care cases (P<0.001). Performance on appropriate triage advice across the 23 individual symptom checkers ranged from 33% (95% confidence interval 19% to 48%) to 78% (64% to 91%) of standardized patient evaluations.

CONCLUSIONS

Symptom checkers had deficits in both triage and diagnosis. Triage advice from symptom checkers is generally risk averse, encouraging users to seek care for conditions where self care is reasonable.

WHAT IS ALREADY KNOWN ON THIS TOPIC

The public is increasingly using the internet for self diagnosis and triage advice, and there has been a proliferation of computerized algorithms called symptom checkers that attempt to streamline this process

Despite the growth in use of these tools, their clinical performance has not been thoroughly assessed

WHAT THIS STUDY ADDS

Our study suggests that symptom checkers have deficits in both diagnosis and triage, and their triage advice is generally risk averse

Introduction

Members of the public are increasingly using the internet to research their health concerns. For example, the United Kingdom's online patient portal for national health information, NHS Choices, reports over 15 million visits per month.[1] More than a third of adults in the United States regularly use the internet to self diagnose their ailments, using it both for non-urgent symptoms and for urgent symptoms such as chest pain.[2 3] While there is a wealth of online resources to learn about specific conditions, self diagnosis usually starts with search engines like Google, Bing, or Yahoo.[2] However, internet search engines can lead users to confusing and sometimes unsubstantiated information, and people with urgent symptoms may not be directed to seek emergent care.[3-6] Recently there has been a proliferation of more sophisticated programs called symptom checkers that attempt to more effectively provide a potential diagnosis for patients and direct them to the appropriate care setting.[3 6-13]

Using computerized algorithms, symptom checkers ask users a series of questions about their symptoms or require users to input details about their symptoms themselves. The algorithms vary and may use branching logic, bayesian inference, or other methods. Private companies and other organizations, including the National Health Service, the American Academy of Pediatrics, and the Mayo Clinic, have launched their own symptom checkers. One symptom checker, iTriage, reports 50 million uses each year.[14] Typically, symptom checkers are accessed through websites, but some are also available as apps for smart phones or tablets.

Symptom checkers serve two main functions: to facilitate self diagnosis and to assist with triage. The self diagnosis function provides a list of diagnoses, usually rank ordered by likelihood. The diagnosis function is typically framed as helping educate patients on the range of diagnoses that might fit their symptoms. The triage function informs patients whether they should seek care at all and, if so, where (that is, emergency department, general practitioner's clinic) and with what urgency (that is, emergently or within a few days).

Symptom checkers may supplement or replace telephone triage lines, which are common in primary care.[15-18] To ensure the safety of medical mobile apps, the US Congress is considering the regulation of apps that "provide a list of possible medical conditions and advice on when to consult a health care provider."[19 20]

Symptom checkers have several potential benefits. They can encourage patients with a life threatening problem such as stroke or heart attack to seek emergency care.[21] For patients with a non-emergent problem that does not require a medical visit, these programs can reassure people and recommend they stay home. For approximately a quarter of visits for acute respiratory illness such as viral upper respiratory tract infection, patients do not receive any intervention beyond over the counter treatment,[22] and over half of patients receive unnecessary antibiotics.[23-25] Reducing the number of visits saves patients' time and money, deters overprescribing of antibiotics, and may decrease demand on primary care providers—a critical problem given that the workload for general practitioners in the United Kingdom increased by 62% from 1995 to 2008.[17] However, there are several key concerns. If patients with a life threatening problem are misdiagnosed and not told to seek care, their health could worsen, increasing morbidity and mortality. Alternatively, if patients with minor illnesses are told to seek care, in particular in an emergency department, such programs could increase unnecessary visits and therefore result in increased time and costs for patients and society.

here,by "visits," they mean "visits to doctors or clinics"

The impact of symptom checkers will depend to a large degree on their clinical performance. To measure the accuracy of diagnosis and triage advice provided by symptom checkers, we used 45 standardized patient vignettes to audit 23 symptom checkers. The vignettes reflected a range of conditions from common to less common and low acuity to life threatening.

Methods

Search strategy for symptom checkers

Between June 2014 and November 2014 we searched for symptoms checkers that were in English, were free, were publicly available, were for humans (compared with veterinary use), and did not focus on a single type of condition (for example, only orthopedic problems). To find symptom checkers that were available as apps in the Apple app store and Google Play, we used two search phrases ("symptom checker", "medical diagnosis") used in a recent study on symptom checkers and examined the first 240 search results by hand.[12] We chose 240 because this cut-off has been used in previous studies that have searched smartphone app stores.[26] To find online symptom checkers, we entered the same two search phrases in Google and Google Scholar and examined the first 300 results. In previous research, the probability of relevant search results identified using Google declines substantially after the first 300 results.[27] 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. We excluded 102 that used the same medical content and logic as another tool (and therefore would have identical performance) (see list in supplementary appendix). We excluded a further 25 that only focused on a single class of illness (for example, orthopedic problems), 14 that only provided medical advice (for example, what symptoms are typically associated with a certain condition) and did not provide diagnosis or triage advice, and two that were not working. After these exclusions, we evaluated 23 symptom checkers.

Symptom checkers' characteristics

We categorized symptom checkers by whether they facilitated self diagnosis, self triage, or both; type of organization that operated the symptom checker; and the maximum number of diagnoses provided and whether they were based on Schmitt or Thompson nurse triage guidelines, which are decision support protocols commonly used in telephone triage for pediatric and adult consultations, respectively.[28 29] We grouped government and health plans together because both may have a financial incentive to deter unnecessary visits. In the supplementary appendix we provide data when available about estimated total visitors to select symptom checkers.

Clinical vignettes

To evaluate the diagnosis and triage performance of the symptom checkers, we used 45 standardized patient vignettes. We used clinical vignettes to assess performance because they are a common method to test physicians and other clinicians on their diagnostic ability and management decisions. We purposefully selected standardized patient vignettes from three categories of triage urgency: 15 vignettes for which emergent care is required, 15 vignettes for which non-emergent care is reasonable, and 15 vignettes for which a medical visit is generally unnecessary and self care is sufficient. We chose vignettes across the severity spectrum because patients use symptom checkers for symptoms that require both urgent and non-urgent care.[3] We included vignettes for both common and uncommon conditions because we believe that the clinical community would be particularly interested in performance for less common but potentially life threatening problems.

The standardized patient vignettes were identified from various clinical sources, including materials used to educate health professionals and a medical resource website with content provided by a panel of physicians.[30] The source for each vignette also provided the associated correct diagnosis. Symptom checkers generally require users to enter a list of symptoms or ask a series of questions about their symptoms. Each vignette was simplified into a core set of symptoms for easy entry, and in some situations we supplemented the data provided by the vignette because a symptom checker asked about a symptom not addressed in the vignette (see the supplementary appendix for details on source, core symptoms, and supplemental symptoms for each vignette).

doi: 10.1136/bmj.h3480 | BMJ 2015;351:h3480 | the bmj

We categorized the 45 vignettes as either "common" or "uncommon" diagnoses based on the prevalence of the diagnosis among ambulatory visits in the United States (for full details see the supplementary appendix).[31]

Assessing diagnosis and triage results

Each standardized patient vignette was entered into each website or app, and we recorded the resulting diagnoses and triage advice. An author (HS) with no clinical training entered all the vignettes. A random sample of 25 vignettes was entered into symptom checkers by another person without clinical training and the inter-rater reliability between the two in capturing the symptom checker's recommendations for diagnosis and triage was high (Cohen's κ 0.90). In some cases we could not evaluate a vignette because some symptom checkers focus only on children or on adults or the symptom checker did not list or ask for the key symptom in the vignette. To avoid penalizing these symptom checkers, we referred to standardized patient vignettes that successfully yielded an output as "standardized patient evaluations."

To assess diagnostic accuracy, we noted whether the correct diagnosis was listed first or listed at all. For several vignettes, two symptom checkers presented a large number of diagnoses (as much as 99). Because such a long list of potential diagnoses is unlikely to be useful for patients, we considered a diagnosis to be listed at all only if it was within the first 20 diagnoses provided by a symptom checker. It is possible that many patients only focus on the top diagnoses listed. Therefore we also looked at whether the correct diagnosis was listed in the first three diagnoses given. We judged the diagnosis incorrect if the symptom checker indicated that the condition could not be identified.

We categorized the triage advice into three groups: emergent, which included advice to call an ambulance, go to the emergency department, or see a general practitioner immediately; non-emergent, which included advice to call a general practitioner or primary care provider, see a general practitioner or primary care provider, go to an urgent care facility, go to a specialist, go to a retail clinic, or have an e-visit; and self care, which included advice to stay at home or go to a pharmacy. If multiple triage locations were suggested (for example, emergency department or specialist), we used the most urgent suggestion. We chose to do so because in almost all of the cases the most urgent triage suggestion was listed first. If a symptom checker was unable to reach a decision on diagnosis for a given standardized patient vignette but provided triage advice, we still assessed the appropriateness of this triage advice. Symptom checkers that required users to select the correct diagnosis before giving triage advice were not included in assessing the accuracy of triage with the exception of iTriage, which always suggested emergent triage advice.

Patient involvement

There was no patient involvement in this study.

Analysis

We calculated summary statistics for diagnostic accuracy and triage advice with 95% confidence intervals based on binomial distribution using Stata/MP 13.0. Given our focus on symptom checkers as a whole, we did not make statistical comparisons of accuracy between individual symptom checkers. We used χ^2 tests to compare the diagnosis and triage accuracy by level and urgency and by type of symptom checker. We conducted a sensitivity analysis of triage advice, excluding several symptom checkers that always or usually recommended emergent care.

Results

Study sample

The 23 identified symptom checkers were based in the United Kingdom, United States, the Netherlands, and Poland (table 1): 11 symptom checkers provided both diagnoses and triage advice, eight only provided diagnoses, and four only provided triage advice. The 45 standardized patient vignettes included 26 common and 19 uncommon diagnoses. Performance was assessed on a total of 770 standardized patient evaluations for diagnosis and 532 standardized patient evaluations for triage. Across the symptom checkers, 10 did not ask for demographics (age and sex).

Accuracy of diagnosis

Overall, the correct diagnosis was listed first in 34% (95% confidence interval 31% to 37%; table 2) of standardized patient evaluations. Performance varied by urgency of condition. The correct diagnosis was listed first for 24% (19% to 30%) of emergent standardized patient evaluations, 38% (32% to 34%) of non-emergent standardized patient evaluations, and 40% (34% to 47%) of self care standardized patient evaluations (P<0.001 for comparison, table 2). There was no difference between symptom checkers that asked for and did not ask for demographic information (34%, 95% confidence interval 30% to 39% and 34%, 28% to 39%, P=0.88; table 3). The correct diagnosis was, however, listed first more often in standardized patient evaluations for common diagnoses than for uncommon diagnoses (38%, 34% to 43% and 28%, 23% to 33%, P=0.004; table 2).

Performance varied across symptom checkers. Listing the correct diagnosis first in standardized patient evaluations ranged from 5% for MEDoctor (95% confidence interval 0% to 13%) to 50% for DocResponse (33% to 67%; table 4). Few differences were observed by the symptom checkers' characteristics (table 3).

Across all symptom checkers the correct diagnosis was listed in the first three diagnoses in 51% (95% confidence interval 47% to 54%) of standardized patient evaluations and in the first 20 diagnoses in 58% (55% to 62%) of standardized patient evaluations (table 2). Diagnostic accuracy for listing the correct diagnosis in the top three and top 20 was higher for self care conditions than for emergent conditions and was also higher for common conditions than for uncommon conditions. There was no significant difference in listing the correct

Table 1 | Symptom checkers included in the study

Symptom checker	Description	Maximum No of diagnoses*	Triage options provided
AskMD (USA)	Online health and wellness platform from Sharecare (www.sharecare.com/askmd/get-started)	15	Not available
BetterMedicine (USA)	Health resource from HealthGrades; symptom checker provides possible diagnoses and information about these conditions (www.bettermedicine.com/symptom-checker/)	46	Not available
DocResponse (USA)	Symptom checker started by a group of certified physicians; user can choose from internal medicine, dermatology, and orthopedic views (www.docresponse.com/)	5	Not available
Doctor Diagnose (USA)	App offered on Google Play; provides potential diagnoses and triage advice in some cases	3	Seek immediate care; call your doctor now; speak with your doctor; home care
Drugs.com (USA)	Online resource for drug and related health information; uses content from Harvard Health Publications (www.drugs.com/symptom-checker/)	10	Emergency department; primary care doctor; home care
EarlyDoc (Netherlands)	For triage criteria, uses Dutch College of General Practitioners (NHG) TriageWijzer and the Australian Triage Scale (used in Australia and New Zealand to assess urgency) (www.earlydoc.com/en/)	3	Don't wait, and call a doctor now; call a doctor preferably today; see your doctor preferably on a weekday; your complaints don't seem urgent
Esagil (USA)	Provides list of likely diagnoses (based on number of entered symptoms that are congruent with diagnosis); user can also enter blood and urine lab results along with symptoms (http://esagil.org/)	65	Not available
Family Doctor (USA)	Displays flow chart to track symptoms to diagnosis and triage option; created by American Academy of Physicians (http://familydoctor.org/familydoctor/en/health-tools/search-by-symptom.html)	7	Emergency room; see your doctor; home care
FreeMD (USA)	Takes user through a series of questions in a "check-up" to finish with "what might be wrong with you" and "where to go for care"; owned by DSHI Systems, which provides triage decision support solutions from emergency medicine physicians to the US government (Department of Veteran Affairs) and private sector companies; program called TriageXpert (www.freemd.com/)	3	Emergency department; urgent care; doctor's office; doctor e-visit; retail clinic; dentist; home care
Harvard Medical School Family Health Guide (USA)	From Harvard Health Publications; available both online and in print† (www.health.harvard.edu/fhg/symptoms/symptoms.shtml)	4	Emergency department; general practice; home care
Healthline (USA)	Health and wellness website that licenses content to employers, health providers, and health plans (www.healthline.com/symptom-checker)	76	Not available
Healthwise (USA)	Non-profit provider for health content and patient education; symptom checker licensed to other organizations; we accessed using Province of Alberta's website (https://myhealth.alberta.ca/health/pages/symptom-checker.aspx)	Not available	Call 911 now; seek care now; seek care today; try home care
Healthy Children (USA)	From American Academy of Pediatrics; use's Barton D Schmitt's "Pediatric HouseCalls Symptom Checker" triage protocol (www.healthychildren.org/English/tips-tools/symptom-checker)	Not available	Call 911 now; call your doctor now (night or day); call your doctor within 24 hours; call your doctor during weekday office hours; parent care at home
Isabel (UK)	Created by Isabel Medical Charity (http://symptomchecker.isabelhealthcare.com/suggest_diagnoses_advanced/landing_page)	10	Walk-in care; family doctor; emergency services
iTriage (USA)	Owned by Aetna; provides clinical sites in user's region with addresses and phone numbers (www.itriagehealth.com/avatar)	5	Emergency department, urgent care, retail clinic, family practice, internal medicine, specialties, prescription medication, over the counter medication
Mayo Clinic (USA)	Health resource website (www.mayoclinic.org/symptom-checker/select-symptom/itt-20009075)	20	Not available
MEDoctor (USA)	Free differential diagnosis system (www.medoctor.com/)	3	Not available
NHS Symptom Checkers (UK)	Available through England's National Health Services (NHS) Choices website (www.nhs.uk/symptomcheckers/pages/symptoms.aspx)	Not available	Emergency department; general practitioner; home care
Steps2Care (USA)	iPhone and Android app; provides symptom care guides from Barton D Schmitt's pediatric telephone triage guidelines and David A Thompson's adult telephone triage guidelines	Not available	Call 911 now; go to emergency room now; call doctor now or go to emergency room; call doctor within 24 hours; call doctor during office hours; self care at home
Symcat (USA)	Triage tool uses data linking specific patient symptoms and physician diagnoses across visits seen in the National Ambulatory Medical Care Survey (NAMCS) (www.symcat.com/)	6	Primary care; retail clinic; emergency room; urgent care
Symptify (USA)	Online self assessment tool and other health services, including, for example, emergency contact list, consultation list (https://symptify.com/)	9	Emergency room; urgent care; home care; inconclusive
Symptomate (Poland)	Uses bayesian network methodology and medical database for diagnoses (https://symptomate.com/)	5	Emergency room; specialist; general practitioner
WebMD (USA)	Medical reference and healthcare resource website (http://symptoms.webmd.com)	99	Not available

*Symptom checkers that provided diagnostic advice presented a list of potential diagnoses. We identified the maximum number of diagnoses presented across applicable vignettes.
†The online tool often refers the user to the book to make a decision on diagnosis and triage. For this study, we assessed the print version of the symptom checker.

Table 2 | Accuracy of diagnosis decision and triage advice for all symptom checkers, stratified by severity of standardized patient (SP) vignette and by frequency of the condition's diagnosis

Type of vignette or diagnosis	No of vignettes (%)	Diagnosis						Triage		
		Listed first			Listed in top 3			Listed in top 20				
		Rate*	% (95% CI)	P value	Rate*	% (95% CI)	P value	Rate*	% (95% CI)	P value	Rate*	% (95% CI)	P value
All vignettes	45 (100)	262/770	34 (31 to 37)	—	394/770	51 (47 to 54)	—	449/770	58 (55 to 62)	—	301/532	57 (52 to 61)	—
Type of SP vignette:													
Emergent	15 (33)	64/263	24 (19 to 30)		104/263	40 (34 to 46)		132/263	50 (44 to 56)		147/183	80 (75 to 86)	
Non-emergent	15 (33)	98/260	38 (32 to 44)	<0.001	148/260	57 (51 to 63)	<0.001	157/260	60 (54 to 66)	0.003	96/175	55 (47 to 63)	<0.001
Self care	15 (33)	100/247	40 (34 to 47)		142/247	57 (51 to 63)		160/247	65 (59 to 71)		58/174	33 (26 to 40)	
Type of diagnosis†:													
Common	26 (58)	174/457	38 (34 to 43)	0.004	254/457	56 (52 to 61)	<0.001	283/457	62 (57 to 66)	0.01	162/313	52 (46 to 57)	0.01
Uncommon	19 (42)	88/313	28 (23 to 33)		140/313	45 (38 to 49)		166/313	53 (47 to 59)		139/219	63 (57 to 70)	

*Number of correct SP evaluations divided by applicable SP evaluations. Some SP vignettes could not be applied to a given symptom checker (see text). For example, an adult SP vignette could not be evaluated if it was a pediatric symptom checker.
†Based on annual number of ambulatory care visits in United States, 2009-10 (see supplementary appendix for further description).

diagnosis in the top 20 between symptom checkers that listed more than 11 diagnoses compared with those that only listed 1-3 diagnoses (59%, 53% to 65% *v* 53%, 46% to 59%, P=0.12; table 3). The accuracy of listing the correct diagnosis in the top 20 across the 23 individual symptom checkers ranged from 34% (95% confidence interval 17% to 52%) to 84% (73% to 95%, table 4).

Accuracy of triage advice

Appropriate triage advice was given in 57% (95% confidence interval 52% to 61%) of standardized patient evaluations (table 2). Performance on triage advice was higher for emergent care standardized patient evaluations than for non-emergent and self-care standardized patient evaluations: 80% (75% to 86%) *v* 55% (47% to 63%) *v* 33% (26% to 40%), P<0.001). Appropriate triage advice was higher for uncommon diagnoses than for common diagnoses: 63% (57% to 70%) *v* 52% (46% to 57%), P=0.01).

iTriage, Symcat, Symptomate, and Isabel always suggested users seek care and therefore never advised self care (table 4). After excluding these four symptom checkers, appropriate triage advice was given in 61% (95% confidence interval 56% to 66%) of standardized patient evaluations (see supplementary table 5).

Symptom checkers that used the Schmitt or Thompson nurse triage protocols were more likely to provide appropriate triage decisions than those that did not: 72% (95% confidence interval 60% to 84%) *v* 55% (50% to 59%), P=0.01; table 3. Accurate triage advice varied by operator of symptom checker (provider groups and physician associations 68% (58% to 77%), private companies 59% (53% to 65%), health plans or governments 43% (34% to 51%), P<0.001).

Discussion

Using standardized patient vignettes we measured the diagnostic and triage accuracy of symptom checkers. Although there was a range of performance across symptom checkers, overall they had deficits in both diagnosis and triage accuracy. On average, symptom checkers provided the correct diagnosis within the first 20 listed in 58% of standardized patient evaluations, with the best performing symptom checker listing the correct diagnosis in 84% of standardized patient evaluations. Symptom checkers advised the appropriate level of care about half the time, but this varied by clinical severity. The correct triage decision was much higher for standardized patient evaluations requiring emergent care (80%) than for those for which self care was appropriate (34%).

Comparisons with other studies

Our results on diagnostic accuracy and appropriate triage are roughly similar to previous work on the performance of single symptom checkers for a limited set of diagnoses.[6-8 32] An orthopedic symptom checker listed the correct diagnosis for knee pain 89% of the time, and Boots WebMD listed the correct diagnosis 70% of the time for ear, nose, and throat symptoms.[7 8] One study that also used two common acute standardized patient

Table 3 | Accuracy of diagnosis given and triage advice for all symptom checkers given certain characteristics of the tools

Symptom checker characteristics	All symptom checkers	Diagnosis						Triage			
		No of symptom checkers (%)	Listed first		P value	Listed in top 20		P value	No of symptom checkers (%)	Rate*	% (95% CI)	P value
			Rate*	% (95% CI)		Rate*	% (95% CI)					
All symptom checkers	23 (100)	19 (100)	262/770	34 (31 to 37)		449/770	58 (55 to 62)	—	15 (100)	301/532	57 (52 to 61)	
Uses nurse triage books?†:												
Yes	2 (9)	0 (0)	0	—‡	—‡	0	—‡	—‡	2 (13)	41/57	72 (60 to 84)	0.01
No	21 (91)	19 (100)	262/770	34 (31 to 37)		449/770	58 (55 to 62)		13 (87)	260/475	55 (50 to 59)	
Asks demographic questions?:												
Yes	14 (61)	12 (63)	157/458	34 (30 to 39)	0.88	275/458	60 (56 to 65)	0.24	9 (60)	162/319	51 (45 to 56)	<0.001
No	9 (39)	7 (37)	105/312	34 (28 to 39)		174/312	56 (50 to 61)		6 (40)	139/213	65 (59 to 72)	
Site owner:												
Health plan or government	3 (13)	1 (5)	16/44	36 (22 to 51)	0.9	34/44	77 (64 to 90)	0.01	3 (20)	56/131	43 (34 to 51)	<0.001
Provider group	4 (17)	5 (26)	56/167	34 (26 to 41)		104/167	62 (55 to 70)		4 (27)	65/96	68 (58 to 77)	
Private Company	14 (61)	13 (68)	190/559	34 (30 to 38)		311/559	56 (52 to 60)		8 (53)	180/305	59 (53 to 65)	
Maximum No of diagnoses listed:												
1-3	6 (32)	6 (32)	78/227	34 (28 to 41)	0.13	120/227	53 (46 to 59)	0.12	5 (33)	87/163	53 (46 to 60)	0.32
4-10	7 (37)	7 (37)	107/283	38 (32 to 43)		175/283	62 (56 to 68)		6 (40)	131/224	58 (52 to 65)	
≥11	6 (32)	6 (32)	77/260	30 (24 to 35)		154/260	59 (53 to 65)		4 (27)	—§	—§	

*Number of correct SP evaluations divided by applicable SP evaluations. Some SP vignettes could not be applied to a given symptom checker. For example, we could not evaluate an SP vignette aimed at adults if the symptom checker was for children.
†From the Schmitt or Thompson protocols.
‡Symptom checker does not provide suggestions for diagnosis.
§Symptom checker does not provide triage advice.

vignettes to evaluate WebMD reported a diagnostic accuracy rate of 50%.[6]

Whether this level of performance for diagnosis and triage we observed is acceptable depends on the standard for comparison. If symptom checkers are seen as a replacement for seeing a physician, they are likely an inferior alternative. It is believed that physicians have a diagnostic accuracy rate of 85-90%, though in some studies using clinical vignettes, performance was lower.[33 34] However, in-person physician visits might be the wrong comparison because patients are likely not using symptom checkers to obtain a definitive diagnosis but for quick and accessible guidance. Also, instead of diagnostic accuracy the key assessment of symptom checkers may be appropriate triage. Distinguishing between Rocky Mountain spotted fever and meningitis may be less important than ensuring patients seek emergent care.

If symptom checkers are seen as an alternative for simply entering symptoms into an online search engine such as Google, then symptom checkers are likely a superior alternative. A recent study found that when typing acute symptoms that would require urgent medical attention into search engines to identify symptom-related web sites, advice to seek emergent care was present only 64% of the time.[3]

Perhaps the most appropriate comparison to symptom checkers is telephone triage lines, which are widely used in developed nations.[15-18] In general patients use symptom checkers and telephone triage for similar complaints.[13] Also, many nurse phone triage lines use the same underlying clinical logic as the symptom checkers evaluated in this study. For example, some health plan nurse triage lines use the Healthwise symptom checker, and the Schmitt and Thompson protocols were originally developed for phone triage and now provide the underlying logic for several symptom checkers that we evaluated. The accuracy of telephone triage recommendations, as compared to in-person physician recommendations, ranged from 61% in a study of pediatric abdominal pain to 69% in a multicenter observational study.[35 36] A recent study of NHS Symptom Checkers and NHS Direct's telephone triage line found triage advice from both to be comparable.[9] Given their similar clinical logic, triage performance, and their negligible operation costs, symptom checkers could potentially be a more cost effective way of providing triage advice than nurse-staffed phone lines.[17]

Implications for using symptom checkers

Both symptom checkers and telephone triage have been promoted as a means of reducing unnecessary office visits.[15-18 37] The impact of symptom checkers on how people seek care depends on how patients respond to advice, and this is unknown. In one study, users expressed skepticism about the diagnosis ultimately suggested by a symptom checker.[6] The risk averse nature of symptom checkers' triage advice is a concern. In two thirds of standardized patient evaluations where medical attention was not necessary, we found symptom

doi: 10.1136/bmj.h3480 | BMJ 2015;351:h3480 | the bmj

Table 4 | Accuracy of diagnosis decision and triage advice for each symptom checker

Symptom checker (n=23)	Diagnosis						Triage							
	Listed first		Listed in top 3		Listed in top 20		All cases		Emergent care required		Non-emergent care reasonable		Self care reasonable	
	Rate*	% (95% CI)	Rate*	% (95% CI)	Rate*	% (95% CI)	Rate*	% (95% CI)	Rate*	% (95% CI)	Rate*	% (95% CI)	Rate*	% (95% CI)
Ask MD	17/40	43 (26 to 59)	27/40	68 (52 to 83)	30/40	75 (61 to 89)	—†	—†	—†	—†	—†	—†	—†	—†
BetterMedicine	11/45	24 (11 to 38)	13/45	29 (15 to 43)	17/45	38 (23 to 53)	—†	—†	—†	—†	—†	—†	—†	—†
DocResponse	18/36	50 (33 to 67)	24/36	67 (50 to 83)	26/36	72 (57 to 88)	—†	—†	—†	—†	—†	—†	—†	—†
Doctor Diagnose	16/39	41 (25 to 57)	17/39	44 (27 to 60)	18/39	46 (30 to 63)	10/16	63 (36 to 89)	8/10	80 (13 to 100)	2/3	67 (0 to 100)	0/3	0 (0 to0)
Drugs.com	17/43	40 (24 to 55)	20/43	47 (31 to 63)	25/43	58 (43 to74)	25/42	60 (44 to75)	8/14	57 (27 to 87)	9/15	60 (32 to 88)	8/13	62 (31 to 92)
EarlyDoc	6/19	32 (9 to 55)	7/19	33 (9 to 57)	7/19	33 (9 to57)	9/17	53 (26 to79)	4/6	67 (12 to 100)	3/5	60 (0 to 100)	2/6	33 (0 to 88)
Esagil	9/44	20 (8 to33)	15/44	34 (24 to 54)	22/44	50 (35 to65)	—†	—†	—†	—†	—†	—†	—†	—†
Family Doctor	20/43	47 (31 to62)	24/43	56 (40 to 71)	24/43	56 (40 to71)	22/41	54 (38 to70)	6/12	50 (17 to 83)	7/14	50 (20 to 80)	9/15	60 (32 to 88)
FreeMD	16/44	36 (22 to51)	20/44	45 (30 to 61)	21/44	48 (32 to63)	26/44	59 (44 to74)	10/15	67 (40 to 94)	13/15	87 (67 to 100)	3/14	21 (0 to 46)
HMS Family Health Guide	13/38	34 (18 to50)	20/38	52 (39 to 72)	21/38	55 (39 to72)	32/40	78 (64 to91)	13/14	93 (77 to 100)	11/13	85 (62 to 100)	8/13	62 (31 to 92)
Healthline	17/45	38 (23 to53)	24/45	53 (37 to 68)	26/45	58 (43 to73)	—†	—†	—†	—†	—†	—†	—†	—†
Healthwise	—‡	—‡	—‡	—‡	—‡	—‡	19/44	43 (28 to58)	15/15	100 (100 to 100)	1/15	7 (0 to 21)	3/14	21 (0 to 46)
Healthy Children	—‡	—‡	—‡	—‡	—‡	—‡	11/15	73 (48 to99)	3/3	100 (100 to 100)	5/5	100 (100 to 100)	3/7	43 (0 to 92)
Isabel	20/45	44 (29 to60)	31/45	69 (55 to 83)	38/45	84 (73 to 95)	23/45	51 (36 to66)	15/15	100 (100 to 100)	8/15	53 (25 to 82)	0/15	0 (0 to 0)
iTriage	16/45	36 (22 to51)	29/45	64 (49 to 78)	34/45	77 (64 to90)	14/43	33 (19 to48)	14/14	100 (100 to 100)	0/14	0 (0 to 0)	0/15	0 (0 to 0)
Mayo Clinic	7/41	17 (5 to29)	24/41	59 (43 to 74)	31/41	76 (62 to89)	—†	—†	—†	—†	—†	—†	—†	—†
MEDoctor	2/37	5 (0 to13)	16/37	43 (26 to 60)	16/37	43 (26 to60)	—†	—†	—†	—†	—†	—†	—†	—†
NHS	—‡	—‡	—‡	—‡	—‡	—‡	23/44	52 (37 to68)	13/15	87 (67 to 100)	3/15	20 (0 to 43)	7/14	50 (20 to 80)
Steps2Care	—‡	—‡	—‡	—‡	—‡	—‡	30/42	71 (57 to86)	12/14	86 (65 to 100)	10/14	71 (44 to 98)	8/14	57 (27 to 87)
Symcat	18/45	40 (25 to55)	32/45	71 (57 to 85)	34/45	76 (62 to89)	20/45	44 (29 to60)	8/15	53 (25 to 82)	12/15	80 (57 to 100)	0/15	0 (0 to 0)
Symptify	13/45	29 (15 to43)	16/45	36 (22 to 51)	20/45	44 (29 to60)	28/40	70 (55 to85)	11/12	92 (73 to 100)	10/14	71 (44 to 98)	7/14	50 (20 to 80)
Symptomate	10/32	31 (14 to48)	11/32	34 (17 to 52)	11/32	34 (17 to52)	9/14	64 (36 to93)	7/9	76 (44 to 100)	2/3	67 (0 to 100)	0/2	0 (0 to 0)
WebMD	16/45	36 (21 to50)	23/45	51 (36 to 66)	28/45	62 (47 to77)	—†	—†	—†	—†	—†	—†	—†	—†

HMS=Harvard Medical School; NHS=National Health Service.
*Number of correct SP evaluations divided by applicable SP evaluations. Some SP vignettes could not be applied to a given symptom checker. For example, we could not evaluate an SP vignette aimed at adults if the symptom checker was for children.
†Symptom checker does not provide triage advice.
‡Symptom checker does not provide diagnosis suggestions.

checkers encouraged care. Overly risk adverse advice is not limited to symptom checkers. Telephone triage advice can also encourage unnecessary care seeking.[32 35] For instance, the NHS's telephone triage line, which is not staffed by health professionals, has been implicated in increasing visits to emergency departments in the UK.[38] Some patients researching health conditions online are motivated by fear, and the listing of concerning diagnoses by symptom checkers could contribute to hypochondriasis and "cyberchondria," which describes the escalated anxiety associated with self diagnosis on the internet.[39-43] Together, confusion, risk adverse triage advice, and cyberchondria could mean that symptom checkers encourage patients to receive care unnecessarily and thus increase healthcare spending. Understanding how patients interpret and use the advice from symptom checkers and the impact of symptom checkers on care seeking should be a key focus for future research.

The symptom checkers in this study represent the first generation of such tools, and there are several potential advances that may improve their performance in future versions. Incorporating local epidemiological data may help inform diagnoses. For instance, addition of real time information about the local incidence of illness in the community greatly improved the performance of a diagnostic tool for group A streptococcal pharyngitis.[10] Diagnosis and triage rates could also be improved if symptom checkers incorporated individual clinical data from medical claims or the electronic medical record. Demographic information is critical for both diagnostic and triage decisions for programs such as symptom checkers.[11] One surprising finding in our study was that symptom checkers that asked for demographic background information did not perform better. However, it is possible that this demographic information was not effectively incorporated into the symptom checkers' algorithms.

Strengths and limitations of this study

Despite the growing use of symptom checkers, we believe our study is the first to assess the clinical performance across a large number of symptom checkers and wide range of conditions.

There were key limitations to this study. We cannot be sure we identified all publicly available symptom checkers, despite a thorough search of relevant databases and consultation with experts in this discipline. We used clinical vignettes in which the symptoms and diagnoses were typically clear, and few vignettes included comorbid conditions, resulting in a possible overestimation of the true clinical accuracy of symptom checkers.[33] Some standardized patient vignettes contained specific clinical language (for example, mouth ulcers, tonsils with exudate), and actual patients with the same condition might struggle with the words to use to describe their symptoms or use different terms. Therefore, our analysis represents an indirect assessment of how well symptom checkers would perform with actual patients. We do not know how well physicians or other providers would diagnose or triage when presented with these standardized patient vignettes, preventing a direct comparison between symptom checkers and physicians. When symptom checkers suggested several care sites (for example, emergency department or general practice), our triage assessment was based only on the highest acuity site of care listed, and this may contribute to our finding that triage advice is risk averse.

Symptom checkers are part of a larger trend of both patients and physicians using the internet for many healthcare tasks and therefore it seems likely that the use of symptom checkers will only increase. Patients are chatting online with physicians,[44] emailing their doctors for medical advice,[45] receiving care through e-visits,[46 47] and downloading health apps to smartphones.[48] In addition to the public, physicians and other practitioners are also using conceptually similar tools to aid in the diagnosis and triage of their patients.[49 50]

Physicians should be aware that an increasing number of their patients are using new internet based tools such as symptom checkers and that the diagnosis and triage advice patients receive may often be inaccurate. For patients, our results imply that in many cases symptom checkers can give the user a sense of possible diagnoses but also provide a note of caution, as the tools are frequently wrong and the triage advice overly cautious. Symptom checkers may, however, be of value if the alternative is not seeking any advice or simply using an internet search engine. Further evaluations and monitoring of symptom checkers will be important to assess whether they help people learn more and make better decisions about their health.

Contributors: All authors conceived and designed the study. HLS acquired the data and drafted the manuscript. HLS and AM analysed and interpreted the data. CG, JAL, and AM critically revised the manuscript for important intellectual content. HLS and AM carried out the statistical analysis. AM provided administrative, technical, and material support and supervised the study. AM acts as guarantor.

Funding: This study was funded by the US National Institute of Health (National Institute of Allergy and Infectious Disease grant No R21 AI097759-01).

Competing interests: All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare: all authors are affiliated with Harvard Medical School. Harvard Medical School's Family Health Guide is used as the basis for one of the symptom checkers evaluated. This symptom checker is available both in print and online (www.health.harvard.edu/family_health_guide/symptoms). None of the authors have been or plan to be involved in the development, evaluation, promotion, or any other facet of a Harvard Medical School related symptom checker; the authors have no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required

Data sharing: No additional data available.

Transparency: The guarantor (AM) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

1 Gann B. Giving patients choice and control: health informatics on the patient journey. *Yearb Med Inform* 2012;7:70-3.
2 Fox S, Duggan M. Health Online 2013. Internet and American life project. Pew Research Center and California Health Care Foundation, 2013:4.
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;18:213-8.
4 Black P. The dangers of using Google as a diagnostic aid. *Br J Nurs* 2009;18:1157.
5 Zhongbo C, Turner MR. The internet for self-diagnosis and prognostication in ALS. *Amyotroph Lateral Scler* 2010;11:566.
6 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:e16.
7 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;42:2371-6.
8 Farmer SE, Bernardotto M, Singh V. How good is Internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? *Clin Otolaryngol* 2011;36:517-8.
9 Elliot AJ, Kara EO, Loveridge P, et al. Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system. *Epidemiol Infect* 2015:1-7.
10 Fine AM, Nizet V, Mandl KD. Participatory medicine: a home score for streptococcal pharyngitis enabled by real-time biosurveillance: a cohort study. *Ann Intern Med* 2013;159:577-83.
11 DocBot: a novel clinical decision support algorithm. Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual international conference of the IEEE; 2014 26-30 Aug 2014.
12 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;133:128-35.
13 North F, Varkey P, Laing B, et al. Are e-health web users looking for different symptom information than callers to triage centers? *Telemed J E Health* 2011;17:19-24.
14 Reuters. Aetna brings new iTriage employer technology to mid-sized businesses, 2013.
15 Lattimer V, George S, Thompson F, et al. Safety and effectiveness of nurse telephone consultation in out of hours primary care: randomised controlled trial. The South Wiltshire Out of Hours Project (SWOOP) Group. *BMJ* 1998;317:1054-9.
16 Bunn F, Byrne G, Kendall S. The effects of telephone consultation and triage on healthcare use and patient satisfaction: a systematic review. *Br J Gen Pract* 2005;55:956-61.
17 Campbell J, Fletcher E, Britten N, et al. Telephone triage for management of same-day consultation requests in general practice (the ESTEEM trial): a cluster randomised controlled trial and cost-sequence analysis. *Lancet* 2014;384:1859-68.
18 Richards D, Meakins J, Tawfik J, et al. Nurse telephone triage for same day appointments in general practice: multiple interrupted time series trial of effect on workload and costs. *BMJ* 2002;325.1214
19 Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. *J Med Internet Res* 2014;16:e210.
20 113th US Congress. Medical Electronic Data Technology Enhancement for Consumers' Health (MEDTECH), 2014.
21 Saczynski J, Yarzebski J, Lessard D, et al. Trends in pre-hospital delay in patients with acute myocardial infarction (from the Worcester Heart Attack Study). *Am J Cardiol* 2008;102:1591.
22 Barnett ML, Linder JA. Antibiotic prescribing for adults with acute bronchitis in the United States, 1996-2010. *JAMA* 2014;311:2020-2.
23 Barnett ML, Linder JA. Antibiotic prescribing to adults with sore throat in the United States, 1997-2010. *JAMA Intern Med* 2014;174:138-40.
24 Gonzales R, Malone DC, Maselli JH, et al. Excessive antibiotic use for acute respiratory infections in the United States. *Clin Infect Dis* 2001;33:757-62.
25 Little P, Rumsby K, Kelly J, et al. Information leaflet and antibiotic prescribing strategies for acute lower respiratory tract infection: a randomized controlled trial. *JAMA* 2005;293:3029-35.
26 Bierbrier R, Lo V, Wu RC. Evaluation of the accuracy of smartphone medical calculation apps. *J Med Internet Res* 2014;16:e32.
27 Orizio G, Merla A, Schulz PJ, et al. Quality of online pharmacies and websites selling prescription drugs: a systematic review. *J Med Internet Res* 2011;13:e74.
28 Schmitt BD. Pediatric telephone protocols: office version. American Academy of Pediatrics, 2012.
29 Thompson DA. Adult telephone protocols, 3rd edn. American Academy of Pediatrics, 2013.
30 Epocrates. 2014. www.epocrates.com/.
31 CDC, NAMCS, NHAMCS. Annual number and percent distribution of ambulatory care visits by setting type according to diagnosis group: United States, 2009-2010, 2010.
32 Poote A, French D, Dale J, et al. A study of automated self-assessment in primary care student health centre setting. *J Telemed Telecare* 2014;20:125.
33 Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013;22(Suppl 2):ii21-ii27.
34 Meyer AN, Payne VL, Meeks DW, et al. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med* 2013;173:1952-8.
35 Staub GM, von Overbeck J, Blozik E. Teleconsultation in children with abdominal pain: a comparison of physician triage recommendations and an established paediatric telephone triage protocol. *BMC Med Inform Decis Mak* 2013;13:110.
36 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;16:181-4.
37 Stacey D, Graham I, O'Connor A, Pomey M. Barriers and facilitators influencing call center nurses' decision support for callers facing values-sensitive decisions: a mixed methods study. *Worldviews Evid Based Nurs* 2005;2:184-95.
38 Donnelly L. A&E Crisis cause by NHS 111 phoneline, senior medic suggests. *Telegraph* 14 Jan, 2015.
39 Usborne S. Cyberchondria: the perils of internet self-diagnosis. Independent. 17 Feb, 2009.
40 Hartzband P, Groopman J. Untangling the web - Patients, doctors, and the internet. *N Engl J Med* 2010;362:1064.
41 Brigo F, Igwe SC, Ausserer H, et al. Why do people Google epilepsy? An infodemiological study of online behavior for epilepsy-related search terms. *Epilepsy Behav* 2014;31:67-70.
42 White RW, Horvitz E. Experiences with web search on medical concerns and self diagnosis. *AMIA Annu Symp Proc* 2009;2009:696-700.
43 Husain I, Spence D. Can healthy people benefit from health apps? *BMJ* 2015;350:h1887.
44 Eminovic N, Wyatt J, Tarpey A, et al. First evaluation of the NHS Direct Online Clinical Enquiry Service: A nurse-led web chat triage service for the public. *J Med Internet Res* 2004;6:e17.
45 Eysenbach G, Diepgen T. Patients looking for information on the internet and seeking teleadvice. *Arch Dermatol* 1999;135:151-6.
46 Mehrotra A, Paone S, Martich GD, et al. A comparison of care at e-visits and physician office visits for sinusitis and urinary tract infection. *JAMA Intern Med* 2013;173:72-4.
47 DeJong C, Santa J, Dudley RA. Websites that offer care over the Internet: is there an access quality tradeoff? *JAMA* 2014;311:1287-8.
48 Edney A. The FDA sets its sights on medical Apps: Bloomberg Businessweek, 20 Sep, 2013.
49 Cook DA, Enders F, Linderbaum JA, et al. Speed and accuracy of a point of care web-based knowledge resource for clinicians: a controlled crossover trial. *Interact J Med Res* 2014;3:e7.
50 Sadeghi S, Barzi A, Sadeghi N, et al. A Bayesian model for triage decision support. *Int J Med Inform* 2006;75:403-11.

Appendix: supplementary information