

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

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

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

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

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review

Journal:	BMJ Open
Manuscript ID	bmjopen-2018-027743
Article Type:	Research
Date Submitted by the Author:	06-Nov-2018
Complete List of Authors:	Chambers, Duncan; University of Sheffield, ScHARR; Cantrell, Anna; University of Sheffield, ScHARR Johnson, Maxine; University of Sheffield, ScHARR; University of Sheffield Preston, Louise; University of Sheffield, ScHARR Baxter, Susan; University of Sheffield, ScHARR Booth, Andrew; University of Sheffield, ScHARR Turner, Janette; University of Sheffield, ScHARR
Keywords:	urgent care, symptom checkers, systematic reviews

SCHOLARONE™ Manuscripts

Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review

Duncan Chambers, Anna Cantrell, Maxine Johnson, Louise Preston, Susan K Baxter, Andrew Booth and Janette Turner

School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, Sheffield S1 4DA, UK Duncan Chambers research fellow, Anna Cantrell research fellow, Maxine Johnson research fellow, Louise Preston research fellow, Susan K Baxter senior research fellow, Andrew Booth reader in evidence-based information practice, Janette Turner reader in emergency and urgent care research

*Correspondence to Duncan Chambers: d.chambers@sheffield.ac.uk

Contributor/guarantor information:

Susan K Baxter (Senior Research Fellow in Public Health): Protocol development; study selection; report writing; co-ordinated PPI meeting; Andrew Booth (Reader in Evidence-Based Information Practice): Information retrieval; study selection; report writing; Anna Cantrell (Research Associate in Health Economics and Decision Science): Information retrieval; study selection; data extraction; quality assessment; report writing; Duncan Chambers (Research Fellow in Public Health): Project co-ordination; study selection; data extraction; quality assessment; report writing; Maxine Johnson (Research Fellow in Public Health): Study selection; data extraction; quality assessment; report writing; Louise Preston (Research Fellow in Health Economics and Decision Science): Study selection; data extraction; quality assessment; report writing; Janette Turner (Reader in Emergency & Urgent Care Research): Topic expert advice: report writing. All authors commented on drafts of the protocol and report. Duncan Chambers is the guarantor for this work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

None of the authors have any competing interests

Copyright

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a non-exclusive worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.



Abstract

Objectives: In England, the NHS111 service provides assessment and triage by telephone for problems that are urgent but not classified as an emergency. A digital version of this service is currently being evaluated. We aimed to systematically review the evidence on digital and online symptom checkers and health assessment/triage services.

Design: Systematic review with narrative synthesis.

Setting: Primary care.

Participants: General population seeking information online or digitally to address an urgent health problem.

Interventions: Any online or digital service designed to assess symptoms, provide health advice and direct patients to appropriate services.

Primary and secondary outcome measures: Safety; clinical effectiveness; costs/cost-effectiveness; accuracy; impact on service use; compliance with advice received; patient/carer satisfaction; and equity and inclusion. Accuracy covered 1) ability to provide a correct diagnosis and 2) ability to distinguish between high and low acuity/urgency problems (and hence direct patients to appropriate services).

Results: We included 29 publications (27 studies). Evidence on patient safety was weak. Diagnostic accuracy varied between different systems but was generally low if health professionals' diagnoses were used as the reference standard. Algorithm-based triage tended to be more risk-averse than that of health professionals. There was very limited evidence on patients' compliance with online triage advice. Study participants generally expressed high levels of satisfaction with digital and online triage services, albeit in mainly uncontrolled studies. Younger and more highly educated people were more likely to use these services.

Conclusions: The English 'digital 111' service is being implemented against a background of uncertainty around the likely impact on important outcomes. The health system may need to respond to short-term increases (or decreases) in demand and/or shifts from one part of the system to another. The popularity of online and digital services with younger and more educated people has implications for health equity.

Registration: PROSPERO (registration number CRD42018093564)

Strengths and limitations of this study

- This systematic review was based on a rigorous search of the literature which maximised efficiency by combining an initial focused search with subsequent rounds of follow-up searching, including searches for named symptom checker systems.
- Our narrative synthesis approach used a mixture of description and tabulation to summarise the evidence, including overall strength of the evidence base for each of the pre-specified outcomes of interest.
- Given the decision to implement a national urgent care service based on digital symptom checkers in the NHS in England, our study highlights areas of uncertainty that will need to be resolved by research and data collection.
- The review inclusion criteria were relatively broad and findings from symptom checker systems for specific conditions may not be applicable to more general systems and vice versa.
- We have also included studies of symptom checkers as part of electronic consultation systems in general practice, which again represents a slightly different setting from a general 'digital 111' service, and this should be kept in mind when interpreting the results.

Introduction

Digital and online symptom checkers and assessment services are used by patients seeking guidance about health problems, including some that may require urgent action. These services generally provide people with possible alternative diagnoses based on their reported symptoms and/or suggest a course of action (e.g. self-care, make a GP appointment or go to an emergency department (ED)).

In England, the NHS111 service provides assessment and triage by telephone for problems that are urgent but not classified as emergencies. The latest data from NHS England¹ show that in September2018 there were over 1.27 million calls to NHS111, an average of 42,400 per day. Outcomes of these calls were that 13.2% had ambulances despatched; 9.5% were recommended to attend an ED; 58.7% were recommended to attend primary care; 4.8% to attend another service; and 13.8% were not recommended to attend another service (e.g. their condition was considered suitable for self-care)

NHS England is planning to introduce a digital platform to make NHS111accessible via a website or smartphone app. A beta version of the service (referred to as 'NHS111 Online') is available at https://l11.nhs.uk/ (accessed 26 October 2018). The 'digital 111' service is seen as key to reducing demand for the telephone 111 service, enabling resources to be redirected to supporting 'integrated urgent and emergency care systems' as outlined in the 'NHS 5-year Forward View' and its 2017 update 'Next Steps on the NHS 5-year Forward View'²³.

There is an expectation that a digital 111 platform will help to manage demand and increase efficiency in the urgent and emergency care system, complementing the agenda of locally based Sustainability and Transformation Partnerships (STPs). However, there is a risk of increasing demand, duplicating healthcare contacts and providing advice that is not safe or clinically appropriate. For example, an evaluation of the NHS111 telephone service at four pilot sites and three control sites found that in its first year the service was not successful in reducing 999 emergency calls or in shifting patients from emergency to urgent care⁴. A recent study of 23 symptom checker algorithms providing diagnostic and triage advice that would form the basis of a 'digital 111' platform found deficiencies in both their diagnostic and triage capabilities⁵.

In 2017, NHS England carried out pilot evaluations of different systems in four regions of England. The evaluations aimed to assess whether digital/online triage was acceptable to users and connected them to appropriate clinical care⁶. The full report of the evaluations was not yet published at the time of writing. The objective of this systematic review was to inform further development of the proposed digital platform by summarising and critiquing the previous research in this area, both from the UK and overseas. The overall research question was: for people seeking guidance about an urgent health problem, what is the effect of digital and online services designed to assess symptoms and signpost patients to appropriate services (compared with non-digital services or no comparator) on important clinical and health service outcomes? Outcomes include safety; clinical and cost-effectiveness; diagnostic and triage accuracy; impact on service use; patient/carer satisfaction; compliance with advice received; and outcomes related to equity and inclusion.

Methods

The review protocol was registered with PROSPERO (registration number CRD42018093564) and is available from the project website (https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/164717/).

Literature search and screening

Initial scoping searches revealed that a highly sensitive search strategy, as typically conducted for systematic reviews, retrieved a disproportionately high number of references on GP decision-making and triage. We therefore devised a three stage retrieval strategy as an acceptable alternative to comprehensive topic-based searching. This involved:

- 1. Targeted searches of precise high specificity terms in seven databases (MEDLINE, EMBASE, the Cochrane Library, CINAHL, HMIC (Health Management Information Consortium), Web of Science and ACM Digital Library). These searches were not restricted by language or date. A sample search strategy is presented in Appendix 1.
- 2. Phrase searching for names of known symptom checkers using a list compiled from Semigran 2015 and other sources

3. Citation searches and reference checking of key included studies and reviews, complemented by contact with service providers (directly and via websites).

Search results were stored in a reference management system (EndNote) and imported into EPPI-Reviewer software for screening, data extraction and quality assessment. The search results were screened against the inclusion criteria by one reviewer, with a 10% sample screened by a second reviewer. Uncertainties were resolved by discussion among the review team.

Inclusion and exclusion criteria

Population: General population seeking information online or digitally to address an urgent health problem, including adults and children and issues arising from both acute and long-term chronic illness.

Intervention: Any online or digital service designed to assess symptoms, provide health advice and direct patients to appropriate services. Services that only provide health advice were excluded, as were those that offer treatment, e.g. online CBT services.

Comparator: The 'gold standard' comparator is current practice of telephone assessment (e.g. NHS111) or face to face assessment (e.g. general practice, urgent care centre or ED). However, studies with other relevant comparators (e.g. comparative performance in tests or simulations) or with no comparator were included if they addressed the research questions.

Outcomes: The main outcomes of interest were safety (e.g. any evidence of adverse events arising from following or ignoring advice from online/digital services); clinical effectiveness; costs/cost-effectiveness; accuracy; impact on service use; compliance with advice received; patient/carer satisfaction; and equity and inclusion. 'Accuracy covered 1) ability to provide a correct diagnosis and 2) ability to distinguish between high and low acuity/urgency problems (and hence direct patients to appropriate services).

Study design: We did not restrict inclusion by study design (and included relevant audits or service evaluations in addition to formal research studies) but included studies had to evaluate (quantitatively or qualitatively) some aspect of an online/digital service **Other**: Studies from any developed country healthcare system were eligible for inclusion

Excluded: Purely descriptive studies, conceptual papers, projections of possible future developments and studies conducted in low or middle income countries were excluded from the review.

Data extraction and quality/strength of evidence assessment

We extracted and tabulated key data from the included studies, including study design, population/setting, results and key limitations. Data extraction was performed by one reviewer, with a 10% sample checked for accuracy and consistency.

To characterise the included digital and online systems as interventions, we identified studies reporting on a particular system and extracted data from all relevant studies using a modification of the TIDieR (Template for Intervention Description and Replication) checklist⁷ which we designated TIDieST (Template for Intervention Description for Systems for Triage). Further details may be found in the full report (Chambers et al., in preparation).

Quality (risk of bias) assessment was undertaken for peer-reviewed full publications only (i.e. not grey literature publications or conference abstracts). Randomised controlled trials were assessed using the Cochrane Collaboration risk of bias tool. For diagnostic accuracy type studies, we used the Cochrane Collaboration version of QUADAS and for other study design we used the National Heart Lung and Blood Institute tool for observational cohort and cross-sectional studies. Quality assessment was performed by one reviewer, with a 10% sample checked for accuracy and consistency.

Assessment of the overall strength (quality and relevance) of evidence for each research question is part of the narrative synthesis. Overall strength of the evidence base for key outcomes was assessed using an adaptation of the method described by Baxter et al.⁸ This involves classifying evidence as 'stronger', 'weaker', 'conflicting' or 'insufficient' based on study numbers and design. Specifically, "stronger evidence" represented generally consistent findings in multiple studies with a comparator group design or comparative diagnostic accuracy studies; "weaker evidence" represented generally consistent findings in one study with a comparator group design and several non-comparator studies or multiple non-comparator studies; "very limited evidence" represented an outcome reported by a single study; and finally, "inconsistent evidence" represented an outcome where fewer than 75% of studies agreed on the direction of effect. All studies in the review, including those that did not meet criteria for risk of bias assessment, were included in the strength of evidence assessment.

Evidence synthesis

We performed a narrative synthesis structured around the pre-specified research questions and outcomes. We did not perform any meta-analyses because the included studies varied widely in terms of design, methodology and outcomes.

Patient and public involvement (PPI)

The review was discussed at two meetings of an existing PPI group covering the programme from which the review was commissioned (Sheffield HS&DR Evidence Synthesis Centre). At the meetings there was discussion regarding the focus of the work, including a presentation on previous research on NHS111 telephone services to provide a context for understanding the current work. The meetings also included presentation and discussion of the findings of the review, in order to explore key messages for patients which could inform dissemination of the findings. Discussion during one meeting was structured using a SWOT (strengths, weaknesses, opportunities and threats) analysis approach, which revealed a number of potential concerns amongst patients as well as potential perceived benefits. Involvement of the advisory group was beneficial in highlighting some issues that had also emerged from the systematic review, and enabled the reviewers to structure the review findings taking this into account.

Results

Results of literature search

Twenty-seven studies (29 publications) were included in the review. Figure 1 presents the flow of studies through the selection process.

Figure 1: PRISMA flow diagram

Characteristics of included studies

Seventeen studies (Table 1) evaluated symptom checkers as a self-contained intervention, of which eight covered a limited range of symptoms, e.g. respiratory⁹⁻¹¹ or gastrointestinal^{12 13} symptoms which we considered to be 'urgent'. The remaining studies in this group evaluated symptom checkers covering a wider range of common urgent care symptoms. Studies either evaluated a single system¹⁴⁻¹⁷ or multiple systems^{5 18}. We found only one study of a symptom checker specifically intended for assessment of children's symptoms, a development of the SORT (Strategy for Off-Site Rapid Triage) system for influenza-like illness¹⁹ Two reports with some overlap of content evaluated the 'babylon check' app^{14 20}

Five studies⁶ ²¹⁻²⁴ evaluated symptom checkers as part of a broader self-assessment and consultation system (often referred to as electronic consultation or e-consultation). Study characteristics are summarised in Table 2. In this type of system, the role of symptom checkers is to help patients decide whether their symptoms require a consultation with a doctor or other health professional or can be dealt with by self-care. If a consultation is required, details of the symptoms and a request for an appointment or call-back can be submitted electronically. This type of study is important because it considers the service within the broader context of the urgent and emergency care system. A limitation is that some studies focused mainly on the 'downstream' elements of the pathway, e.g. consultation with GPs, and provided limited data on the symptom checker element of the system.

A final group of five studies examined patient and/or public attitudes to online self-diagnosis in the context of urgent care²⁵⁻²⁹. See the full report for further details (Chambers et al. in preparation).

Table 1: Studies of symptom checkers as a self-contained intervention

Reference	Study design	System type	Comparator >	Population/sample
Babylon	• Uncontrolled	• Digital	• Health professiona	General population
Health 2017 ²⁰	observational	Smartphone app	performance on real-world	Participants in the London
	No control group but		data	pilot evaluation of 'digital 111'
	some comparison with	X	• Other	services
	NHS111 telephone		<i>NHS111 data for 12</i> <u>≦</u>	
	data		months from Februa	
		' /	2017	
Berry	Simulation	• Online	• None	Specific condition(s)
201612	Evaluation of	17 symptom checkers	• None http://bmjope	Gastrointestinal symptoms
	symptom checker		D://b	
	performance on		mjc	
	clinical vignettes		pen	
Berry	Controlled	• Online	Health professionals	• Specific condition(s)
2017 ³⁰	observational	Three online symptom checkers (WebMD, iTriage and	performance on real-world	Patients with a cough
		FreeMD)	data	presenting to an internal
			on A	medicine clinic
Berry 2017 ¹³	Controlled	• Online	• Health professiona <u>É</u> .	• Specific condition(s)
	observational	Three online symptom checkers (WebMD, iTriage,	performance on real-world	Abdominal pain
		FreeMD)	data	
Kellermann	• Simulation	• Online	Health professional	• Specific condition(s)
2010 9	The developed	SORT was available on 2 interactive websites	performance on real world	Influenza symptoms
	algorithm was tested		data 🥳	
	against past patient		The algorithm was tested	
	records		against clinicians'	
			decision on past patient	
			records.	

Little 2016 ¹⁰	Experimental	• Online	• Other	Specific condition(s)
	Randomised	'Internet Doctor' website	Usual GP care without	Respiratory infections and
	controlled trial (RCT)		access to the Interne	associated symptoms
			Doctor website ≥	
Luger et al.	Simulation	• Online	• Other	General population
2014 31	Described as	Google and WebMD	Comparing two internet	Older adults (50 years or
	"human-computer		health tools.	older)
	interaction study"		Do	
	using think-aloud		wnic	
	protocols.		pade	
Marco-Ruiz	• Qualitative	• Online	health tools. 19. Downloaded from http://bmjopen.bmj.com/ on	General population
et al. 2017 ³²	Qualitative element	Erdusyk	om .	Internet tool users
	• Other	\Q_L	http	
	1. Online evaluation)://b	
	by users (problem		njo	
	detection) 2. Think		pen	
	aloud technique by	• Online Erdusyk	.bm	
	smaller sample of	'61	j. co	
	participants		m/ c	
	(usability)		on A	
Middleton	Simulation	• Digital	• Health professiona .	General population
2016^{14}		'babylon check' automatic triage system	performance on , o	
			performance on test/simulation 8	
			Twelve 'clinicians' 💆	
			(doctors) and 17 nurses	
Nagykaldi	• Uncontrolled	• Online	• None	Specific condition(s)
2010 33	observational	Customised practice website including a bilingual	None St. Protected by	Influenza
		influenza self-triage module, a downloadable	lecte	
		influenza toolkit and electronic messaging capability.	ed a	

		A bilingual seasonal influenza telephone hotline was	277.	
		available as an alternative.	743 (
Nijland	• Uncontrolled	• Online	None None August 2019	 General population
2016^{17}	observational	Web-based triage system (http://www.dokterdokter.nl)	Au	
	Retrospective		gus	
	analysis of 15 months'		t 20	
	data		19.	
Poote 2014	• Uncontrolled	• Online	• Health professional	General population
5	observational	Prototype self-assessment triage system	performance on real-world	Students attending a
			data a	University Student Health
			GPs triage rating was	Centre with new acute
		700	compared with rating	symptoms
		0,6	from the self-assessment	
		- / h	system	
Price 2013	• Uncontrolled	• Online	• Health professional	Specific condition(s)
9	observational	A web-based decision support tool - Strategy for Off-	performance on real-world	Influenza in children
		site Rapid Triage (SORT) for Kids designed to help	data	
		parents and adult caregivers decide whether a child	The sensitivity of the	
		with possible influenza symptoms needs to visit the	algorithm was compared	
		emergency department for immediate care.	with a gold standard 5 →	
			evidence form child's \(\frac{1}{2}\).	
			medical records that they	
			received 1 or more of \S	
			ED-specific intervent E ns.	
Semigran	• Experimental	Multiple	• Other	General population
2015 5	Described as an audit	23 symptom checkers were evaluated. Symptom	Vignettes had a diag <mark>n</mark> osis	Where a single class of
	study	checkers available as apps (via the App Store and	and triage attached t	illness was examined by the
		Google Play) were identified through searching for	them and these were	symptom checker, the
		"symptom checker" and "medical diagnosis" and	compared against the	symptom checker was
		screened the first 240 results. Symptom checkers	<u>٧</u>	

		available online were identified through searching Google and Google Scholar for "symptom checker" and "medical diagnosis" and screened the first 300 results.	symptom checker advice. 34 00 11 2	excluded from the study.
Semigran 2016 ¹⁸	• Experimental Comparison of physician and symptom checker diagnoses based on clinical vignettes	• Multiple "Human Dx is a web-and app based platform"	• Health professional performance on test/simulation Clincial vignettes - Comparison of 23 symptom checkers with physician diagnosis for 45 vignettes	• General population Of the 45 condition vignettes - there were 15 low, 15 medium and 15 huigh acuity vignettes - there were 26 common and 19 uncommon condition vignettes
Sole 2006 ¹⁶	• Uncontrolled observational Descriptive comparative study	Online A web-based triage system (24/7 WebMed)	• Health professional performance on real-world data Data was evaluated from students who had used the web based triage and then requested an appointment via email (so triage data was available for comparison).	General population
Yardley 2010 ¹¹	• Experimental Exploratory randomised trial	Online 'Internet Doctor' website	• Other Self-care information provided as a static web page with no symptom checker or triage advice	• Specific condition(s) Minor respiratory symptoms, e.g. cough, sore throat, fever, runny nose

Table 2: Studies of symptom checkers as part of an electronic consultation system

Table 2: St	udies of symptom c	BMJ Open heckers as part of an electronic consultation	system Comparator 2018-027743 on 1 August 20	
Reference	Study design	System type	Comparator g	Population/sample
Carter	Uncontrolled	• Online	• Other	General population
2018 21	observational	webGP(subsequently known as eConsult)	Investigate patient experience \overline{by}	General practices in NHS
	Mixed-methods		surveying patients who had use	Northern, Eastern and
	evaluation	O _A	webGP and comparing their	Western Devon Clinical
		- / h	experience with controls (patients	Commissioning Group's
		$\mathcal{O}_{\mathcal{O}}$	who had received a face-to-face	area
		1 60	consultation during the same time	
		· C/-	period) matched for age and gender	
Cowie	• Uncontrolled	• Online	• None	General population
2018 ²²	observational	eConsult, accessed via GP surgery websites.	jope	Patients registered with
	6-month evaluation	Service provides self-care assessment and	en.b	participating GP practices
	at 11 GP practices	advice, including symptom checkers; triage and	om _{j.c}	
	in Scotland	signposting to alternative services; access to	Oom	
		NHS24 (phone service); and e-consults allowing	/ on	
		submission of details by e-mail)	• None • None • None • None	
Madan	• Uncontrolled	• Online	• None	General population
2014 ²³	observational	webGP (subsequently known as eConsult)	, 20	
	Report of 6-month		24 1	
	pilot study		9	
NHS	• Uncontrolled	• Multiple	• None E	General population
England ⁶	observational	Pilots featured NHS Pathways (Web-based;	Authors stated it was not appropriate	
	Analysis of data	West Yorkshire); Sense.ly ('voice-activated	to compare pilot sites because 餐	
	from four pilot	avatar'; West Midlands); Espert 24 (Web-based;	differences in starting date,	
	studies together	Suffolk) and babylon (app; North Central	'footprints' covered, method of uptake	

				<u> </u>	
	with data from other	London)	and underlying population	277	
	sources			43	
Nijland	• Other	• Online	• None) 1	General population
2009 ²⁴	Online survey	Responses of interest relate to 'indirect e-		Au	Patients with Internet access
		consultation' (consulting a GP via secure e-mail	1	gus	but no experience of e-
		with intervention of a Web-based triage system)		t 20	consultation

Results by outcome

Safety

None of the six included studies that reported on safety outcomes identified any problems or differences in outcomes between symptom checkers and health professionals. Most of the studies compared system performance with that of health professionals using real or simulated data. The only study with no comparison group was the 6-month pilot study of webGP²³, which reported 'no major incidents'.

Limitations of the studies included not being based on real patient data¹⁴; covering only a limited range of conditions⁹ ¹⁹; and sampling a young healthy population (students) not representative of the general population of users of the urgent care system¹⁵. Studies of e-consultation systems did not generally collect data on those respondents who decided not to seek an appointment, limiting their ability to assess any impact on safety for this group. Overall, the evidence should be interpreted cautiously as indicating no evidence of a detrimental impact on safety rather than evidence of no detrimental effect.

Clinical effectiveness

Only two studies reported on clinical effectiveness outcomes, making it difficult to draw any firm conclusions. In the study by Little et al., those who used the Internet Doctor website experienced longer illness duration and more days of illness rated moderately bad or worse than the usual care group¹⁰ The pilot study of the webGP system²³ reported that several patients received advice to seek treatment for serious symptoms that might otherwise have been ignored. However, no details or quantitative data were provided.

Costs/cost-effectiveness

Two included studies provided limited data on possible cost savings. Based on 6 months of pilot data, Madan²³estimated savings of £11,000 annually for an average general practice (6,500 patients) compared with current practice. The report also suggested a saving to commissioners equivalent to £414,000 annually for a CCG covering 250,000 patients. These savings were specifically related to self-reported diversion of patients from GP appointments to self-care and from urgent care to e-consultation. Using similar methodology, the manufacturers of the 'babylon check' app claimed average savings of over £10/triage

compared with NHS111 by telephone, based on a higher proportion of patients being recommended to self-care²⁰.

Diagnostic accuracy

Eight studies reported at least some data on the diagnostic accuracy of symptom checkers. In spite of the diverse methods and comparisons in the included studies, almost all agreed that the diagnostic accuracy of symptom checkers was poor in absolute terms (e.g. in evaluating 'vignettes' designed to test knowledge of specific conditions, where the correct diagnosis was already known by definition) or relative to that of health professionals. In the most comprehensive evaluation, Semigran et al. evaluated 23 symptom checkers across 770 standardised patient evaluations⁵. Overall the correct diagnosis was made in 34% of cases (95% CI 31%-37%), although performance varied widely between symptom checkers, high and low acuity conditions and common and rare conditions. When the same authors compared the 23 symptom checkers with physicians using 43 vignettes, physicians were more likely to list the correct diagnosis first (out of three differential diagnoses) (72.1% vs. 34% p<0.001) as well as among the top three diagnoses (84.3% vs. 51.2% p<0.001)¹⁸.

The only exception to the rule was an evaluation carried out at a student health centre¹⁶. Using data from 59 participants who used the 24/7 WebMed system and who were subsequently treated at the health centre, the study found good agreement between chief complaint, 24/7 WebMed classification and provider diagnosis (kappa values 0f 0.89 to 0.94). This study differed from the others in using data from students rather than a general population sample. In addition, the students' complaints were generally common and uncomplicated, a scenario in which symptom checkers performed relatively well in the study by Semigran et al.¹⁸.

Accuracy of disposition (triage and signposting to appropriate services)

Six included studies reported on this outcome, all except one of which¹³ evaluated a 'general purpose' symptom checker. As with diagnostic accuracy, diverse methodologies and outcome measures were used.

The results overall presented a mixed picture but most studies indicated that symptom checkers were inferior and/or more cautious in their triage advice compared with doctors or

other health professionals. In their review of 23 symptom checkers, Semigran et al. found that the systems provided appropriate triage advice in 57% (95% CI 52% to 61%) of cases⁵. Performance varied across the systems evaluated, correct triage ranging from 33% to 78%. The NHS England pilot evaluation of four systems⁶ found that agreement with clinical experts varied from 30% to 95%, although the number of responses also varied, reducing the comparability of the results.

For abdominal pain, Berry et al. evaluated three symptom checkers and found that 33% of diagnoses were at the same level of urgency as physician diagnoses (emergency, non-emergency or self-care); 39% were diagnosed as more serious and 30% less serious than the physician's judgement¹³. A similar level of agreement between algorithm and clinician (39%) was reported by Poote et al.¹⁵, while the system evaluated by Nijland et al. advised patients to visit a doctor in 85% of cases, even when the symptoms were appropriate for self-care¹⁷.

The only studies to report clearly equal or superior accuracy of disposition using an automated system were the evaluations of Babylon check by the company that developed the system. Middleton et al. ¹⁴ reported that using patient vignettes, the app gave an accurate triage outcome in 88.2% of cases, compared with 75.5% for doctors and 73.5% for nurses. When vignettes were delivered by a medical professional rather than actors, the accuracy of Babylon check increased to over 90%. A later report looked at triage results obtained as part of the NHS England pilot evaluation, concluding that all of 74 referrals to urgent or emergency care were appropriate²⁰.

Impact on service use/diversion

Eight studies reported on this outcome, although one of them⁹ merely stated that it was not possible to assess the effect of the intervention (a web-based influenza triage system) on patients' use of health services.

The pilot evaluation of the webGP system reported that 18% of users planned to book an appointment but chose not to do so²³ In addition, 14% of users reported that they would have attended a walk-in centre or other urgent care service if they had not had access to the webGP system.

The NHS England pilot evaluation of four online/digital systems in different regions of England⁶ compared the recommendations of the digital systems with those of the NHS111 telephone service over a similar time period (the first months of 2017). Compared with the telephone service, the online and digital services directed a slightly higher proportion of patients to self-care (18% vs. 14%) and a lower proportion to other primary care services such as GPs, dental and pharmacy (40 vs. 60%). The manufacturer's data on the 'babylon check' app collected as part of the NHS England evaluation indicated that patients were more likely to be triaged to self-care by the app compared with NHS111 by telephone (40 vs. 14%)²⁰. This figure includes people who received information leaflets on self-care as well as those who were actively triaged. If the former group is excluded, the figures for the two services are similar (14% for NHS111 and 15.6% for 'babylon check'²⁰.

In their study of self-assessment for students attending a university health centre, Poote et al. found that the prototype system they studied was able to identify a proportion of cases that doctors considered appropriate for self-care, suggesting a potential to reduce service use¹⁵. Similarly, Little et al's RCT of a web-based symptom checker designed to support self-care for respiratory symptoms¹⁰ reported that patients in the intervention group had fewer contacts with doctors than the usual care control group despite having a longer duration of illness and more days with relatively severe symptoms. This was balanced by an increase in contacts with the NHS Direct telephone service (which preceded NHS 111) and it should be noted that the system under evaluation recommended people needing treatment to contact NHS Direct rather than go directly to a doctor. Finally, a study of young adults (students) found that intention to seek treatment for a hypothetical illness was stronger when the diagnosis was made with the aid of WebMD or Google than with no electronic aid²⁸.

Patient compliance with triage advice

Only two of the included studies reported specifically on patients' compliance (or intention to comply) with advice received. The NHS England pilot evaluation in four regions asked participants in two of those regions (Suffolk and London) what they intended to do based on the advice received⁶. It appears that the question was asked when patients were aware of the advice from the system but it was unclear whether the evaluation covered real or hypothetical cases. No quantitative data were provided but the report stated that in the Suffolk pilot,

'overall users would have followed the advice given'. However, those who were recommended to call 999 or attend an ED were more likely to seek advice from primary care or self-management. Similarly, in the London region there was generally good agreement between advice and intended action but patients recommended to call 999 or go to an ED indicated that they would seek advice from a GP. In a study of a web-based triage system in the Netherlands, 192 patients were asked about their intention to comply immediately after receiving advice from the system¹⁷. Thirty-five patients responded to a follow-up survey on actual compliance, of whom 20 (57%) reported that they had followed the advice.

Compliance was correlated with intention to comply, which in turn was correlated with the patient's attitude towards the advice received.

Equity and inclusion

Fourteen studies investigated the outcome of equity and inclusion or compared users and non-users. One study ¹⁰reported that patients who were classed as less deprived were more likely to agree to use "Internet Doctor" than decline participation, although no relationship was found between deprivation and results in this study or between e-Consult use and deprivation in another study ²². Association between e-consultation use and education levels was explored in a third study. Patients with low to medium levels of education tended to be motivated toward indirect e-consultation (which involves contact with a health professional via e-mail), mainly to reduce uncertainty ²⁴

Evidence from included studies suggests that users of e-consultation were more likely to be young⁶ ²¹⁻²³, employed¹⁷ ²¹ ²³ and female⁶ ¹⁷ ²² ²³ than non-users. One study also found a significantly larger use by white patients (78%) than other ethnicities²².

Risk of bias assessment

We assessed risk of bias in the two included RCTs¹⁰ ¹¹ using the Cochrane risk of bias tool. Thirteen studies⁹ ¹⁶ ¹⁷ ²¹ ²² ²⁴ ²⁷ ²⁹ ³¹ ³³ were assessed with the tool for cross-sectional and cohort studies and four (six publications⁵ ¹⁵ ¹⁸ ¹⁹ ³⁴ ³⁵) with the modified QUADAS tool. Seven grey literature reports and conference abstracts were not formally assessed for risk of bias⁶ ¹² ¹⁴ ²³ ²⁸ ³⁰. Identified limitations were extracted for all included studies.

Risk of bias results are presented in Appendix 2. With the possible exception of the two randomised trials, the included studies generally had at least a moderate risk of bias. However, the diverse designs and objectives of the studies made risk of bias difficult to assess in some cases with the available tools. Grey literature reports containing relevant data were included in the review but not formally assessed for risk of bias. Reports prepared by individuals with a commercial interest in a specific system and published without independent peer review^{14 23} should be treated with particular caution because of possible conflicts of interest.

Overall strength of evidence assessment/evidence map

The overall strength of evidence for key outcomes is summarised in Table 3. We found relatively strong evidence that the diagnostic accuracy of digital and online symptom checkers tends to be lower than that of health professionals; and that patients who have used these systems generally show high levels of satisfaction (mainly in non-comparative studies). Areas where evidence is lacking or inconsistent include clinical and cost-effectiveness, accuracy of disposition to appropriate services and patient compliance with advice received. For safety, we found no evidence of an increased risk with digital/online systems but the available evidence was weak.

Table 3: Overall strength of evidence by outcome

0-4	D-1	F-::1	C4	1 C
Outcome	Relevant studies	Evidence statement	Strength of	Comments g
C - C-4	17 . 11	N:1	evidence Weaker	D.ti.,
Safety	=Kellermann 2010 ⁹	No evidence of a difference in risk between	weaker	Rating changed from stronger based on study
		health professionals and symptom checkers		numbers and design to weaker because of low
	=Little 2016 ¹⁰			numbers of adverse events reported
	=Middleton 2016 ¹⁴			19.
				Do
	=Poote 2014 ¹⁵			w _n
	=Price 2013 ¹⁹			oac
CI: : 1	Madan 2014 ²³	T CC : 1 1 C	T7 1' ', 1	<u>v</u>
Clinical	-Little 2016 ¹⁰	Insufficient evidence to draw any firm	Very limited	from
effectiveness	?Madan 2014 ²³	conclusions		
Costs/cost-	+Babylon Health	Insufficient evidence to draw any firm	Inconsistent	http://bm
effectiveness	2017^{20}	conclusions		://b
	+/-Cowie 2018 ²²			mj.
	+Madan 2014 ²³		_	<u> </u>
Diagnostic	?Berry 2016 ¹²	Symptom checkers appear inferior to health	Stronger	Mainly for specific conditions or pre-prepared
accuracy	-Berry 2017 ³⁰	professionals in terms of diagnostic accuracy		vignettes <u>3</u>
	-Berry 2017 ¹³			CO
	-Price 2013 ¹⁹			0
	?Semigran 2015 ⁵			D D
	-Semigran			April 9
	2016 ¹⁸			
	=Sole 2006 ¹⁶			20
Disposition	=Babylon Health	Inconsistent findings on accuracy of	Inconsistent	Performanc variable between different systems
accuracy	2017^{20}	disposition		oy (
	-Berry 2017 ¹³			Jue
	=Middleton			by guest. Protected by cap
	201614			Pro
	?Nijland 2010 ¹⁷			ytec
	-Poote 2014 ¹⁵			i ded
	+/-Semigran			by
	20155	I .	1	

Outcome	Relevant studies	Evidence statement	Strength of	Comments $\frac{8}{7}$	
			evidence	743	
	+/-NHS			on	
	England 2017 ⁶			1	
Service	?Kellermann	Inconsistent findings on effects on service	Inconsistent	\ugust 2019. Downloaded from http://k	
use/diversion	20109	use		ust	
	+/-Little 2016 ¹⁰			20	
	+/- Poote 2014 ¹⁵			19.	
	?Carter 2018 ²¹			Do	
	?Cowie 2018 ²²			w <u>n</u>	
	+Madan 2014 ²³			oac	
	+/- NHS			ed	
	England 2017 ⁶	100		fro	
	+Babylon Health			m = m	
	2017 ²⁰	· O/A		l ttp	
	-Luger 2011 ³¹				
Compliance	?Nijland ¹⁷	No comparative data on compliance	Very limited	mjope	
	?NHS England			per	
	20176			ı.b	
Patient/carer	?Nagykaldi	Most studies report high rates of patient	Weaker	Few studies	ith comparator data
satisfaction	2010 ³³	satisfaction with symptom checkers and e-	1/1	mox	
	?Nijland ¹⁷	consultation systems generally		0	
	?Price 2013 ¹⁹			Α ·	
	+Yardley ¹¹			pril pril	
	?Carter 2018 ²¹			9,	
	?Cowie 2018 ²²			202	
	?Madan 2014 ²³			4 6	
	?NHS England			9 9	
	20176			pril 9, 2024 by guest.	
O . 11 1 . 1	?Lanseng 2007 ²⁷	1.00	1	: : : : : : : : : : : : : : : : : : :	1 1

Controlled studies in bold; = means no significant difference in outcomes; + means better outcome with symptom checker; +/- varying results within study; ? results difficult to interpret in comparative terms

Discussion

Main findings

The literature search identified 29 publications describing 27 studies that met the inclusion criteria. The overall strength of the evidence base varied between outcomes (Table 3), but in absolute terms the evidence is weak, being based largely on observational studies. A substantial component of grey literature of uncertain quality complicates the interpretation of the evidence. Interpretation of the evidence should also take into account risks of bias in individual studies.

We found little evidence to indicate whether or not digital and online symptom checkers are detrimental to patient safety. The studies that reported on the outcome were mostly short-term and involved relatively small samples and hence reported few or no adverse events. Some were limited to people with specific types of symptoms and others recruited from specific population groups not representative of typical users of urgent care services. This body of evidence should therefore be interpreted cautiously and not extrapolated to the possible impact of a nationally available digital urgent care service being used by millions of people annually.

The evidence on patient satisfaction with digital and online systems also had some limitations but these findings appear more likely to be generalisable. Study participants generally expressed high levels of satisfaction, albeit in uncontrolled studies. For example, in the NHS England pilot evaluation, 70–80% of users were satisfied with their experience at each of the pilot sites⁶. This evidence, together with the increasing reliance on digital technology in all areas of life, suggests that any national digital urgent care service may be popular and well-used.

Digital and online systems have yet to achieve a high level of accuracy in the diagnosis of specific conditions. This finding applies both to 'general purpose' symptom checkers and to those limited to particular conditions. Although the evidence was classified as relatively strong, several caveats should be applied. Some of the included studies did not recruit representative populations and others were based on standardised vignettes rather than real-

world data. In addition, studies that compared symptom checkers with health professionals tended to use the doctors' clinical diagnosis as the reference standard, which would bias the comparison in favour of the health professionals.

Accuracy of signposting of patients to the most appropriate level of service is closely related to diagnostic accuracy, but results for this outcome were inconsistent between studies. In general, algorithm-based triage tended to be more risk-averse than that of health professionals, with 85% of respondents being advised to visit their doctor in one study¹⁷. While there is considerable uncertainty about the magnitude of the effect, a national digital urgent care service could result in considerable numbers of patients receiving inappropriate advice to visit the ED or request an urgent GP appointment. Middleton and colleagues¹⁴ claimed that the 'babylon check' app had a high degree of triage accuracy for vignettes compared with health professionals, but this non-peer-reviewed report requires further validation.

We also found inconsistent evidence on effects on service use. There was some indication that symptom checkers can influence the pattern of service use but the magnitude and direction of the effect varied between studies. Patients' reactions to online triage advice and whether they follow the advice or seek further help or information would have implications for service use but we found limited evidence for this outcome. Preliminary findings from the NHS England evaluation suggest that patients may be more likely to seek further advice for more urgent conditions⁶ but further confirmation is required.

Over half of the included studies considered equity and inclusion issues either directly or by comparing users and non-users of digital triage systems. Not surprisingly, studies revealed a clear consensus that younger and more highly educated people are more likely to use these services while older and less educated patients are more likely to prefer telephone or face-to-face contact. This could have implications for health equity if urgent care pathways prioritise (or appear to prioritise) requests originating from digital sources. Problems have arisen in primary care because patients using e-consultation systems to request an appointment following online triage may be seen more quickly than those contacting the practice by telephone.

Strengths and limitations

This systematic review was undertaken on a short timescale using a relatively large team of experienced researchers, including both methodological and topic experts. We performed a rigorous search of the literature including reference checking and citation searching. Rather than a conventional highly sensitive search (which would have resulted in inefficiencies in the screening process), we combined an initial focused search with subsequent rounds of follow-up searching, including searches for named symptom checker systems. We assessed risk of bias in individual studies using a variety of appropriate checklists as well as summarising the overall strength of evidence for key outcomes (Table 3).

The heterogeneous and descriptive nature of the included studies meant that meta-analysis was not feasible for any of the outcomes of interest. Our narrative synthesis approach used a mixture of description and tabulation to summarise the evidence for each of the pre-specified outcomes of interest. This was a review of published (including non-peer-reviewed) literature and the coverage of systems is not exhaustive; for example, we did not extract data from websites. We also did not carry out any original analyses of raw data even where such data were available. The timing of the review meant that final results of NHS England's pilot evaluation were not available to us. We were able to make use of a draft report that was published online⁶ but we acknowledge that the findings of the final evaluation report, when available, will supersede those of the 2017 draft.

The review inclusion criteria were relatively broad and findings from symptom checker systems for specific conditions may not be applicable to more general systems and vice versa. We have also included studies of symptom checkers as part of electronic consultation systems in general practice, which again represents a slightly different setting from a general 'digital 111' service, and this should be kept in mind when interpreting the results.

Implications for service delivery and research

The implications of this systematic review for service delivery should be considered in the context that a decision has already been taken to introduce a 'digital 111' service and implementation of the service is in progress. Achieving a high level of diagnostic accuracy will be key to the success of a 'digital 111' service. Failure to provide an accurate diagnosis

may result in outcomes including patient dissatisfaction and unwillingness to use the service again; increased use of other urgent and emergency care services; and possible risks to patient safety (although the cautious approach characteristic of most existing systems may help to mitigate this).

The studies included in the review suggest a high level of uncertainty about the impact of 'digital 111' on the urgent care system and the wider healthcare system. Some of these uncertainties can be addressed by research and data collection but the health service may need to respond to short-term increases (or decreases) in demand and/or shifts from one part of the system to another. This may increase pressure on the system, at least in the short-term. In the longer-term, if usage of the 111 telephone service decreases as planned, there may be opportunities to reconfigure the workforce to support the integrated urgent care agenda.

Based on the areas of limited evidence identified by the review, priorities for research (in addition to ongoing collection of data to monitor usage and safety of the 'digital 111' service) include studies to compare the performance of different systems directly; rigorous economic evaluations based on real-world data; research to investigate the pathways followed by patients using the service; evaluation of systems designed for childhood illnesses; and investigation of the possible role of behaviour change theory in the development and implementation of symptom checkers.

Ethical approval

Ethical approval was not required for this work

Funding

This report presents independent research funded by the National Institute for Health Research (NIHR) Health Services & Delivery Research Programme (project number HSDR16/47/17). The funding programme approved the review protocol but had no role in the collection, analysis and interpretation of the data, the writing of this paper or the decision to submit the paper for publication. The views and opinions expressed are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

Data sharing

No new data have been created in the preparation of this report and therefore there is nothing available for access and further sharing. All queries should be submitted to the corresponding author.

References

- 1. NHS England. NHS 111 minimum data set 2018-19 2018 [Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/nhs-111-minimum-data-set-nhs-111-minimum-data-set-2018-19/accessed 29 October 2018.
- 2. NHS England. Five year forward view. Leeds: NHS England, 2014.
- 3. NHS England. Next steps on the NHS Five Year Forward View. Leeds: NHS England, 2017.
- 4. Turner J, O'Cathain A, Knowles E, et al. Impact of the urgent care telephone service NHS 111 pilot sites: a controlled before and after study. *BMJ Open* 2013;3(11):e003451. doi: 10.1136/bmjopen-2013-003451
- 5. Semigran HL, Linder JA, Gidengil C, et al. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015;351:h3480.
- 6. NHS England. NHS111 online evaluation. Leeds: NHS England, 2017.
- 7. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687. doi: 10.1136/bmj.g1687
- 8. Baxter S, Johnson M, Chambers D, et al. The effects of integrated care: a systematic review of UK and international evidence. *BMC Health Serv Res* 2018;18(1):350. doi: 10.1186/s12913-018-3161-3
- 9. Kellermann AL, Isakov AP, Parker R, et al. Web-based self-triage of influenza-like illness during the 2009 H1N1 influenza pandemic. *Annals of Emergency Medicine* 2010;56(3):288-94.e6.
- Little P, Stuart B, Andreou P, et al. Primary care randomised controlled trial of a tailored interactive website for the self-management of respiratory infections (Internet Doctor). [Erratum appears in BMJ Open. 2017 Mar 21;7(3):e009769corr1; PMID: 28325861]. BMJ Open 2016;6(4):e009769.
- 11. Yardley L, Joseph J, Michie S, et al. Evaluation of a Web-based intervention providing tailored advice for self-management of minor respiratory symptoms: exploratory randomized controlled trial. *Journal of Medical Internet Research* 2010;12(4):e66.
- 12. Berry AC, Berry BB, Nakshabendi R, et al. Evaluation of Accuracy Between Online Symptom Checkers for Diagnosis of Gastrointestinal Symptoms from MKSAP Clinical Vignette Board Review Questions. *Gastroenterology* 2016;150(4):S849-S50. doi: 10.1016/s0016-5085(16)32869-4
- 13. Berry AC, Cash BD, Mulekar MS, et al. Symptom checkers vs. Doctors, the ultimate test: a prospective study of patients presenting with abdominal pain. *Gastroenterology* 2017;152(5):S852-S53. doi: 10.1016/s0016-5085(17)32937-2
- 14. Middleton K, Butt M, Hammerla N, et al. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. London: Babylon Health, 2016.
- 15. Poote AE, French DP, Dale J, et al. A study of automated self-assessment in a primary care student health centre setting. *Journal of Telemedicine & Telecare* 2014;20(3):123-7.
- 16. Sole ML, Stuart PL, Deichen M. Web-based triage in a college health setting. *Journal of American College Health* 2006;54(5):289-94.

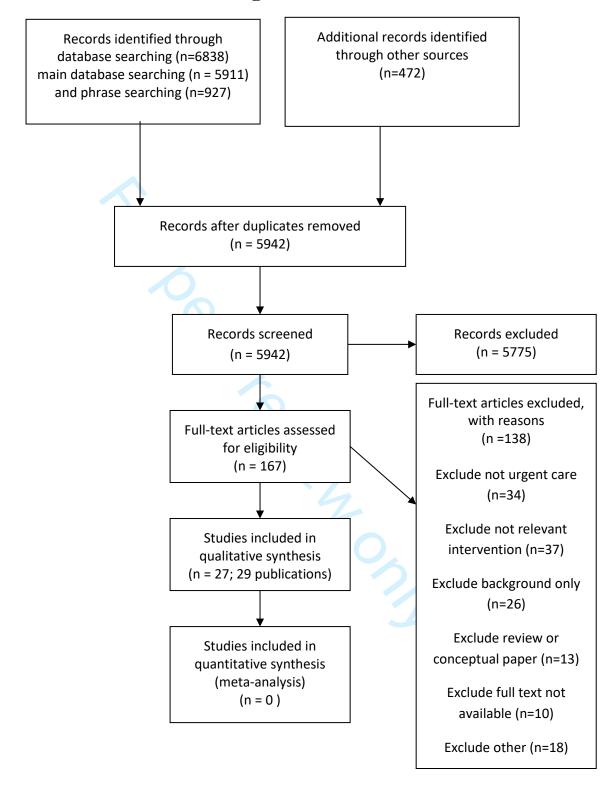
- 17. Nijland N, Cranen K, Boer H, et al. Patient use and compliance with medical advice delivered by a web-based triage system in primary care. *Journal of Telemedicine and Telecare* 2010;16(1):8-11. doi: 10.1258/jtt.2009.001004
- 18. Semigran HL, Levine DM, Nundy S, et al. Comparison of Physician and Computer Diagnostic Accuracy. *JAMA Intern Med* 2016;176(12):1860-61. doi: 10.1001/jamainternmed.2016.6001
- 19. Price RA, Fagbuyi D, Harris R, et al. Feasibility of web-based self-triage by parents of children with influenza-like illness: A cautionary tale. *JAMA Pediatrics* 2013;167(2):112-18.
- 20. Babylon Health. NHS111 powered by babylon: outcomes evaluation. London: Babylon Health, 2017.
- 21. Carter M, Fletcher E, Sansom A, et al. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. *BMJ Open* 2018;8(2):e018688.
- 22. Cowie J, Calveley E, Bowers G, et al. Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. *International Journal of Environmental Research and Public Health* 2018;15(896)
- 23. Madan A. WebGP: the Virtual general practice. London: Hurley Group, 2014.
- 24. Nijland N, van Gemert-Pijnen J, Boer H, et al. Increasing the use of e-consultation in primary care: Results of an online survey among non-users of e-consultation. *International Journal of Medical Informatics* 2009;78(10):688-703. doi: 10.1016/j.ijmedinf.2009.06.002
- 25. Backman AS, Lagerlund M, Svensson T, et al. Use of healthcare information and advice among non-urgent patients visiting emergency department or primary care. *Emergency Medicine Journal* 2012;29(12):1004-06.
- 26. Joury AU, Alshathri M, Alkhunaizi M, et al. Internet Websites for Chest Pain Symptoms Demonstrate Highly Variable Content and Quality. *Academic Emergency Medicine* 2016;23(10):1146-52.
- 27. Lanseng EJ, Andreassen TW. Electronic healthcare: a study of people's readiness and attitude toward performing self-diagnosis. *International Journal of Service Industry Management* 2007;18(3-4):394-417. doi: 10.1108/09564230710778155
- 28. Luger TM, Suls J. Online health information and intentions to seek healthcare. *Psychosomatic Medicine* 2011;73 (3):A59.
- 29. North F, Varkey P, Laing B, et al. Are e-health web users looking for different symptom information than callers to triage centers? *Telemedicine Journal & E-Health* 2011;17(1):19-24.
- 30. Berry AC, Berry NA, Wang B, et al. Symptom checkers versus doctors: A prospective, head-to-head comparison for GERD vs. Non-GERD Cough. *American Journal of Gastroenterology* 2017;112 (Supplement 1):S190. doi: http://dx.doi.org/10.1038/ajg.2017.299
- 31. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. *Journal of Medical Internet Research* 2014;16(1):e16.
- 32. Marco-Ruiz L, Bones E, de la Asuncion E, et al. Combining multivariate statistics and the thinkaloud protocol to assess Human-Computer Interaction barriers in symptom checkers. *Journal of Biomedical Informatics* 2017;74:104-22.
- 33. Nagykaldi Z, Calmbach W, Dealleaume L, et al. Facilitating patient self-management through telephony and web technologies in seasonal influenza. *Informatics in Primary Care* 2010;18(1):9-16.
- 34. Fraser HSF, Clamp S, Wilson CJ. Limitations of Study on Symptom Checkers. *JAMA Internal Medicine* 2017;177(5):740-41.
- 35. Mehrotra A, Semigran HL, Levine DM, et al. Limitations of Study on Symptom Checkers. *JAMA Internal Medicine* 2017;177(5):741-41.



Identification

Screening

PRISMA 2009 Flow Diagram





Appendix 1: Highly focused specific MEDLINE search strategy (adapted for other databases)

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present> Search Strategy:

- 1. (symptom checker or symptoms checker or symptom checkers or symptoms checkers).tw.
- 2. ("self diagnosis" or "self referral" or "self triage" or "self assessment").tw. (10403)
- 3. TRIAGE/
- 4. 2 or 3
- 5. (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone).tw.
- 6. 4 and 5
- 7. ("online diagnosis" or "web based triage" or "electronic triage" or etriage).tw.
- 8. 1 or 6 or 7

Risk of bias results for randomised trials

Appendix 2: Risk of bias tables

Study ID Random sequence generation Low risk Selection and attrition bills Reporting and other his	Cl 4 TP41 -	D.f	C-14:		D 1 -4h h
*Reference Little 2016¹² *Allocation concealment •Low risk *Blinded assessment of primerry care records *Incomplete outcome data* *Low risk *Incomplete outcome data* *Low risk *Anything else, ideally prespecified •Low risk *Yardley (2010) *Reference Yardley 2010¹³ *Allocation concealment •Low risk *Incomplete outcome data* *Unclear *Allocation concealment •Low risk *Incomplete outcome data* *Incomplete outcome data* *Anything else, ideally prespecified •Unclear	Short Title	Reference	Selection and performance bias	Detection and attrition bias	Reporting and other bias
Allocation concealment • Low risk Low risk Low risk Low risk	Little (2016)	Study ID		Blinding of outcome $\vec{\varphi}$	
Allocation concealment • Low risk Blinding of participants and personnel* • Unclear Pardley (2010) • Reference Yardley 2010 ¹³ Allocation concealment • Low risk Anything else, ideally prespecified • Low risk Blinding of participants and personnel* • Unclear • Unclear Blinding of outcome assessment* • Unclear Allocation concealment • Low risk Incomplete outcome data • Unclear Anything else, ideally prespecified • Anything else, ideally prespecified			• Low risk	assessment*	• Unclear
Allocation concealment • Low risk Low risk Low risk Low risk		Little 2016 ¹²		• Low risk	
Allocation concealment • Low risk Low risk Low risk Low risk				Blinded assessment of primary	
		•	Allocation concealment	care records $\frac{\alpha}{2}$	Anything else, ideally
		4	• Low risk	rom	prespecified
				- htt	• Low risk
				Incomplete outcome data	
			Blinding of participants and	• Low risk	
				Эре	
			_	1.br	
			/6	nj.o	
	Yardley (2010)	Study ID	Random sequence generation	Blinding of outcome	Selective reporting
	101010)	_		assessment*	
			Low risk	• Unclear	Chercui
		Turatey 2010		Sileicai <u>si</u>	
			Allogation concealment	200	Anything else ideally
				Incomplete outcome detakt	
Blinding of participants and personnel* • Low risk			LOW FISK	_	
Blinding of participants and personnel* • Low risk				• Low risk	• Low risk
Personnel* • Low risk					
personnel* • Low risk				rote	
• Low risk			_	ecte	
			• Low risk	් . ව	
				y cc	

Risk of bias results for cohort/cross-sectional studies

Risk of bias res	sults for cohort/cross-sectional studies	BMJ Open	6/bmjopen-2018-027743 on 1 Au
Reference	Questions 1-4	Questions 5-7	Questions 8-10g
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Unclear
Backman A-S et al. 2012 ³⁰	The aims refer to "non-urgent" but the information is sought prior to visiting ED.	• No	"Health care information use in the past"
		6. Did the study	9. Were outcome assessors blinded?
	2. Was the study population clearly specified and	examine exposure	• Not applicable
	defined?	levels?	
	• Yes	• Yes	;//b
	4	Health advice seeking	10. Were confounders adjusted for?
			• Yes
	3. Was the participation rate at least 50%?		To some extent: participant and physician attributes
	• Yes	7. Were exposure	assessed for influence on the results.
	79%	measures clearly	η/ ο
		defined?	ă ≽
		• Unclear	P/) / pril :
	4. Were all the subjects selected or recruited from	Measures are vague, e.g.	, <u>o</u>
	the same or similar populations?	"previous use" of	024
	• Yes	information Also,	by (
	Primary care and ED attendees	discriminating between	on April 9, 2024 by guest.
		types of information	st. Pro
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Yes g
Carter 2018 ²⁶		• No	Attitudes and experiences of practice staff and

6/bmjopen-2018-

			<u> </u>
			patients on web P.
	2. Was the study population clearly specified and	6. Did the study	3 on
	defined?	examine exposure	9. Were outcome assessors blinded?
	• Yes	levels?	• Not applicable
	GPs, practice staff and their patients at 6 practices in	Not applicable	st 2
	Devon		019
			10. Were confounders adjusted for?
	' O.	7. Were exposure	• Not applicable€
	3. Was the participation rate at least 50%?	measures clearly	oa d.
	• No	defined?	ed fr
	Postal survey only had response rate of 35.1% but also	Not applicable	rom
	GPs judgement of webGP requests and 5GPs and 5		http
	administrators were interviewed.	6)://b
			njoj
		evien.	oe n
	4. Were all the subjects selected or recruited from		bmj.
	the same or similar populations? • Yes	. 6/7	.cg
			0
	GPs, practice staff and their patients at 6 practices in Devon		n Ar
	Devon		oaded from http://bmjopen.bmj.com/ on April 9,
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	4 v
Cowie 2018 ²⁷		• No	• Yes by gues
			
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• No btected
	• Yes	examine exposure	ied d
		levels?	у
		• No	10. Were confounders adjusted for?

	3. Was the participation rate at least 50%?		• Yes
	• No		743
	No for patient surveys	7. Were exposure	0 2
		measures clearly	<u> </u>
		defined?	igus
	4. Were all the subjects selected or recruited from	 Not applicable 	20
	the same or similar populations?		119.
	• Yes		• Yes
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes $\frac{\alpha}{2}$
Joury et al.	100	• No	Scores used for geadability, popularity, content and
2016 US ³¹	0,6		quality
	2. Was the study population clearly specified and)://b
	defined?	6. Did the study	jo
	Not applicable	examine exposure	9. Were outcome assessors blinded?
		levels?	• Not applicable
		 Not applicable 	J. 00
	3. Was the participation rate at least 50%?		Ď.
	Not applicable		10. Were confounders adjusted for?
		7. Were exposure	• Unclear $\frac{\dot{\nabla}}{2}$
		measures clearly	9,
	4. Were all the subjects selected or recruited from	defined?	2022
	the same or similar populations?	 Not applicable 	4 by
	Not applicable		• Unclear pril 9, 2024 by gues
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcond measures clearly defined?
• Reference	• Unclear	justification provided?	• Not applicable
Kellermann		• Not applicable	, if ed
2010 11			by
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?

BMJ Open

	defined?	6 Did the standar	• Not applicable
	• Unclear	6. Did the study	Not applicable
		examine exposure	ω ο
	Patients with influenza-like illness in US that accessed	levels?	10 XX
	one of 2 websites http://www.flu.gov and www.H1N2	 Not applicable 	10. Were confounders adjusted for?
	ResponseCenter.com		• Not applicable
			20
		7. Were exposure	19.
	3. Was the participation rate at least 50%?	measures clearly	Ρον
	Not applicable	defined?	Nnla
		 Not applicable 	oa d
			ed of
	4. Were all the subjects selected or recruited from		rom
	the same or similar populations?		2019. Downloaded from http://bmjopen.bm
	• Unclear		p:///
	Only counted web hits, no demographic data available		<u>, </u>
	on patients. No data on usage of algorithm by	81.	ope Ope
	clinicians or call centers.		n.b
		10,	ற். c
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	
Lanseng &		• No	• Yes On April 99
Andreassen			⊖ 55 0) 11tt = :: • • • • • • • • • • • • • • • • •
2007	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
Norway ³²	defined?	6. Did the study	Wara outcome assassars blinded?
Norway	• Yes	examine exposure	
	168	levels?	• No Quest
			st. F
	2 11/ 41	• No	10. Were conformaters adjusted for?
	3. Was the participation rate at least 50%?	Readiness	1 4
	• Unclear		• Unclear
			• Unclear $\frac{6}{9}$
ı		7. Were exposure	ı O

		BMJ Open	6/bmjopen-2018-
	4 Work all the subjects selected on reconsited from	maggung gloodly	2018-02:
	4. Were all the subjects selected or recruited from	measures clearly defined?	7743
	the same or similar populations? • Yes	• Not applicable	3 on 1 <i>t</i>
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes
Luger et al. 2014^{23}		• No	019. Do
2017	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable
	• Yes	examine exposure	e d
	\sim	levels?	from
	(0)	• No	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?		• Unclear
	• Unclear		• Unclear //bmjopen.bmj.com
		7. Were exposure	per
		measures clearly	1.bm
	4. Were all the subjects selected or recruited from	defined?	nj. cc
	the same or similar populations?	Not applicable	
	• Yes		on Apri
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
 Reference 	• Yes	justification provided?	• Not applicable
Marco-Ruiz et		• No	4 by
al. 2017			, âm
Norway ²⁴	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable o
	• No	examine exposure	tected
		levels?	
		• No	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?		• Unclear ପ୍ରି

			<u> </u>
	• Yes		27743 on 1 August 2019
	53%		43 0
		7. Were exposure	on 1
		measures clearly	≥ P
	4. Were all the subjects selected or recruited from	defined?	l gu
	the same or similar populations?	 Not applicable 	st 20
	• Unclear)19. E
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes
Nagykaldi	'	• Not applicable	Web hits on customised practice website influenza
2010 ²⁵			self-management webpages. Downloads of self-
	2. Was the study population clearly specified and		management in ueza toolkit. Completion of Iflueza
	defined?	6. Did the study	self-triage modifie sessions. Volume of calls to
	• Yes	examine exposure	telephone hotlings. Qualitative feedback from patient
	Study population was patients from 12 primary care	levels?	on statisfaction with and utility of self-management
	practices in US.	 Not applicable 	websites and telephone hotline. Qualitative feedback
			from clinicians ground their involvement and their
		\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	perceptions of patient self-management techniques.
	3. Was the participation rate at least 50%?	7. Were exposure	1 1 01 0
	• Not applicable	measures clearly	Apri
		defined?	9. Were outcome assessors blinded?
		 Not applicable 	• Not applicable
	4. Were all the subjects selected or recruited from		4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
	the same or similar populations?		א פר
	• Yes		10. Were confounders adjusted for?
	All participants were patients from 12 primary care		• Not applicable
	practices that accessed customised practice website or		otec
	telephone helpline		∺ ed
			by
	•		• Not applicable Protected by copyright.
			yrigi
			t.

		BMJ Open	6/bmjopen-201
• Reference Nijland 2009 ²⁹	1. Was the research question clearly stated? • Yes	5. Was a sample size justification provided? • No	8. Were outcome measures clearly defined? • Yes 9 1
	 2. Was the study population clearly specified and defined? Yes 3. Was the participation rate at least 50%? Unclear 4. Were all the subjects selected or recruited from the same or similar populations? Yes 	 6. Did the study examine exposure levels? Not applicable 7. Were exposure measures clearly defined? Not applicable 	9. Were outcome assessors blinded? • No 10. Were confounders adjusted for? • Yes Methods not very clearly reported but appears to be multiple regression
• Reference Nijland 2016 ¹⁹	1. Was the research question clearly stated? • Yes	5. Was a sample size justification provided?No	8. Were outcome measures clearly defined? • Yes 9 Apple
	2. Was the study population clearly specified and defined?Yes	6. Did the study examine exposure levels?Not applicable	9. Were outcome assessors blinded? • No 10. Were confounders adjusted for?
	 3. Was the participation rate at least 50%? No Low participation rate in survey relative to users of triage system (though unclear how many were invited to participate) 	7. Were exposure measures clearly defined?	• Unclear Protected by copyrig

			Ó
		Not applicable	27743 on 1
	4. Were all the subjects selected or recruited from		o n
	the same or similar populations?		
	• Yes		August
	165		ust 2
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes
North et. al.		• Not applicable	Wn .
2011^{34}		тист прристего	loac
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable
	• Yes	examine exposure	at at
		levels?	tp://
		• Yes	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?	Self-exposure	• Unclear
	Not applicable		Some discussion potential confounders.
		10,	₽.
		7. Were exposure	com/ on April 9
	4. Were all the subjects selected or recruited from	measures clearly	on on
	the same or similar populations?	defined?	Αp
	• Not applicable	• Not applicable	rii 9
	Two applicable	Tvot applicable	, 202
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Not applicable
Sole 2006 ¹⁸	"The primary purpose of this study was to identify and	• No	Jest
	describe the demographic profile of students who used		Pr
	the newly implemented Web-based triage system. A		9. Were outcome assessors blinded?
	secondary purpose was to compare Web-based triage	6. Did the study	• Not applicable
	diagnoses to the diagnoses made in clinic for a subset	examine exposure	<u> </u>
	and the second of the analytics of many in current for a subset	- inposure	CO
			J Sy Figi

	, , , , , , , , , , , , , , , , , , , ,		0 10 777 a N 7 7 7 1 2 2 2
of stud	dents who requested appointments"	levels?	10. Were confounders adjusted for?
		• Yes	• Not applicable
			on
2. Wa	s the study population clearly specified and		> -
define		7. Were exposure	August 2019.
• Yes	· · ·	measures clearly	ist A
	ets who used the web based triage over a four	defined?	201
	nts who used the web based triage over a four		
	implementation period (1290 students). Then of	• Yes	Oow
those	students, those who requested an appointment via		vnlo
email	(143 students), then of those 59 who attended the		ă d
health	centre after requesting an email appointment.		ed d
	20		Downloaded from http://bmjopen.bmj.com
			- htt
3. Wa	s the participation rate at least 50%?		p://
	applicable		<u>a</u> ,
			9
			yn .b
4 We	re all the subjects selected or recruited from		<u>ä</u> .
	_		on Co
	me or similar populations?		٥ کې
• Yes			Ď Þ
			Apri

Risk of bias results for diagnostic studies

Reference	Questions 1 to 4	Questions 5 to 8	Questions 9 to 11 g
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical Information?
•	• No	verification	• Yes $\overset{\smile}{\mathfrak{Q}}$
Reference	Study participants were all patients registered at a student	avoided?	fe d
Poote	health centre in England attending with new acute	Not applicable?	by a
			10. Were uninterpretable results reported?

		BMJ Open	6/bmjop
)en-2018-
2014 17	symptoms. If the self-assessment triage system was only for students to be representative the study population would have needed to include range of student health	6. Was the	• Not applicable • Not applicable
	centres in different areas. If the system was for any UK	reference standard	11. Were withdrawals from the study explained?
	general practices the study population would have needed	independent of the	
	to include patients of all ages, ethnicity, gender etc from a	index test?	st 2
	range GP practices in different areas.	• Unclear	019
		Patients took the	D
	10.	assessment from self-	w _n l.
	2. Acceptable reference standard?	triage through to	oad
	• Yes	their GP	ed fr
	700	consultation.	rom
	3. Acceptable delay between tests?		ittp://
	• Yes	7. Index test results	bm _j ;
		blinded?	ор er
		• No	1.bm
	4. Partial verification avoided?	Patients took the). CC
	• Yes	assessment from self-	/mc
	All patients that completed self-triage also had a GP	triage through to	on /
	consultation where the GP rated the urgency of their	their GP	Pri
	consultation.	consultation.	gust 2019. Downloaded from http://bmjopen.bmj.com/ on April 9, 2024 by guest. Protected • Yes
		8. Reference	by g
		standard results	Jues
		blinded?	ن ت
		• Yes	rote
			cted
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?

		BMJ Open	6/bmjopen-2018-0
	• No	verification	
Reference	SORT was only trialled in 2 Emergency Departments in	avoided?	• Yes 277743
Price 2013	US, a larger range would be needed for a representative	• Not applicable?	on Signature
20	spectrum. Also, patients were from ED not home so	Tiot applicable.	10. Were uninterpretable results reported?
	potentially sicker patients in the sample.		
	potentially steker patterns in the sample.	6. Was the	• Not applicable
	A	reference standard	019
	2. Acceptable reference standard?	independent of the	11. Were withdraw as from the study explained?
	• Yes	index test?	• No \$
	Sensitivity of SORT for kids algorithm in identifying the	• Yes	oad
	need for ED care was based on an explicit gold standard:		e d
	documented evidence that the child received 1 or more of		rom
	5 ED-specific interventions.	7. Index test results	htt
		blinded?	p://t
		• Yes	om _j c
	3. Acceptable delay between tests?	C//	loaded from http://bmjopen.bmj.com/ on April 9
	• Yes		, b
		8. Reference	<u></u> co
		standard results	m/ c
	4. Partial verification avoided?	blinded?	on »
	• Yes	• Yes	Pri:
			, , ,
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?
•	• Unclear	verification	• Yes
Reference	There were 45 standardised patient vignettes which were	avoided?	This is the clinical information that would be supplied by
Semigran	divided into three levels of triage urgency and included	• Not applicable?	the patient which man or may not differ from the
2015 4	more and less common conditions. It is not clear how		information given by The vignette.
	closely this replicates the spectrum of conditions that	6 Wag tha	[#548 Semigran 2015, pdf] Page 8: ion of the true clinical
	people use symptom checkers for.	6. Was the	accuracy of symptom checkers.33 Some standardized
		reference standard independent of the	patient vignettes constained specific clinical language (for
		maependent of the	Dyrig

		BMJ Open	6/bmjopen-20
	2. Acceptable reference standard?	index test?	example, mouth ulcers, tonsils with exudate), and actual
	• Yes	• Yes	patients with the sante condition might struggle with the
	[#548 Semigran 2015.pdf] Page 2: The source for each		words to use to describe their symptoms or use diferent
	vignette also provided the associated correct diagnosis.		terms. Therefore, oupanalysis represents an indirect
	rightene unso provided the dissociated correct anaghesis.	7. Index test results	assess- ment of how well symptom checkers would perform
		blinded?	with actual patients \aleph
	3. Acceptable delay between tests?	• Yes	00
	• Not applicable		D. D.
			10. Were uninterpretable results reported?
		8. Reference	• Yes
	4. Partial verification avoided?	standard results	[#548 Semigran 2019 pdf] Page 3: ns for diagnosis and
	Not applicable	blinded?	triage was high (Colen's κ 0.90). In some cases we could
		• Yes	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 symp- tom in the vignette. To
		C1	avoid penalizing the se symptom checkers, we referred to
			standardized patient vignettes that successfully yielded an
		(0)	output as "standardized patient evaluations."
) m
		•	on /
			11. Were withdraw is from the study explained?
			• Not applicable 9
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?
•	• Unclear	verification	• Yes ©
Reference	There were 45 standardised patient vignettes which were	avoided?	The physicians and t \mathbf{n} e symptom checkers used the same
Semigran	divided into three levels of triage urgency and included	• Not applicable?	vignettes 7
2016 8	more and less common conditions. It is not clear how		rtect
	closely this replicates the spectrum of conditions that		ēd
	people use symptom checkers for.	6. Was the	10. Were uninterpretable results reported?

	BMJ Open	6/bmjopen-2018-
	reference standard	• Not applicable
	independent of the	743
2. Acceptable reference standard?	index test?	on 1
• Yes	 Not applicable 	11. Were withdrawals from the study explained?
		• No gu
		It is unclear why the physicians chose to solve the specific
3. Acceptable delay between tests?	7. Index test results	vignettes $\vec{\Theta}$
Not applicable	blinded?	Do
	• Yes	Downloaded from
		bad
4. Partial verification avoided?		ed f
• No	8. Reference	rom
There was a total of 234 physicians involved in the study	standard results	
and of the 45 vignettes, each was solved by at least 20	blinded?	p://k
physicians but it is not clear why they chose the specific	• Yes	mjc
vignettes to solve.	Chi	http://bmjopen.bm

j.com/ on April 9, 2024 by guest. Protected by copyright.



PRISMA 2009 Checklist

·		01	
Section/topic	#	Checklist item -0277 43	Reported on page #
TITLE		On	
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT		- Just	
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION		hioac	
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
8 Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, in reference, comparisons, outcomes, and study design (PICOS).	5
METHODS	,		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
7 Information sources 8	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplion tensor and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
2 Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I²) for each meta analysis. http://bmjopen.bmj.com/site/about/guidelines.xhtml	N/A



PRISMA 2009 Checklist

4		Page 1 of 2	
Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
10 Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7
RESULTS		——————————————————————————————————————	
14 Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reach stage, ideally with a flow diagram.	9
17 Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOs, follow-up period) and provide the citations.	10-16
19 Risk of bias within studies 20	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Appendix 2
Results of individual studies 23	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summare data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	17-22
24 Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	22-23
29 DISCUSSION		D Ti.	
30 3∱ Summary of evidence 32	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	26-27
33 Limitations 34	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	28
35 Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implication of the research.	28-29
FUNDING		· · ·	
39 Funding 40	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	29

43 doi:10.1371/journal.pmed1000097

BMJ Open

Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review

Journal:	BMJ Open
Manuscript ID	bmjopen-2018-027743.R1
Article Type:	Research
Date Submitted by the Author:	02-Apr-2019
Complete List of Authors:	Chambers, Duncan; The University of Sheffield, ScHARR Cantrell, Anna; The University of Sheffield, ScHARR Johnson, Maxine; The University of Sheffield, ScHARR Preston, Louise; The University of Sheffield, ScHARR Baxter, Susan; The University of Sheffield, ScHARR Booth, Andrew; The University of Sheffield, ScHARR Turner, Janette; The University of Sheffield, ScHARR
Primary Subject Heading :	Health services research
Secondary Subject Heading:	Diagnostics
Keywords:	urgent care, symptom checkers, systematic reviews

SCHOLARONE™ Manuscripts

Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review

Duncan Chambers¹, Anna Cantrell¹, Maxine Johnson¹, Louise Preston¹, Susan K Baxter¹, Andrew Booth¹ and Janette Turner¹

¹School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, Sheffield S1 4DA, UK

*Correspondence to Duncan Chambers: d.chambers@sheffield.ac.uk

Contributor/guarantor information:

DC contributed to the planning (project co-ordination and protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. AC contributed to the planning (protocol development), conduct (information retrieval, study selection, data extraction and quality assessment) and reporting (report writing) of the study. MJ contributed to the planning (protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. LP contributed to the planning (protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. SB contributed to the planning (protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. AB contributed to the planning (protocol development), conduct (information retrieval and study selection) and reporting (report writing) of the study. JT contributed to the planning, conduct and reporting of the study by providing expert topic advice at all stages. All the authors contributed to the study conception and design (protocol development), acquisition of data (study selection and data extraction) and analysis or interpretation of data (writing sections and/or commenting on drafts of the report). Duncan Chambers is the guarantor for this work. The corresponding author attests

that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

None of the authors have any competing interests

Copyright

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a non-exclusive worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.



Abstract

Objectives: In England, the NHS111 service provides assessment and triage by telephone for urgent health problems. A digital version of this service has recently been introduced. We aimed to systematically review the evidence on digital and online symptom checkers and similar services.

Design: Systematic review.

Data sources: We searched MEDLINE, EMBASE, the Cochrane Library, CINAHL, HMIC (Health Management Information Consortium), Web of Science and ACM Digital Library up to April 2018, supplemented by phrase searches for known symptom checkers and citation searching of key studies.

Eligibility criteria: Studies of any design that evaluated a digital or online symptom checker or health assessment service for people seeking advice about an urgent health problem.

Data extraction and synthesis: Data extraction and quality assessment (using the Cochrane Collaboration version of QUADAS for diagnostic accuracy studies and the National Heart, Lung and Blood Institute tool for observational studies) -were done by one reviewer with a sample checked for accuracy and consistency. We performed a narrative synthesis of the included studies structured around pre-defined research questions and key outcomes.

Results: We included 29 publications (27 studies). Evidence on patient safety was weak. Diagnostic accuracy varied between different systems but was generally low. Algorithm-based triage tended to be more risk-averse than that of health professionals. There was very limited evidence on patients' compliance with online triage advice. Study participants generally expressed high levels of satisfaction, albeit in mainly uncontrolled studies. Younger and more highly educated people were more likely to use these services.

Conclusions: The English 'digital 111' service has been implemented against a background of uncertainty around the likely impact on important outcomes. The health system may need to respond to short-term changes and/or shifts in demand. The popularity of online and digital services with younger and more educated people has implications for health equity.

Registration: PROSPERO (registration number CRD42018093564)

Strengths and limitations of this study

- This systematic review was based on a rigorous search of the literature which
 maximised efficiency by combining an initial focused search with subsequent rounds
 of follow-up searching, including searches for named symptom checker systems.
- Our narrative synthesis approach used a mixture of description and tabulation to summarise the evidence, including overall strength of the evidence base for each of the pre-specified outcomes of interest.
- Given the decision to implement a national urgent care service based on digital symptom checkers in the NHS in England, our study highlights areas of uncertainty that will need to be resolved by research and data collection.
- The review inclusion criteria were relatively broad and findings from symptom checker systems for specific conditions may not be applicable to more general systems and vice versa.
- We have also included studies of symptom checkers as part of electronic consultation systems in general practice, which again represents a slightly different setting from a general 'digital 111' service, and this should be kept in mind when interpreting the results.

Introduction

Digital and online symptom checkers and assessment services are used by patients seeking guidance about health problems, including some that may require urgent action. These services generally provide people with possible alternative diagnoses based on their reported symptoms and/or suggest a course of action (e.g. self-care, make a GP appointment or go to an emergency department (ED)).

In England, the NHS111 service provides assessment and triage by telephone for problems that are urgent but not classified as emergencies. The latest data from NHS England¹ show that in September2018 there were over 1.27 million calls to NHS111, an average of 42,400 per day. Outcomes of these calls were that 13.2% had ambulances despatched; 9.5% were recommended to attend an ED; 58.7% were recommended to attend primary care; 4.8% to attend another service; and 13.8% were not recommended to attend another service (e.g. their condition was considered suitable for self-care)

NHS England has recently introduced a digital platform to make NHS111accessible via a website or smartphone app. A beta version of the service (referred to as 'NHS111 Online') is available at https://111.nhs.uk/ (accessed 1 April 2019). The 'digital 111' service is seen as key to reducing demand for the telephone 111 service, enabling resources to be redirected to supporting 'integrated urgent and emergency care systems' as outlined in the 'NHS 5-year Forward View' and its 2017 update 'Next Steps on the NHS 5-year Forward View'^{2 3}.

There is an expectation that a digital 111 platform will help to manage demand and increase efficiency in the urgent and emergency care system, complementing the agenda of locally based Sustainability and Transformation Partnerships (STPs) which involve the health service and local government working together to integrate and co-ordinate care⁴. However, there is a risk of increasing demand, duplicating healthcare contacts (by increasing the number of potential access routes into the system) and providing advice that is not safe or clinically appropriate. For example, an evaluation of the NHS111 telephone service at four pilot sites and three control sites found that in its first year the service was not successful in reducing 999 emergency calls or in shifting patients from emergency to urgent care⁵. A recent study of 23 symptom checker algorithms providing diagnostic and triage advice that would form the

basis of a 'digital 111' platform found deficiencies in both their diagnostic and triage capabilities (based on patient vignettes)⁶.

In 2017, NHS England carried out pilot evaluations of different systems in four regions of England. The evaluations aimed to assess whether digital/online triage was acceptable to users and connected them to appropriate clinical care⁷. The full report of the evaluations was not yet published at the time of writing. The objective of this systematic review was to inform further development of the proposed digital platform by summarising and critiquing the previous research in this area, both from the UK and overseas. The overall research question was: for people seeking guidance about an urgent health problem, what is the effect of digital and online services designed to assess symptoms and signpost patients to appropriate services (compared with non-digital services or no comparator) on important clinical and health service outcomes? Outcomes include safety; clinical and cost-effectiveness; diagnostic and triage accuracy; impact on service use; patient/carer satisfaction; compliance with advice received; and outcomes related to equity and inclusion.

Methods

The review protocol was registered with PROSPERO (registration number CRD42018093564) and is available from the project website (https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/164717/).

Literature search and screening

Initial scoping searches revealed that a highly sensitive search strategy, as typically conducted for systematic reviews, retrieved a disproportionately high number of references on GP decision-making and triage as demonstrated by examination of sample search results (e.g. first 100). We therefore devised a three stage retrieval strategy as an acceptable alternative to comprehensive topic-based searching. This involved:

1. Targeted searches of precise high specificity terms in seven databases (MEDLINE, EMBASE, the Cochrane Library, CINAHL, HMIC (Health Management Information Consortium), Web of Science and ACM Digital Library). These searches were not restricted by language or date. Search strategies are presented in Appendix 1.

- 2. Phrase searching for names of known symptom checkers using a list compiled from Semigran 2015 and other sources
- 3. Citation searches and reference checking of key included studies and reviews, complemented by contact with service providers (directly and via websites).

The main literature search was completed in April 2018 and follow-up searches in May 2018. Search results were stored in a reference management system (EndNote) and imported into EPPI-Reviewer software for screening, data extraction and quality assessment. The search results were screened against the inclusion criteria by one reviewer, with a 10% random sample screened by a second reviewer. Uncertainties were resolved by discussion among the review team.

Inclusion and exclusion criteria

Population: General population seeking information online or digitally to address an urgent health problem, including adults and children and issues arising from both acute and long-term chronic illness.

Intervention: Any online or digital service designed to assess symptoms, provide health advice and direct patients to appropriate services. Services that only provide health advice were excluded, as were those that offer treatment, e.g. online CBT services.

Comparator: The 'gold standard' comparator is current practice of telephone assessment (e.g. NHS111) or face to face assessment (e.g. general practice, urgent care centre or ED). However, studies with other relevant comparators (e.g. comparative performance in tests or simulations) or with no comparator were included if they addressed the research questions.

Outcomes: The main outcomes of interest were safety (e.g. any evidence of adverse events arising from following or ignoring advice from online/digital services); clinical effectiveness; costs/cost-effectiveness; accuracy; impact on service use; compliance with advice received; patient/carer satisfaction; and equity and inclusion. 'Accuracy covered 1) ability to provide a correct diagnosis and 2) ability to distinguish between high and low acuity/urgency problems (and hence direct patients to appropriate services).

Study design: We did not restrict inclusion by study design (and included relevant audits or service evaluations in addition to formal research studies) but included studies had to evaluate (quantitatively or qualitatively) some aspect of an online/digital service

Other: Studies from any developed country healthcare system were eligible for inclusion

Excluded: Purely descriptive studies, conceptual papers, projections of possible future developments and studies conducted in low or middle income countries were excluded from the review.

Data extraction and quality/strength of evidence assessment

We extracted and tabulated key data from the included studies, including study design, population/setting, results and key limitations. Data extraction was performed by one reviewer, with a 10% random sample checked for accuracy and consistency.

To characterise the included digital and online systems as interventions, we identified studies reporting on a particular system and extracted data from all relevant studies using a modification of the TIDieR (Template for Intervention Description and Replication) checklist⁸ which we designated TIDieST (Template for Intervention Description for Systems for Triage). Further details may be found in the full report (Chambers et al., *Health Services & Delivery Research* 2019 (in press)).

Quality (risk of bias) assessment was undertaken for peer-reviewed full publications only (i.e. not grey literature publications (such as research reports, working papers, or reports produced by government departments, academics, business and industry) or conference abstracts). Randomised controlled trials were assessed using the Cochrane Collaboration risk of bias tool. For diagnostic accuracy type studies, we used the Cochrane Collaboration version of QUADAS⁹ and for other study designs we used the National Heart Lung and Blood Institute tool for observational cohort and cross-sectional studies (https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools, accessed 25th March 2019). Quality assessment was performed by one reviewer, with a random 10% sample checked for accuracy and consistency.

Assessment of the overall strength (quality and relevance) of evidence for each research question is part of the narrative synthesis. Overall strength of the evidence base for key outcomes was assessed using an adaptation of the method described by Baxter et al. ¹⁰ This involves classifying evidence as 'stronger', 'weaker', 'conflicting' or 'insufficient' based on study numbers and design. Specifically, "stronger evidence" represented generally consistent findings in multiple studies with a comparator group design or comparative diagnostic

accuracy studies; "weaker evidence" represented generally consistent findings in one study with a comparator group design and several non-comparator studies or multiple non-comparator studies; "very limited evidence" represented an outcome reported by a single study; and finally, "inconsistent evidence" represented an outcome where fewer than 75% of studies agreed on the direction of effect. All studies in the review, including those that did not meet criteria for risk of bias assessment, were included in the strength of evidence assessment.

Evidence synthesis

We performed a narrative synthesis structured around the pre-specified research questions and outcomes. We did not perform any meta-analyses because the included studies varied widely in terms of design, methodology and outcomes.

Patient and public involvement (PPI)

The review was discussed at two meetings of an existing PPI group covering the programme from which the review was commissioned (Sheffield HS&DR Evidence Synthesis Centre). At the meetings there was discussion regarding the focus of the work, including a presentation on previous research on NHS111 telephone services to provide a context for understanding the current work. The meetings also included presentation and discussion of the findings of the review, in order to explore key messages for patients which could inform dissemination of the findings. Discussion during one meeting was structured using a SWOT (strengths, weaknesses, opportunities and threats) analysis approach, which revealed a number of potential concerns amongst patients (e.g. reliability and consistency; high costs of programming and development; whether patients would follow advice given; and threats to equity) as well as potential perceived benefits (e.g. improved access to care at all hours; value to those who might feel embarrassed discussing their problem with a health professional). Involvement of the advisory group was beneficial in highlighting some issues that had also emerged from the systematic review, and enabled the reviewers to structure the review findings taking this into account. For example, the group's uncertainty about the likely impact of 'digital 111' was reflected in the review findings and recommendations for ongoing evaluation and further research. The review report also reflects the group's relatively cautious attitude (while recognising the need to update the way services are accessed) which contrasts with the strong belief in some quarters that 'digital 111' will help to ensure that patients

receive appropriate care more quickly while reducing 'inappropriate' visits to EDs and GP appointments.

Results

Results of literature search

Twenty-seven studies (29 publications) were included in the review. Figure 1 presents the flow of studies through the selection process. Inter-rater agreement on initial study selection was moderate (Kappa = 0.582). This reflects a degree of learning by the review team: our initial sift of the search results consciously favoured inclusivity and items found not to meet the inclusion criteria on detailed examination were subsequently discarded.



Figure 1: PRISMA flow diagram

Characteristics of included studies

Seventeen studies (Table 1) evaluated symptom checkers as a self-contained intervention, of which eight covered a limited range of symptoms, e.g. respiratory¹¹⁻¹³ or gastrointestinal^{14 15} symptoms which we considered to be 'urgent'. The remaining studies in this group evaluated symptom checkers covering a wider range of common urgent care symptoms. Studies either evaluated a single system¹⁶⁻¹⁹ or multiple systems^{6 20}. We found only one study of a symptom checker specifically intended for assessment of children's symptoms, a development of the SORT (Strategy for Off-Site Rapid Triage) system for influenza-like illness²¹ Two reports with some overlap of content evaluated the 'babylon check' app^{16 22}

Five studies^{7 23-26} evaluated symptom checkers as part of a broader self-assessment and consultation system (often referred to as electronic consultation or e-consultation). Study characteristics are summarised in Table 2. In this type of system, the role of symptom checkers is to help patients decide whether their symptoms require a consultation with a doctor or other health professional or can be dealt with by self-care. If a consultation is required, details of the symptoms and a request for an appointment or call-back can be submitted electronically. This type of study is important because it considers the service within the broader context of the urgent and emergency care system. A limitation is that some studies focused mainly on the 'downstream' elements of the pathway, e.g. consultation with GPs, and provided limited data on the symptom checker element of the system.

A final group of five studies examined patient and/or public attitudes to online self-diagnosis in the context of urgent care²⁷⁻³¹. See the full report for further details (Chambers et al. *Health Services & Delivery Research* 2019 (in press)).

Table 1: Studies of symptom checkers as a self-contained intervention

Reference	Study design	System type	Comparator >	Population/sample
Babylon	• Uncontrolled	• Digital	• Health professiona	General population
Health 2017 ²²	observational	Smartphone app	performance on real-world	Participants in the London
	No control group but		data	pilot evaluation of 'digital 111'
	some comparison with		• Other	services
	NHS111 telephone		NHS111 data for 12 <u>\s</u>	
	data		months from Februa	
		'	2017	
Berry 2016 ¹⁴	Simulation	• Online	• None	Specific condition(s)
	Evaluation of	17 symptom checkers	None http://bmjope	Gastrointestinal symptoms
	symptom checker		o://b	
	performance on		omjc	
	clinical vignettes		pper	
Berry 2017 ³²	Controlled	• Online	Health professional	Specific condition(s)
	observational	Three online symptom checkers (WebMD, iTriage and	performance on real-world	Patients with a cough
		FreeMD)	data	presenting to an internal
			on A	medicine clinic
Berry 2017 ¹⁵	Controlled	• Online	• Health professiona <u>É</u> .	Specific condition(s)
	observational	Three online symptom checkers (WebMD, iTriage,	performance on real-world	Abdominal pain
		FreeMD)	data	
Kellermann	Simulation	• Online	• Health professiona	Specific condition(s)
2010 11	The developed	SORT was available on 2 interactive websites	performance on real world	Influenza symptoms
	algorithm was tested		data 🤔	
	against past patient		The algorithm was tested	
	records		against clinicians' 💆	
			decision on past patient	
			records.	

Little 2016 ¹²	Experimental	• Online	• Other	Specific condition(s)
	Randomised	'Internet Doctor' website	Usual GP care with out	Respiratory infections and
	controlled trial (RCT)		access to the Internet	associated symptoms
			Doctor website ≥	
Luger et al.	Simulation	• Online	• Other	General population
2014 ³³	Described as	Google and WebMD	Comparing two integret	Older adults (50 years or
2014		Google und Wedivid		older)
	"human-computer			older)
	interaction study"		l wo	
	using think-aloud	Uh	nloa	
	protocols.		1d	
Marco-Ruiz	• Qualitative	• Online	• None $\frac{\alpha}{7}$	General population
et al. 2017 ³⁴	Qualitative element	Erdusyk	Jm -	Internet tool users
	• Other	Ch	http	
	1. Online evaluation	- / h	://br	
	by users (problem	10.	nj. Og	
	detection) 2. Think	• Online Erdusyk	Downloaded from http://bmjopen.bmj.com/ on None	
	aloud technique by		.bm	
	smaller sample of	'01.	j.co	
	participants		m/ o	
	(usability)		on t	
Middleton	Simulation	• Digital	• Health professiona	General population
2016^{16}		'babylon check' automatic triage system	performance on "O	
			test/simulation	
			Twelve 'clinicians' \$	
			(doctors) and 17 nurges	
Nagykaldi	Uncontrolled	• Online		Specific condition(s)
2010 ³⁵	observational	Customised practice website including a bilingual	Pa	Influenza
_010	ood i wionai	influenza self-triage module, a downloadable influenza	otec	
		toolkit and electronic messaging capability. A	 	
		bilingual seasonal influenza telephone hotline was	• None	
		oungua seasona injuenza telephone nottine was	<u> </u>	1

		BMJ Open	6/bmjopen-2018-C	
		available as an alternative.	018-027	
Nijland 2016 ¹⁹	• Uncontrolled observational Retrospective analysis of 15 months' data	Online Web-based triage system (http://www.dokterdokter.nl)	• None August 20	General population
Poote 2014	Uncontrolled observational	Online Prototype self-assessment triage system	• Health professional performance on real-world data GPs triage rating was compared with rating from the self-assessment system	General population Students attending a University Student Health Centre with new acute symptoms
Price 2013	Uncontrolled observational	• Online A web-based decision support tool - Strategy for Offsite Rapid Triage (SORT) for Kids designed to help parents and adult caregivers decide whether a child with possible influenza symptoms needs to visit the emergency department for immediate care.	• Health professional performance on real-world data The sensitivity of the algorithm was compaged with a gold standard evidence form child's medical records that they received 1 or more of ED-specific interventions.	Specific condition(s) Influenza in children
Semigran 2015 ⁶	• Experimental Described as an audit study	• Multiple 23 symptom checkers were evaluated. Symptom checkers available as apps (via the App Store and Google Play) were identified through searching for "symptom checker" and "medical diagnosis" and screened the first 240 results. Symptom checkers available online were identified through searching Google and Google Scholar for "symptom checker"	• Other Vignettes had a diagnosis and triage attached to them and these were compared against the symptom checker advice.	General population Where a single class of illness was examined by the symptom checker, the symptom checker was excluded from the study.

6/bmjopen-2018-

			<u>, </u>	
		and "medical diagnosis" and screened the first 300	277	
		results.	743 (
Semigran	Experimental	Multiple	• Health professiona	General population
$2016^{\ 20}$	Comparison of	"Human Dx is a web-and app based platform"	performance on test/simulation Clincial vignettes - as	Of the 45 condition vignettes
	physician and		test/simulation	- there were 15 low, 15
	symptom checker		Clincial vignettes - og	medium and 15 huigh acuity
	diagnoses based on		comparison of 23 symptom	vignettes - there were 26
	clinical vignettes		checkers with physican	common and 19 uncommon
		Oh	diagnosis for 45 vigrettes	condition vignettes
Sole 2006 18	• Uncontrolled	• Online	• Health professiona	General population
501C 2000	observational		1 2	General population
	Descriptive	A web-based triage system (24/7 Weblieu)	data	
	comparative study		Data was evaluated #rom	
	comparative study	' /	students who had used the	
		· 01	web based triage and then	
		A web-based triage system (24/7 WebMed)	requested an appointment	
		(0)	via email (so triage data	
			was available for	
			comparison).	
Yardley	Experimental	• Online	• Other	Specific condition(s)
2010^{13}	Exploratory	'Internet Doctor' website	Self-care information	Minor respiratory symptoms,
	randomised trial		provided as a static Reb	e.g. cough, sore throat, fever,
			page with no symptom	runny nose
			checker or triage ad ce	
			st.	
			Prot	
			ecte	
			ф ф	
			y c	
			Protected by copyright	
			ight	15
			• *	1.

Table 2: Studies of symptom checkers as part of an electronic consultation system

Reference	Study design	System type	Comparator 5	Population/sample
Carter	Uncontrolled	• Online	• Other	General population
2018 23	observational	webGP(subsequently known as eConsult)	Investigate patient experience by	General practices in NHS
	Mixed-methods		surveying patients who had use	Northern, Eastern and
	evaluation	Oh	webGP and comparing their $\frac{3}{6}$	Western Devon Clinical
		-	experience with controls (patients who	Commissioning Group's
		\mathcal{O}_{\triangle}	had received a face-to-face	area
		700	consultation during the same time	
		· C/	period) matched for age and gander	
Cowie	Uncontrolled	• Online	• None	General population
2018 ²⁴	observational	eConsult, accessed via GP surgery websites.	nji Op	Patients registered with
	6-month evaluation	Service provides self-care assessment and	en.r	participating GP practices
	at 11 GP practices	advice, including symptom checkers; triage and	om.	
	in Scotland	signposting to alternative services; access to	com	
		NHS24 (phone service); and e-consults allowing	mjopen.bmj.com/ on	
		submission of details by e-mail)	→	
Madan	Uncontrolled	• Online	• None $\frac{\overline{z}}{\omega}$	General population
2014 ²⁵	observational	webGP (subsequently known as eConsult)	, 2024	
	Report of 6-month		24 k	
	pilot study) 9	
NHS	• Uncontrolled	• Multiple	• None 6	General population
England ⁷	observational	Pilots featured NHS Pathways (Web-based;	Authors stated it was not appropriate	
	Analysis of data	West Yorkshire); Sense.ly ('voice-activated	to compare pilot sites because 🏽	
	from four pilot	avatar'; West Midlands); Espert 24 (Web-based;	differences in starting date, $\frac{Q}{Q}$	
	studies together with	Suffolk) and babylon (app; North Central	'footprints' covered, method of brake	
	data from other	London)	and underlying population	

BMJ Open

				3-0	
	sources			277	
Nijland	• Other	• Online	• None	43	General population
2009^{26}	Online survey	Responses of interest relate to 'indirect e-		on 1	Patients with Internet access
		consultation' (consulting a GP via secure e-mail		Au	but no experience of e-
		with intervention of a Web-based triage system)		gus	consultation

Results by outcome

Safety

None of the six included studies that reported on safety outcomes identified any problems or differences in outcomes between symptom checkers and health professionals. Most of the studies compared system performance with that of health professionals using real or simulated data. The only study with no comparison group was the 6-month pilot study of webGP²⁵, which reported 'no major incidents'.

Limitations of the studies included not being based on real patient data¹⁶; covering only a limited range of conditions^{11 21}; and sampling a young healthy population (students) not representative of the general population of users of the urgent care system¹⁷. Studies of e-consultation systems did not generally collect data on those respondents who decided not to seek an appointment, limiting their ability to assess any impact on safety for this group. Overall, the evidence should be interpreted cautiously as indicating no evidence of a detrimental impact on safety rather than evidence of no detrimental effect.

Clinical effectiveness

Only two studies reported on clinical effectiveness outcomes, making it difficult to draw any firm conclusions. In the study by Little et al., those who used the Internet Doctor website experienced longer illness duration and more days of illness rated moderately bad or worse than the usual care group¹² The pilot study of the webGP system²⁵ reported that several patients received advice to seek treatment for serious symptoms that might otherwise have been ignored. However, no details or quantitative data were provided.

Costs/cost-effectiveness

Two included studies provided limited data on possible cost savings. Based on 6 months of pilot data, Madan²⁵estimated savings of £11,000 annually for an average general practice (6,500 patients) compared with current practice. The report also suggested a saving to commissioners equivalent to £414,000 annually for a CCG (Clinical Commissioning Group, responsible for specifying and purchasing most health services in the NHS in England) covering 250,000 patients. These savings were specifically related to self-reported diversion of patients from GP appointments to self-care and from urgent care to e-consultation. Using

similar methodology, the manufacturers of the 'babylon check' app claimed average savings of over £10/triage compared with NHS111 by telephone, based on a higher proportion of patients being recommended to self-care²².

Diagnostic accuracy

Eight studies reported at least some data on the diagnostic accuracy of symptom checkers. In spite of the diverse methods and comparisons in the included studies, almost all agreed that the diagnostic accuracy of symptom checkers was poor in absolute terms (e.g. in evaluating 'vignettes' designed to test knowledge of specific conditions, where the correct diagnosis was already known by definition) or relative to that of health professionals. In the most comprehensive evaluation, Semigran et al. evaluated 23 symptom checkers across 770 standardised patient evaluations⁶. Overall the correct diagnosis was made in 34% of cases (95% CI 31%-37%), although performance varied widely between symptom checkers, high and low acuity conditions and common and rare conditions. When the same authors compared the 23 symptom checkers with physicians using 43 vignettes, physicians were more likely to list the correct diagnosis first (out of three differential diagnoses) (72.1% vs. 34% p<0.001) as well as among the top three diagnoses (84.3% vs. 51.2% p<0.001)²⁰.

The only exception to the rule was an evaluation carried out at a student health centre¹⁸. Using data from 59 participants who used the 24/7 WebMed system and who were subsequently treated at the health centre, the study found good agreement between chief complaint, 24/7 WebMed classification and provider diagnosis (kappa values 0f 0.89 to 0.94). This study differed from the others in using data from students rather than a general population sample. In addition, the students' complaints were generally common and uncomplicated, a scenario in which symptom checkers performed relatively well in the study by Semigran et al.²⁰.

Accuracy of disposition (triage and signposting to appropriate services)

Six included studies reported on this outcome, all except one of which¹⁵ evaluated a 'general purpose' symptom checker. As with diagnostic accuracy, diverse methodologies and outcome measures were used.

The results overall presented a mixed picture but most studies indicated that symptom checkers were inferior and/or more cautious in their triage advice compared with doctors or other health professionals. In their review of 23 symptom checkers, Semigran et al. found that the systems provided appropriate triage advice in 57% (95% CI 52% to 61%) of cases⁶. Performance varied across the systems evaluated, correct triage ranging from 33% to 78%. The NHS England pilot evaluation of four systems⁷ found that agreement with clinical experts varied from 30% to 95%, although the number of responses also varied, reducing the comparability of the results.

For abdominal pain, Berry et al. evaluated three symptom checkers and found that 33% of diagnoses were at the same level of urgency as physician diagnoses (emergency, non-emergency or self-care); 39% were diagnosed as more serious and 30% less serious than the physician's judgement¹⁵. A similar level of agreement between algorithm and clinician (39%) was reported by Poote et al.¹⁷, while the system evaluated by Nijland et al. advised patients to visit a doctor in 85% of cases, even when the symptoms were appropriate for self-care¹⁹.

The only studies to report clearly equal or superior accuracy of disposition using an automated system were the evaluations of Babylon check by the company that developed the system. Middleton et al. 16 reported that using patient vignettes, the app gave an accurate triage outcome in 88.2% of cases, compared with 75.5% for doctors and 73.5% for nurses (unaware of the 'correct' diagnosis for the vignettes). When vignettes were delivered by a medical professional rather than actors, the accuracy of Babylon check increased to over 90%. A later report looked at triage results obtained as part of the NHS England pilot evaluation, concluding that all of 74 referrals to urgent or emergency care were appropriate²².

Impact on service use/diversion

Eight studies reported on this outcome, although one of them¹¹ merely stated that it was not possible to assess the effect of the intervention (a web-based influenza triage system) on patients' use of health services.

The pilot evaluation of the webGP system reported that 18% of users planned to book an appointment but chose not to do so²⁵ In addition, 14% of users reported that they would have

attended a walk-in centre or other urgent care service if they had not had access to the webGP system.

The NHS England pilot evaluation of four online/digital systems in different regions of England⁷ compared the recommendations of the digital systems with those of the NHS111 telephone service over a similar time period (the first months of 2017). Compared with the telephone service, the online and digital services directed a slightly higher proportion of patients to self-care (18% vs. 14%) and a lower proportion to other primary care services such as GPs, dental and pharmacy (40 vs. 60%). The manufacturer's data on the 'babylon check' app collected as part of the NHS England evaluation indicated that patients were more likely to be triaged to self-care by the app compared with NHS111 by telephone (40 vs. 14%)²². This figure includes people who received information leaflets on self-care as well as those who were actively triaged. If the former group is excluded, the figures for the two services are similar (14% for NHS111 and 15.6% for 'babylon check'²².

In their study of self-assessment for students attending a university health centre, Poote et al. found that the prototype system they studied was able to identify a proportion of cases that doctors considered appropriate for self-care, suggesting a potential to reduce service use¹⁷. Similarly, Little et al's RCT of a web-based symptom checker designed to support self-care for respiratory symptoms¹² reported that patients in the intervention group had fewer contacts with doctors than the usual care control group despite having a longer duration of illness and more days with relatively severe symptoms. This was balanced by an increase in contacts with the NHS Direct telephone service (which preceded NHS 111) and it should be noted that the system under evaluation recommended people needing treatment to contact NHS Direct rather than go directly to a doctor. Finally, a study of young adults (students) found that intention to seek treatment for a hypothetical illness was stronger when the diagnosis was made with the aid of WebMD or Google than with no electronic aid³⁰.

Patient compliance with triage advice

Only two of the included studies reported specifically on patients' compliance (or intention to comply) with advice received. The NHS England pilot evaluation in four regions asked participants in two of those regions (Suffolk and London) what they intended to do based on the advice received⁷. No quantitative data were provided but the report stated that in the

Suffolk pilot, 'overall users would have followed the advice given'. However, those who were recommended to call 999 or attend an ED were more likely to seek advice from primary care or self-management. Similarly, in the London region there was generally good agreement between advice and intended action but patients recommended to call 999 or go to an ED indicated that they would seek advice from a GP. In a study of a web-based triage system in the Netherlands, 192 patients were asked about their intention to comply immediately after receiving advice from the system¹⁹. Thirty-five patients responded to a follow-up survey on actual compliance, of whom 20 (57%) reported that they had followed the advice. Compliance was correlated with intention to comply, which in turn was correlated with the patient's attitude towards the advice received.

Equity and inclusion

Fourteen studies investigated the outcome of equity and inclusion or compared users and non-users. One study ¹²reported that patients who were classed as less deprived were more likely to agree to use "Internet Doctor" than decline participation, although no relationship was found between deprivation and results in this study or between e-Consult use and deprivation in another study ²⁴. Association between e-consultation use and education levels was explored in a third study. Patients with low to medium levels of education tended to be motivated toward indirect e-consultation (which involves contact with a health professional via e-mail), mainly to reduce uncertainty ²⁶

Evidence from included studies suggests that users of e-consultation were more likely to be young⁷ ²³⁻²⁵, employed¹⁹ ²³ ²⁵ and female⁷ ¹⁹ ²⁴ ²⁵than non-users. One study also found a significantly larger use by white patients (78%) than other ethnicities²⁴.

Risk of bias assessment

We assessed risk of bias in the two included RCTs¹² 13 using the Cochrane risk of bias tool. Thirteen studies¹¹ 18 19 23 24 26-29 31 33-35 were assessed with the tool for cross-sectional and cohort studies and four (six publications⁶ 17 20 21 36 37) with the modified QUADAS tool. Seven grey literature reports and conference abstracts were not formally assessed for risk of bias⁷ 14-16 25 30 32. Identified limitations were extracted for all included studies.

Risk of bias results are presented in Appendix 2. With the possible exception of the two randomised trials, the included studies generally had at least a moderate risk of bias. However, the diverse designs and objectives of the studies made risk of bias difficult to assess in some cases with the available tools. Grey literature reports containing relevant data were included in the review but not formally assessed for risk of bias. Reports prepared by individuals with a commercial interest in a specific system and published without independent peer review^{16 25} should be treated with particular caution because of possible conflicts of interest.

Overall strength of evidence assessment/evidence map

The overall strength of evidence for key outcomes is summarised in Table 3. We found relatively strong evidence that the diagnostic accuracy of digital and online symptom checkers tends to be lower than that of health professionals; and that patients who have used these systems generally show high levels of satisfaction (mainly in non-comparative studies). Areas where evidence is lacking or inconsistent include clinical and cost-effectiveness, accuracy of disposition to appropriate services and patient compliance with advice received. For safety, we found no evidence of an increased risk with digital/online systems but the available evidence was weak.

Table 3: Overall strength of evidence by outcome

Outcome	Relevant studies	Evidence statement	Strength of evidence	Comments &
Safety	=Kellermann 2010 ¹¹ =Little 2016 ¹² =Middleton	No evidence of a difference in risk between health professionals and symptom checkers	Weaker	Rating changed from stronger based on study numbers and design to weaker because of low numbers of adverse events reported
	2016 ¹⁶ =Poote 2014 ¹⁷ =Price 2013 ²¹ Madan 2014 ²⁵	FO _F		3. Downloade
Clinical effectiveness	-Little 2016 ¹² ?Madan 2014 ²⁵	Insufficient evidence to draw any firm conclusions	Very limited	ad from
Costs/cost- effectiveness	+Babylon Health 2017 ²² +/-Cowie 2018 ²⁴ +Madan 2014 ²⁵	Insufficient evidence to draw any firm conclusions	Inconsistent	http://bmjop
Diagnostic accuracy	?Berry 2016 ¹⁴ -Berry 2017 ³² -Berry 2017 ¹⁵ -Price 2013 ²¹ ?Semigran 2015 ⁶ -Semigran 2016 ²⁰ =Sole 2006 ¹⁸	Symptom checkers appear inferior to health professionals in terms of diagnostic accuracy	Stronger	Mainly for specific conditions or pre-prepared vignettes 3.
Disposition accuracy	=Babylon Health 2017 ²² -Berry 2017 ¹⁵ =Middleton 2016 ¹⁶ ?Nijland 2010 ¹⁹ -Poote 2014 ¹⁷ +/-Semigran 2015 ⁶	Inconsistent findings on accuracy of disposition	Inconsistent	Performanc@variable between different systems by guest. Protected by coppy

Outcome	Relevant studies	Evidence statement	Strength of evidence	Comments 27743
	+/-NHS England 2017 ⁷		0.130130	3 on 1 /
Service use/diversion	?Kellermann 2010 ¹¹ +/-Little 2016 ¹² +/-Poote 2014 ¹⁷ ?Carter 2018 ²³ ?Cowie 2018 ²⁴ +Madan 2014 ²⁵ +/- NHS England 2017 ⁷ +Babylon Health 2017 ²² -Luger 2011 ³³	Inconsistent findings on effects on service use	Inconsistent	August 2019. Downloaded from http://b
Compliance	?Nijland ¹⁹ ?NHS England 2017 ⁷	No comparative data on compliance	Very limited	dmjopen.b
Patient/carer satisfaction	?Nagykaldi 2010 ³⁵ ?Nijland ¹⁹ ?Price 2013 ²¹ +Yardley ¹³ ?Carter 2018 ²³ ?Cowie 2018 ²⁴ ?Madan 2014 ²⁵ ?NHS England 2017 ⁷ ?Lanseng 2007 ²⁹	Most studies report high rates of patient satisfaction with symptom checkers and econsultation systems generally	Weaker	Few studies with comparator data On April 9, 2024 by guest.

Controlled studies in bold; = means no significant difference in outcomes; + means better outcome with symptom checker; +/- varying results within study; ? results difficult to interpret in comparative terms

Discussion

Main findings

The literature search identified 29 publications describing 27 studies that met the inclusion criteria. The overall strength of the evidence base varied between outcomes (Table 3), but in absolute terms the evidence is weak, being based largely on observational studies. A substantial component of grey literature of uncertain quality complicates the interpretation of the evidence. Interpretation of the evidence should also take into account risks of bias in individual studies.

We found little evidence to indicate whether or not digital and online symptom checkers are detrimental to patient safety. The studies that reported on the outcome were mostly short-term and involved relatively small samples and hence reported few or no adverse events. Some were limited to people with specific types of symptoms and others recruited from specific population groups not representative of typical users of urgent care services. This body of evidence should therefore be interpreted cautiously and not extrapolated to the possible impact of a nationally available digital urgent care service being used by millions of people annually.

The evidence on patient satisfaction with digital and online systems also had some limitations but these findings appear more likely to be generalisable. Study participants generally expressed high levels of satisfaction, albeit in uncontrolled studies. For example, in the NHS England pilot evaluation, 70–80% of users were satisfied with their experience at each of the pilot sites⁷. This evidence, together with the increasing reliance on digital technology in all areas of life, suggests that any national digital urgent care service may be popular and well-used, although different sections of the population may differ in their degree of engagement (see the discussion of equity and inclusion below)...

Digital and online systems have yet to achieve a high level of accuracy in the diagnosis of specific conditions. This finding applies both to 'general purpose' symptom checkers and to those limited to particular conditions. Although the evidence was classified as relatively strong, several caveats should be applied. Some of the included studies did not recruit

representative populations and others were based on standardised vignettes rather than real-world data. In addition, studies that compared symptom checkers with health professionals tended to use the doctors' clinical diagnosis as the reference standard, which would bias the comparison in favour of the health professionals.

Accuracy of signposting of patients to the most appropriate level of service is closely related to diagnostic accuracy, but results for this outcome were inconsistent between studies. In general, algorithm-based triage tended to be more risk-averse than that of health professionals, with 85% of respondents being advised to visit their doctor in one study¹⁹. While there is considerable uncertainty about the magnitude of the effect, a national digital urgent care service could result in considerable numbers of patients receiving inappropriate advice to visit the ED or request an urgent GP appointment. Middleton and colleagues¹⁶ claimed that the 'babylon check' app had a high degree of triage accuracy for vignettes compared with health professionals, but this non-peer-reviewed report requires further validation.

We also found inconsistent evidence on effects on service use. There was some indication that symptom checkers can influence the pattern of service use but the magnitude and direction of the effect varied between studies. Patients' reactions to online triage advice and whether they follow the advice or seek further help or information would have implications for service use but we found limited evidence for this outcome. Preliminary findings from the NHS England evaluation suggest that patients may be more likely to seek further advice for more urgent conditions⁷ but further confirmation is required.

Over half of the included studies considered equity and inclusion issues either directly or by comparing users and non-users of digital triage systems. Not surprisingly, studies revealed a clear consensus that younger and more highly educated people are more likely to use these services while older and less educated patients are more likely to prefer telephone or face-to-face contact. This could have implications for health equity if urgent care pathways prioritise (or appear to prioritise) requests originating from digital sources. Problems have arisen in primary care because patients using e-consultation systems to request an appointment following online triage may be seen more quickly than those contacting the practice by telephone.

Strengths and limitations

This systematic review was undertaken on a short timescale using a relatively large team of experienced researchers, including both methodological and topic experts. We performed a rigorous search of the literature including reference checking and citation searching. Rather than a conventional highly sensitive search (which would have resulted in inefficiencies in the screening process), we combined an initial focused search with subsequent rounds of follow-up searching, including searches for named symptom checker systems. We assessed risk of bias in individual studies using a variety of appropriate checklists as well as summarising the overall strength of evidence for key outcomes (Table 3).

The heterogeneous and descriptive nature of the included studies meant that meta-analysis was not feasible for any of the outcomes of interest. Our narrative synthesis approach used a mixture of description and tabulation to summarise the evidence for each of the pre-specified outcomes of interest. This was a review of published (including non-peer-reviewed) literature and the coverage of systems is not exhaustive; for example, we did not extract data from websites. We also did not carry out any original analyses of raw data even where such data were available. The timing of the review meant that final results of NHS England's pilot evaluation were not available to us. We were able to make use of a draft report that was published online⁷ but we acknowledge that the findings of the final evaluation report, when available, will supersede those of the 2017 draft.

The review inclusion criteria were relatively broad and findings from symptom checker systems for specific conditions may not be applicable to more general systems and vice versa. We have also included studies of symptom checkers as part of electronic consultation systems in general practice, which again represents a slightly different setting from a general 'digital 111' service, and this should be kept in mind when interpreting the results.

A systematic review in such a topical area of research will require regular updating to keep track of new studies. For example, Verzantvoort et al.³⁸ published a study of self-triage using a smartphone app for out-of-hours primary care in the Netherlands shortly after our literature searches were completed. The app was rated highly for clarity and patient satisfaction. Sensitivity and specificity (using nurse telephone triage as reference standard) were 84% and 74% respectively, although diagnostic accuracy was only evaluated in a sample of

participants (126/4456). Inclusion of this study would not have affected the main conclusions of our review.

Implications for service delivery and research

The implications of this systematic review for service delivery should be considered in the context that a decision has already been taken to introduce a 'digital 111' service and the service became available across England by December 2018. Achieving a high level of diagnostic accuracy will be key to the success of a 'digital 111' service. Failure to provide an accurate diagnosis may result in outcomes including patient dissatisfaction and unwillingness to use the service again; increased use of other urgent and emergency care services; and possible risks to patient safety (although the cautious approach characteristic of most existing systems may help to mitigate this).

The studies included in the review suggest a high level of uncertainty about the impact of 'digital 111' on the urgent care system and the wider healthcare system. Some of these uncertainties can be addressed by research and data collection but the health service may need to respond to short-term increases (or decreases) in demand and/or shifts from one part of the system to another. This may increase pressure on the system, at least in the short-term. In the longer-term, if usage of the 111 telephone service decreases as planned, there may be opportunities to reconfigure the workforce to support the integrated urgent care agenda.

Based on the areas of limited evidence identified by the review, priorities for research (in addition to ongoing collection of data to monitor usage and safety of the 'digital 111' service) include studies to compare the performance of different systems directly; rigorous economic evaluations based on real-world data; research to investigate the pathways followed by patients using the service; evaluation of systems designed for childhood illnesses; and investigation of the possible role of behaviour change theory in the development and implementation of symptom checkers. Qualitative research to investigate perceptions of symptom checkers and barriers to their use by people who are less familiar with digital technology would also be of value.

Ethical approval

Ethical approval was not required for this work

Funding

This report presents independent research funded by the National Institute for Health Research (NIHR) Health Services & Delivery Research Programme (project number HSDR16/47/17). The funding programme approved the review protocol but had no role in the collection, analysis and interpretation of the data, the writing of this paper or the decision to submit the paper for publication. The views and opinions expressed are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

Data sharing

No new data have been created in the preparation of this report and therefore there is nothing available for access and further sharing. All queries should be submitted to the corresponding author.

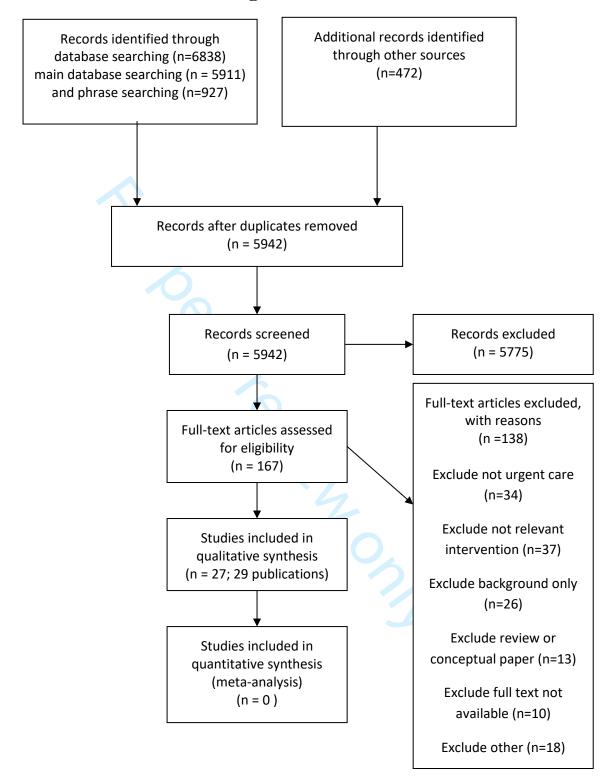
References

- 1. NHS England. NHS 111 minimum data set 2018-19 2018 [Available from: <a href="https://www.england.nhs.uk/statistics/statistical-work-areas/nhs-111-minimum-data-set/statistical-work-areas-nhs-111-minimum-data-set-nhs-111-minimum-data-set-2018-19/accessed 29 October 2018.
- 2. NHS England. Five year forward view. Leeds: NHS England, 2014.
- 3. NHS England. Next steps on the NHS Five Year Forward View. Leeds: NHS England, 2017.
- 4. NHS England. Sustainability and transformation partnerships [Available from: https://www.england.nhs.uk/integratedcare/stps/ accessed March 25 2019.
- 5. Turner J, O'Cathain A, Knowles E, et al. Impact of the urgent care telephone service NHS 111 pilot sites: a controlled before and after study. *BMJ Open* 2013;3(11):e003451. doi: 10.1136/bmjopen-2013-003451
- 6. Semigran HL, Linder JA, Gidengil C, et al. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015;351:h3480.
- 7. NHS England. NHS111 online evaluation. Leeds: NHS England, 2017.
- 8. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687. doi: 10.1136/bmj.g1687

- 9. Reitsma JB, Rutjes AWS, Whiting P, et al. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 100: The Cochrane Collaboration 2009.
- 10. Baxter S, Johnson M, Chambers D, et al. The effects of integrated care: a systematic review of UK and international evidence. *BMC Health Serv Res* 2018;18(1):350. doi: 10.1186/s12913-018-3161-3
- 11. Kellermann AL, Isakov AP, Parker R, et al. Web-based self-triage of influenza-like illness during the 2009 H1N1 influenza pandemic. *Annals of Emergency Medicine* 2010;56(3):288-94.e6.
- 12. Little P, Stuart B, Andreou P, et al. Primary care randomised controlled trial of a tailored interactive website for the self-management of respiratory infections (Internet Doctor).[Erratum appears in BMJ Open. 2017 Mar 21;7(3):e009769corr1; PMID: 28325861]. *BMJ Open* 2016;6(4):e009769.
- 13. Yardley L, Joseph J, Michie S, et al. Evaluation of a Web-based intervention providing tailored advice for self-management of minor respiratory symptoms: exploratory randomized controlled trial. *Journal of Medical Internet Research* 2010;12(4):e66.
- 14. Berry AC, Berry BB, Nakshabendi R, et al. Evaluation of Accuracy Between Online Symptom Checkers for Diagnosis of Gastrointestinal Symptoms from MKSAP Clinical Vignette Board Review Questions. *Gastroenterology* 2016;150(4):S849-S50. doi: 10.1016/s0016-5085(16)32869-4
- 15. Berry AC, Cash BD, Mulekar MS, et al. SYMPTOM CHECKERS VS. DOCTORS, THE ULTIMATE TEST: A PROSPECTIVE STUDY OF PATIENTS PRESENTING WITH ABDOMINAL PAIN. *Gastroenterology* 2017;152(5):S852-S53. doi: 10.1016/s0016-5085(17)32937-2
- 16. Middleton K, Butt M, Hammerla N, et al. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. London: Babylon Health, 2016.
- 17. Poote AE, French DP, Dale J, et al. A study of automated self-assessment in a primary care student health centre setting. *Journal of Telemedicine & Telecare* 2014;20(3):123-7.
- 18. Sole ML, Stuart PL, Deichen M. Web-based triage in a college health setting. *Journal of American College Health* 2006;54(5):289-94.
- 19. Nijland N, Cranen K, Boer H, et al. Patient use and compliance with medical advice delivered by a web-based triage system in primary care. *Journal of Telemedicine and Telecare* 2010;16(1):8-11. doi: 10.1258/jtt.2009.001004
- 20. Semigran HL, Levine DM, Nundy S, et al. Comparison of Physician and Computer Diagnostic Accuracy. *JAMA Intern Med* 2016;176(12):1860-61. doi: 10.1001/jamainternmed.2016.6001
- 21. Price RA, Fagbuyi D, Harris R, et al. Feasibility of web-based self-triage by parents of children with influenza-like illness: A cautionary tale. *JAMA Pediatrics* 2013;167(2):112-18.
- 22. Babylon Health. NHS111 powered by babylon: outcomes evaluation. London: Babylon Health, 2017.
- 23. Carter M, Fletcher E, Sansom A, et al. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. *BMJ Open* 2018;8(2):e018688.
- 24. Cowie J, Calveley E, Bowers G, et al. Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. *International Journal of Environmental Research and Public Health* 2018;15(896)
- 25. Madan A. WebGP: the Virtual general practice. London: Hurley Group, 2014.
- 26. Nijland N, van Gemert-Pijnen J, Boer H, et al. Increasing the use of e-consultation in primary care: Results of an online survey among non-users of e-consultation. *International Journal of Medical Informatics* 2009;78(10):688-703. doi: 10.1016/j.ijmedinf.2009.06.002
- 27. Backman AS, Lagerlund M, Svensson T, et al. Use of healthcare information and advice among non-urgent patients visiting emergency department or primary care. *Emergency Medicine Journal* 2012;29(12):1004-06.
- 28. Joury AU, Alshathri M, Alkhunaizi M, et al. Internet Websites for Chest Pain Symptoms Demonstrate Highly Variable Content and Quality. *Academic Emergency Medicine* 2016;23(10):1146-52.

- 29. Lanseng EJ, Andreassen TW. Electronic healthcare: a study of people's readiness and attitude toward performing self-diagnosis. *International Journal of Service Industry Management* 2007;18(3-4):394-417. doi: 10.1108/09564230710778155
- 30. Luger TM, Suls J. Online health information and intentions to seek healthcare. *Psychosomatic Medicine* 2011;73 (3):A59.
- 31. North F, Varkey P, Laing B, et al. Are e-health web users looking for different symptom information than callers to triage centers? *Telemedicine Journal & E-Health* 2011;17(1):19-24.
- 32. Berry AC, Berry NA, Wang B, et al. Symptom checkers versus doctors: A prospective, head-to-head comparison for GERD vs. Non-GERD Cough. *American Journal of Gastroenterology* 2017;112 (Supplement 1):S190. doi: http://dx.doi.org/10.1038/ajg.2017.299
- 33. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. *Journal of Medical Internet Research* 2014;16(1):e16.
- 34. Marco-Ruiz L, Bones E, de la Asuncion E, et al. Combining multivariate statistics and the thinkaloud protocol to assess Human-Computer Interaction barriers in symptom checkers. *Journal of Biomedical Informatics* 2017;74:104-22.
- 35. Nagykaldi Z, Calmbach W, Dealleaume L, et al. Facilitating patient self-management through telephony and web technologies in seasonal influenza. *Informatics in Primary Care* 2010;18(1):9-16.
- 36. Fraser HSF, Clamp S, Wilson CJ. Limitations of Study on Symptom Checkers. *JAMA Internal Medicine* 2017;177(5):740-41.
- 37. Mehrotra A, Semigran HL, Levine DM, et al. Limitations of Study on Symptom Checkers. *JAMA Internal Medicine* 2017;177(5):741-41.
- 38. Verzantvoort NCM, Teunis T, Verheij TJM, et al. Self-triage for acute primary care via a smartphone application: Practical, safe and efficient? *PLoS One* 2018;13(6):e0199284. doi: 10.1371/journal.pone.0199284 [published Online First: 2018/06/27]

PRISMA 2009 Flow Diagram





Appendix 1: Database search strategies

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present> Search Strategy:

- 1 (symptom checker or symptoms checker or symptom checkers or symptoms checkers).tw. (21)
- 2 ("self diagnosis" or "self referral" or "self triage" or "self assessment").tw. (10438)
- 3 TRIAGE/ (10017)
- 4 2 or 3 (20415)
- 5 (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone).tw. (658190)
- 6 4 and 5 (1568)
- 7 ("online diagnosis" or "web based triage" or "electronic triage" or etriage).tw. (42)
- 8 1 or 6 or 7 (1608)

Embase

- 1 (symptom checker or symptoms checker or symptom checkers or symptoms checkers).tw.
- 2 ("self diagnosis" or "self referral" or "self triage" or "self assessment").tw.
- 3 emergency health service/
- 4 2 or 3
- 5 (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone).tw.
- 6 4 and 5
- 7 ("online diagnosis" or "web based triage" or "electronic triage" or etriage).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
- 8 1 or 6 or 7

Cochrane Library

#1 symptom checker or symptoms checker or symptom checkers or symptoms checkers:ti,ab,kw (Word variations have been searched)

#2 "self diagnosis" or "self referral" or "self triage" or "self assessment":ti,ab,kw (Word variations have been searched)

#3 MeSH descriptor: [Triage] explode all trees

#4 #2 or #3

#5 online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone:ti,ab,kw (Word variations have been searched)

#6 #4 and #5

#7 "online diagnosis" or "web based triage" or "electronic triage" or etriage:ti,ab,kw (Word variations have been searched)

#8 #1 or #6 or #7

CINAHL

S8 (S1 OR S6 OR S7)

S7 TI ("online diagnosis" or "web based triage" or "electronic triage" or etriage) OR AB ("online diagnosis" or "web based triage" or "electronic triage" or etriage) S6 S4 AND S5

S5 TI (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone) OR AB (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone)

S4 (S2 OR S3)

S3 (MH "Triage")

S2 TI ("self diagnosis" or "self referral" or "self triage" or "self assessment") OR AB ("self diagnosis" or "self referral" or "self triage" or "self assessment")

S1 TI (symptom checker or symptoms checker or symptom checkers or symptoms checkers) OR AB (symptom checker or symptoms checker or symptoms checkers)

ACM digital library

WOS

#8 #7 OR #6 OR #1

#7 TS=("online diagnosis" OR "web based triage" OR "electronic triage" OR etriage)

#6 #5 AND #4

#5 TS=(online OR on-line OR web OR electronic OR automated OR internet OR digital OR app OR mobile OR smartphone)

#4 #3 OR #2

#3 TS=triage

#2 TS=("self diagnosis" or "self referral" or "self triage" or "self assessment")

#1 (symptom checker or symptoms checker or symptom checkers or symptoms checkers)

HMIC

- 1 (symptom checker OR symptoms checker OR symptom checkers OR symptoms checkers).ti,ab
- 2 ("self diagnosis" OR "self referral" OR "self triage" OR "self assessment").ti,ab
- 3 TRIAGE/
- 4 (2 OR 3)
- 5 (online OR on-line OR web OR electronic OR automated OR internet OR digital OR app OR mobile OR smartphone).ti,ab
- 6 (4 AND 5)
- 7 ("online diagnosis" OR "web based triage" OR "electronic triage" OR etriage).ti,ab 8 (1 OR 6 OR 7)

Risk of bias results for randomised trials

Appendix 2: Risk of bias tables

Study ID Random sequence generation Low risk Selection and attrition bills Reporting and other his	Cl 4 TP41 -	D.f	C-14:		D 1 -4h h
*Reference Little 2016¹² *Allocation concealment •Low risk *Blinded assessment of primetry care records *Incomplete outcome data* *Low risk *Incomplete outcome data* *Low risk *Anything else, ideally prespecified •Low risk *Yardley (2010) *Reference Yardley 2010¹³ *Allocation concealment •Low risk *Incomplete outcome data* *Unclear *Allocation concealment •Low risk *Incomplete outcome data* *Incomplete outcome data* *Anything else, ideally prespecified •Unclear	Short Title	Reference	Selection and performance bias	Detection and attrition bias	Reporting and other bias
Allocation concealment • Low risk Low risk Low risk Low risk	Little (2016)	Study ID		Blinding of outcome $\vec{\varphi}$	
Allocation concealment • Low risk Blinding of participants and personnel* • Unclear Pardley (2010) • Reference Yardley 2010 ¹³ Allocation concealment • Low risk Anything else, ideally prespecified • Low risk Blinding of participants and personnel* • Unclear • Unclear Blinding of outcome assessment* • Unclear Allocation concealment • Low risk Incomplete outcome data • Unclear Anything else, ideally prespecified • Anything else, ideally prespecified			• Low risk	assessment*	• Unclear
Allocation concealment • Low risk Low risk Low risk Low risk		Little 2016 ¹²		• Low risk	
Allocation concealment • Low risk Low risk Low risk Low risk				Blinded assessment of primary	
		•	Allocation concealment	care records $\frac{\alpha}{2}$	Anything else, ideally
		4	• Low risk	rom	prespecified
				- htt	• Low risk
				Incomplete outcome data	
			Blinding of participants and	• Low risk	
				Эре	
			_	1.br	
			/6	nj.o	
	Yardley (2010)	Study ID	Random sequence generation	Blinding of outcome	Selective reporting
	101010)	_		assessment*	
			Low risk	• Unclear	Chercui
		Turatey 2010		Sileicai <u>si</u>	
			Allogation concealment	200	Anything else ideally
				Incomplete outcome detakt	
Blinding of participants and personnel* • Low risk			• LOW FISK	_	
Blinding of participants and personnel* • Low risk				• Low risk	• Low risk
Personnel* • Low risk					
personnel* • Low risk				rote	
• Low risk			_	ecte	
			• Low risk	් . ව	
				y cc	

Risk of bias results for cohort/cross-sectional studies

Risk of bias res	sults for cohort/cross-sectional studies	BMJ Open	6/bmjopen-2018-027743 on 1 Au
Reference	Questions 1-4	Questions 5-7	Questions 8-10g
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Unclear
Backman A-S et al. 2012 ³⁰	The aims refer to "non-urgent" but the information is sought prior to visiting ED.	• No	"Health care information use in the past"
		6. Did the study	9. Were outcome assessors blinded?
	2. Was the study population clearly specified and	examine exposure	• Not applicable
	defined?	levels?	
	• Yes	• Yes	;//b
	4	Health advice seeking	10. Were confounders adjusted for?
			• Yes
	3. Was the participation rate at least 50%?		To some extent: participant and physician attributes
	• Yes	7. Were exposure	assessed for influence on the results.
	79%	measures clearly	η/ ο
		defined?	ă ≽
		• Unclear	P/) / pril :
	4. Were all the subjects selected or recruited from	Measures are vague, e.g.	, <u>o</u>
	the same or similar populations?	"previous use" of	024
	• Yes	information Also,	by (
	Primary care and ED attendees	discriminating between	on April 9, 2024 by guest.
		types of information	st. Pro
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Yes g
Carter 2018 ²⁶		• No	Attitudes and experiences of practice staff and

		BMJ Open	6/bmjopen-2018
			patients on web P.
	 2. Was the study population clearly specified and defined? Yes GPs, practice staff and their patients at 6 practices in Devon 	6. Did the study examine exposure levels?Not applicable	9. Were outcome assessors blinded? • Not applicable 120
	 3. Was the participation rate at least 50%? No Postal survey only had response rate of 35.1% but also GPs judgement of webGP requests and 5GPs and 5 administrators were interviewed. 	7. Were exposure measures clearly defined?Not applicable	10. Were confounders adjusted for? • Not applicable oaded from http://b
	 4. Were all the subjects selected or recruited from the same or similar populations? Yes GPs, practice staff and their patients at 6 practices in Devon 	evien	• Not applicable from http://bmjopen.bmj.com/ on April 9, 20
• Reference Cowie 2018 ²⁷	1. Was the research question clearly stated? • Yes	5. Was a sample size justification provided? • No	8. Were outcome measures clearly defined? • Yes • Yes
	2. Was the study population clearly specified and defined?Yes	6. Did the study examine exposure levels? • No	9. Were outcome assessors blinded? • No to t

			<u>. </u>
	3. Was the participation rate at least 50%?		• Yes Pes 27743 on 1 August 2019. Dow
	• No		43
	No for patient surveys	7. Were exposure	on ,
		measures clearly	A
		defined?	n Gr
	4. Were all the subjects selected or recruited from	 Not applicable 	st 20
	the same or similar populations?		019
	• Yes		. Do
			law
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes $\frac{\Phi}{\Delta}$
Joury et al.		• No	Scores used for Readability, popularity, content and
2016 US ³¹			quality #
	2. Was the study population clearly specified and		p://k
	defined?	6. Did the study	om _{je}
	Not applicable	examine exposure	9. Were outcome assessors blinded?
		levels?	• Not applicable
		 Not applicable 	, <u>j.</u> Q
	3. Was the participation rate at least 50%?	\ ///	om/
	Not applicable		10. Were confounders adjusted for?
		7. Were exposure	1 × *
		measures clearly	9,
	4. Were all the subjects selected or recruited from	defined?	203
	the same or similar populations?	 Not applicable 	124 b
	Not applicable		• Unclear April 9, 2024 by gues
	**		Le s
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Unclear	justification provided?	
Kellermann		• Not applicable	• Not applicable 2
2010 11		11	Ъу
-	2. Was the study population clearly specified and		9. Were outcome assessors blinded?

	1		<u> </u>
	defined?	6. Did the study	• Not applicable
	• Unclear	examine exposure	43 .
	Patients with influenza-like illness in US that accessed	levels?	on 1
	one of 2 websites http://www.flu.gov and www.H1N2	 Not applicable 	10. Were confounders adjusted for?
	ResponseCenter.com		• Not applicable
		7. Were exposure	019
	3. Was the participation rate at least 50%?	measures clearly	Do
	Not applicable	defined?	n w
		 Not applicable 	oac.
	1		e d
	4. Were all the subjects selected or recruited from		fron
	the same or similar populations?		2019. Downloaded from http://bmjopen.bm
	• Unclear		tp://
	Only counted web hits, no demographic data available		<u>b</u>
	on patients. No data on usage of algorithm by	81.	ope
	clinicians or call centers.		n. B
		10,	ற். c
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Yes
Lanseng &		• No	• Yes • Yes Use of TRI • Yes 9. Were outcome assessors blinded?
Andreassen			<u> </u>
2007	2. Was the study population clearly specified and		20
Norway ³²	defined?	6. Did the study	9 Were outcome assessors blinded?
1401 way	• Yes	examine exposure	
	103	levels?	• No guest
		• No	; *
	3. Was the participation rate at least 50%?	Readiness	10. Were conformders adjusted for?
	• Unclear	Remainess	1 4
	Oncion		• Unclear & by copy
		7 Word oversome	08
		7. Were exposure	<u>P</u>

	4. Were all the subjects selected or recruited from	measures clearly	8-0277
	the same or similar populations?	defined?	743 on
	• Yes	Not applicable	on 1
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes
Luger et al. 2014 ²³		• No	019. Do
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable
	• Yes	examine exposure	ed
		levels?	rom
		• No	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?		• Unclear
	• Unclear		omjo
		7. Were exposure	• Unclear //bmjopen.bmj.com/ on Apri
		measures clearly	n.bm
	4. Were all the subjects selected or recruited from	defined?	J j. cc
	the same or similar populations?	Not applicable	/mc
	• Yes		on ,
			Apri
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
 Reference 	• Yes	justification provided?	• Not applicable
Marco-Ruiz et		• No	4 b)
al. 2017			, gu
Norway ²⁴	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable
	• No	examine exposure)teci
		levels?	red_
		• No	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?		• Unclear $\frac{\ddot{\theta}}{\ddot{\theta}}$

	• Yes		27.
	53%		743
		7. Were exposure	027743 on 1 August 2019
		measures clearly	1 /
	4. Were all the subjects selected or recruited from	defined?	ρυγ
			ust
	the same or similar populations?	 Not applicable 	201
	• Unclear		9. [
	1 W 4 1 1 4 19	F 137	0 177 4 \$ 1 1 1 6 10
D 0	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes a
Nagykaldi		 Not applicable 	Web hits on customised practice website influenza
2010 ²⁵			self-managemen webpages. Downloads of self-
	2. Was the study population clearly specified and		management influeza toolkit. Completion of Iflueza
	defined?	6. Did the study	self-triage modite sessions. Volume of calls to
	• Yes	examine exposure	telephone hotlings. Qualitative feedback from patient
	Study population was patients from 12 primary care	levels?	on statisfaction with and utility of self-management
	practices in US.	 Not applicable 	websites and telephone hotline. Qualitative feedback
			from clinicians ground their involvement and their
		\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	perceptions of patient self-management techniques.
	3. Was the participation rate at least 50%?	7. Were exposure	1
	Not applicable	measures clearly	Apr
		defined?	9. Were outcome assessors blinded?
		 Not applicable 	• Not applicable
	4. Were all the subjects selected or recruited from		24 6
	the same or similar populations?		Ψ 9
	• Yes		10. Were confounders adjusted for?
	All participants were patients from 12 primary care		• Not applicable
	practices that accessed customised practice website or		
	telephone helpline		ctec
			• Not applicable Protected by copyright.
			C C C C C C C C C C C C C C C C C C C
			oyri ₍
			ght

	14	T =	0
_	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Yes $\frac{4}{\omega}$
Nijland 2009 ²⁹		• No	00 1
			Ac Ac
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• No 8
	• Yes	examine exposure	• No 2019
		levels?	D
		 Not applicable 	10. Were confognders adjusted for?
	3. Was the participation rate at least 50%?		• Yes
	• Unclear		Methods not very clearly reported but appears to be
	$\mathcal{N}_{\mathcal{O}}$	7. Were exposure	multiple regresson
		measures clearly	_ · ° <u>-</u>
	4. Were all the subjects selected or recruited from	defined?	nttp://bmjop
	the same or similar populations?	• Not applicable	ma'
	• Yes	Tvot applicable	jope
	100	Vi	n.bn
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Yes
Nijland 2016 ¹⁹		• No	on .
			Apr
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• No
	• Yes	examine exposure	• No 2024 by
		levels?	ý g
		Not applicable	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?	11	
	• No		rote
	Low participation rate in survey relative to users of	7. Were exposure	ctec
	triage system (though unclear how many were invited	measures clearly	d by
	to participate)	defined?	• Unclear Protected by copyrig
	το ραπισιραίε)	ucinicu:	

	T	T-	<u>, </u>
		Not applicable	027743 on 1 August
	4. Were all the subjects selected or recruited from		On On
	the same or similar populations?		
	• Yes		ngu
			Ist 2
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
 Reference 	• Yes	justification provided?	• Yes 💆
North et. al. 2011 ³⁴	Or	Not applicable	wnloac
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable
	• Yes	examine exposure	htt
		levels?	p:///
		• Yes	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?	Self-exposure	• Unclear
	• Not applicable		Some discussion of potential confounders.
		7. Were exposure	com/ on April 9
	4. Were all the subjects selected or recruited from	measures clearly	on ,
	the same or similar populations?	defined?	P h → ^A pri
	• Not applicable	Not applicable	19, 202
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Not applicable
Sole 2006 ¹⁸	"The primary purpose of this study was to identify and	• No	le st
	describe the demographic profile of students who used		. Pr
	the newly implemented Web-based triage system. A		9. Were outcome assessors blinded?
	secondary purpose was to compare Web-based triage	6. Did the study	• Not applicable
	diagnoses to the diagnoses made in clinic for a subset	examine exposure	, , b
			сору

of students who requested appointments"	levels?	10. Were confounders adjusted for?
	• Yes	• Not applicable
2. Was the study population clearly specified and		n 1 A
defined?	7. Were exposure	August 2019.
• Yes	measures clearly	st 2
Students who used the web based triage over a four	defined?	019
month implementation period (1290 students). Then of	• Yes	
those students, those who requested an appointment via		w _{nl}
email (143 students), then of those 59 who attended the		oadd
health centre after requesting an email appointment.		ed fr
7000.		Downloaded from http://bmjopen.bmj.com/
3. Was the participation rate at least 50%?		.p://k
Not applicable		omje
	C//.	ppen
		. <u>b</u>
4. Were all the subjects selected or recruited from	(6/2	00
the same or similar populations?		
• Yes		on A
		Apri

Risk of bias results for diagnostic studies

Reference	Questions 1 to 4	Questions 5 to 8	Questions 9 to 11 g
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical Information?
•	• No	verification	• Yes $\overset{\smile}{\mathfrak{Q}}$
Reference	Study participants were all patients registered at a student	avoided?	te d
Poote	health centre in England attending with new acute	Not applicable?	by c
			10. Were uninterpretable results reported?

		BMJ Open	6/bmjop
			• Not applicable
2014 17	symptoms. If the self-assessment triage system was only for students to be representative the study population		• Not applicable 777 43
	would have needed to include range of student health	6. Was the	On
	centres in different areas. If the system was for any UK	reference standard	11. Were withdrawals from the study explained?
	general practices the study population would have needed	independent of the	
	to include patients of all ages, ethnicity, gender etc from a	index test?	st 2
	range GP practices in different areas.	• Unclear	019
	, , , , , , , , , , , , , , , , , , ,	Patients took the	. Do
		assessment from self-	IUMC
	2. Acceptable reference standard?	triage through to	loac
	• Yes	their GP	de d
	\sim	consultation.	fron
	CO.		htt
	3. Acceptable delay between tests?		p:///
	• Yes	7. Index test results	om _j .
		blinded?	ope ₁
		• No	n.bn
	4. Partial verification avoided?	Patients took the	nj. Ω
	• Yes	assessment from self-)m/
	All patients that completed self-triage also had a GP	triage through to	On .
	consultation where the GP rated the urgency of their	their GP	Apri
	consultation.	consultation.	<u>,</u>
			202
			4 by
		8. Reference	/ gu
		standard results	est.
		blinded?	Pro
		• Yes	gust 2019. Downloaded from http://bmjopen.bmj.com/ on April 9, 2024 by guest. Protected
			<u> —</u>
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?

		BMJ Open	6/bmjopen-2018-
	• No	verification	• Yes 27
Reference	SORT was only trialled in 2 Emergency Departments in	avoided?	7743
Price 2013	US, a larger range would be needed for a representative	• Not applicable?	3 or
20	spectrum. Also, patients were from ED not home so	- Not applicable:	10. Were uninterpretable results reported?
	potentially sicker patients in the sample.		1 3 6
	potentially steker patients in the sample.	6. Was the	• Not applicable
		reference standard	2019
	2. Acceptable reference standard?	independent of the	11. Were withdraw sis from the study explained?
	• Yes	index test?	• No §
	Sensitivity of SORT for kids algorithm in identifying the	• Yes	loac
	need for ED care was based on an explicit gold standard:		lloaded from http://bmjopen.bmj.com/ on April 9
	documented evidence that the child received 1 or more of		fron
	5 ED-specific interventions.	7. Index test results	n ht
		blinded?	tp://
		• Yes	bmj:
	3. Acceptable delay between tests?	01	ope
	• Yes		n.bn
		8. Reference	nj. Co
		standard results	/mc
	4. Partial verification avoided?	blinded?	on
	• Yes	• Yes	April Pri
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?
•	• Unclear	verification	• Yes &
Reference	There were 45 standardised patient vignettes which were	avoided?	This is the clinical information that would be supplied by
Semigran	divided into three levels of triage urgency and included	Not applicable?	the patient which mag or may not differ from the
2015 4	more and less common conditions. It is not clear how		information given by the vignette.
	closely this replicates the spectrum of conditions that		[#548 Semigran 201 § pdf] Page 8: ion of the true clinical
	people use symptom checkers for.	6. Was the	accuracy of sympton Checkers.33 Some standardized
		reference standard	patient vignettes constained specifc clinical language (for
		independent of the	<u> </u>

	2. Acceptable reference standard?	index test?	example, mouth ulcers, tonsils with exudate), and actual
	• Yes	• Yes	patients with the sante condition might struggle with the
	[#548 Semigran 2015.pdf] Page 2: The source for each		words to use to describe their symptoms or use diferent
	vignette also provided the associated correct diagnosis.		terms. Therefore, ounanalysis represents an indirect
		7. Index test results	assess- ment of how well symptom checkers would perform
		blinded?	with actual patients No
	3. Acceptable delay between tests?	• Yes	7 10
	• Not applicable		D
			10. Were uninterpretable results reported?
		8. Reference	• Yes
	4. Partial verification avoided?	standard results	[#548 Semigran 2018 pdf] Page 3: ns for diagnosis and
	• Not applicable	blinded?	triage was high (Cohen's κ 0.90). In some cases we could
		• Yes	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 symp- tom in the vignette. To
			avoid penalizing these symptom checkers, we referred to
			standardized patient gignettes that successfully yielded an
		10,	output as "standardized patient evaluations."
			Standar diget pattern evaluations.
			on on
			11. Were withdrawals from the study explained?
			• Not applicable 9
			202
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?
•	• Unclear	verification	• Yes ©
Reference	There were 45 standardised patient vignettes which were	avoided?	The physicians and $t = \frac{1}{2}$ symptom checkers used the same
Semigran	divided into three levels of triage urgency and included	• Not applicable?	vignettes
2016 8	more and less common conditions. It is not clear how		otec
	closely this replicates the spectrum of conditions that		it ed.
	people use symptom checkers for.	6. Was the	10. Were uninterpretable results reported?
	1		
			yrigl
			Jt.

	BMJ Open	6/bmjopen-2018-
	reference standard	• Not applicable
	independent of the	743
2. Acceptable reference standard?	index test?	on `
• Yes	 Not applicable 	11. Were withdrawals from the study explained?
		• No gu
		It is unclear why the physicians chose to solve the specific
3. Acceptable delay between tests?	7. Index test results	vignettes $\frac{3}{6}$
Not applicable	blinded?	Do
	• Yes	Downloaded
		oade.
4. Partial verification avoided?		d fr
• No	8. Reference	from
There was a total of 234 physicians involved in the study	standard results	httt
and of the 45 vignettes, each was solved by at least 20	blinded?	o://b
physicians but it is not clear why they chose the specific	• Yes	mjo
vignettes to solve.	Chi	http://bmjopen.br

ij.com/ on April 9, 2024 by guest. Protected by copyright.



PRISMA 2009 Checklist

{		01 8	
Section/topic	#	Checklist item -027743	Reported on page #
TITLE		on 	
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT	<u> </u>	gust :	
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION		n loac	
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
8 Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, in reference, comparisons, outcomes, and study design (PICOS).	5
METHODS		9://b	
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
7 Information sources 8	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplion tensor and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
2 Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I²) for each meta-analysis. http://bmjopen.bmj.com/site/about/guidelines.xhtml	N/A



8 9

PRISMA 2009 Checklist

		Page 1 of 2		
Section/topic	#	Checklist item	Reported on page #	
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7	
RESULTS		9. D		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reach stage, ideally with a flow diagram.	9	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	10-16	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Appendix 2	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summare data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	17-22	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	22-23	
DISCUSSION				
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	26-27	
3 Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	28	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implication g for future research.	28-29	
FUNDING				
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	29	

42 From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The RISMA Statement. PLoS Med 6(7): e1000097. 43 doi:10.1371/journal.pmed1000097

The PRISMA for Abstracts Checklist

	Ž	
TITLE	CHECKLIST ITEM 743 on 1	REPORTED ON PAGE #
1. Title:	Identify the report as a systematic review, meta-analysis, or both.	1 (also in 'Design')
BACKGROUND	2019.	
2. Objectives:	The research question including components such as participants, interventions, comparators, and outcomes.	3 (Objectives)
METHODS	loaded ed	
3. Eligibility criteria:	Study and report characteristics used as criteria for inclusion.	3 (Eligibility criteria)
4. Information sources:	Key databases searched and search dates.	3 (Data sources)
5. Risk of bias:	Study and report characteristics used as criteria for inclusion. Key databases searched and search dates. Methods of assessing risk of bias.	3 (DE and synthesis)
RESULTS	om/ c	
6. Included studies:	Number and type of included studies and participants and relevant characteristics of studies.	3 (Results)
7. Synthesis of results:	Results for main outcomes (benefits and harms), preferably indicating the number of studies and participants for each. If meta-analysis was done, include summary measures and confidence intervals.	3 (Results)
8. Description of the effect:	Direction of the effect (i.e. which group is favoured) and size of the effect in terms meaningful to clinicians and patients.	3 (Results)
DISCUSSION	st. P	
9. Strengths and Limitations of evidence:	Brief summary of strengths and limitations of evidence (e.g. inconsistency, imprecisions indirectness, or risk of bias, other supporting or conflicting evidence)	3 (Results)
10. Interpretation:	General interpretation of the results and important implications	3 (Conclusions)

OTHER		018-02	
11. Funding:	Primary source of funding for the review.	7743	In text
12. Registration:	Registration number and registry name.	on 1	3

August 2019. Downloaded from http://bmjopen.bmj.com/ on April 9, 2024 by guest. Protected by copyright

BMJ Open

Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review

Journal:	BMJ Open
Manuscript ID	bmjopen-2018-027743.R2
Article Type:	Research
Date Submitted by the Author:	12-Jun-2019
Complete List of Authors:	Chambers, Duncan; The University of Sheffield, ScHARR Cantrell, Anna; The University of Sheffield, ScHARR Johnson, Maxine; The University of Sheffield, ScHARR Preston, Louise; The University of Sheffield, ScHARR Baxter, Susan; The University of Sheffield, ScHARR Booth, Andrew; The University of Sheffield, ScHARR Turner, Janette; The University of Sheffield, ScHARR
Primary Subject Heading :	Health services research
Secondary Subject Heading:	Diagnostics
Keywords:	urgent care, symptom checkers, systematic reviews

SCHOLARONE™ Manuscripts

Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review

Duncan Chambers¹, Anna Cantrell¹, Maxine Johnson¹, Louise Preston¹, Susan K Baxter¹, Andrew Booth¹ and Janette Turner¹

¹School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, Sheffield S1 4DA, UK

*Correspondence to Duncan Chambers: d.chambers@sheffield.ac.uk

Contributor/guarantor information:

DC contributed to the planning (project co-ordination and protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. AC contributed to the planning (protocol development), conduct (information retrieval, study selection, data extraction and quality assessment) and reporting (report writing) of the study. MJ contributed to the planning (protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. LP contributed to the planning (protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. SB contributed to the planning (protocol development), conduct (study selection, data extraction and quality assessment) and reporting (report writing) of the study. AB contributed to the planning (protocol development), conduct (information retrieval and study selection) and reporting (report writing) of the study. JT contributed to the planning, conduct and reporting of the study by providing expert topic advice at all stages. All the authors contributed to the study conception and design (protocol development), acquisition of data (study selection and data extraction) and analysis or interpretation of data (writing sections and/or commenting on drafts of the report). Duncan Chambers is the guarantor for this work. The corresponding author attests

that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

None of the authors have any competing interests

Copyright

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a non-exclusive worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.



Abstract

Objectives: In England, the NHS111 service provides assessment and triage by telephone for urgent health problems. A digital version of this service has recently been introduced. We aimed to systematically review the evidence on digital and online symptom checkers and similar services.

Design: Systematic review.

Data sources: We searched MEDLINE, EMBASE, the Cochrane Library, CINAHL, HMIC (Health Management Information Consortium), Web of Science and ACM Digital Library up to April 2018, supplemented by phrase searches for known symptom checkers and citation searching of key studies.

Eligibility criteria: Studies of any design that evaluated a digital or online symptom checker or health assessment service for people seeking advice about an urgent health problem.

Data extraction and synthesis: Data extraction and quality assessment (using the Cochrane Collaboration version of QUADAS for diagnostic accuracy studies and the National Heart, Lung and Blood Institute tool for observational studies) -were done by one reviewer with a sample checked for accuracy and consistency. We performed a narrative synthesis of the included studies structured around pre-defined research questions and key outcomes.

Results: We included 29 publications (27 studies). Evidence on patient safety was weak. Diagnostic accuracy varied between different systems but was generally low. Algorithm-based triage tended to be more risk-averse than that of health professionals. There was very limited evidence on patients' compliance with online triage advice. Study participants generally expressed high levels of satisfaction, albeit in mainly uncontrolled studies. Younger and more highly educated people were more likely to use these services.

Conclusions: The English 'digital 111' service has been implemented against a background of uncertainty around the likely impact on important outcomes. The health system may need to respond to short-term changes and/or shifts in demand. The popularity of online and digital services with younger and more educated people has implications for health equity.

Registration: PROSPERO (registration number CRD42018093564)

Strengths and limitations of this study

- This systematic review was based on a rigorous search of the literature which
 maximised efficiency by combining an initial focused search with subsequent rounds
 of follow-up searching, including searches for named symptom checker systems.
- Our narrative synthesis approach used a mixture of description and tabulation to summarise the evidence, including overall strength of the evidence base for each of the pre-specified outcomes of interest.
- Given the decision to implement a national urgent care service based on digital symptom checkers in the NHS in England, our study highlights areas of uncertainty that will need to be resolved by research and data collection.
- The review inclusion criteria were relatively broad and findings from symptom checker systems for specific conditions may not be applicable to more general systems and vice versa.
- We have also included studies of symptom checkers as part of electronic consultation systems in general practice, which again represents a slightly different setting from a general 'digital 111' service, and this should be kept in mind when interpreting the results.

Introduction

Digital and online symptom checkers and assessment services are used by patients seeking guidance about health problems, including some that may require urgent action. These services generally provide people with possible alternative diagnoses based on their reported symptoms and/or suggest a course of action (e.g. self-care, make a GP appointment or go to an emergency department (ED)).

In England, the NHS111 service provides assessment and triage by telephone for problems that are urgent but not classified as emergencies. The latest data from NHS England¹ show that in September2018 there were over 1.27 million calls to NHS111, an average of 42,400 per day. Outcomes of these calls were that 13.2% had ambulances despatched; 9.5% were recommended to attend an ED; 58.7% were recommended to attend primary care; 4.8% to attend another service; and 13.8% were not recommended to attend another service (e.g. their condition was considered suitable for self-care)

NHS England has recently introduced a digital platform to make NHS111accessible via a website or smartphone app. A beta version of the service (referred to as 'NHS111 Online') is available at https://111.nhs.uk/ (accessed 1 April 2019). The 'digital 111' service is seen as key to reducing demand for the telephone 111 service, enabling resources to be redirected to supporting 'integrated urgent and emergency care systems' as outlined in the 'NHS 5-year Forward View' and its 2017 update 'Next Steps on the NHS 5-year Forward View'²³.

There is an expectation that a digital 111 platform will help to manage demand and increase efficiency in the urgent and emergency care system, complementing the agenda of locally based Sustainability and Transformation Partnerships (STPs) which involve the health service and local government working together to integrate and co-ordinate care⁴. However, there is a risk of increasing demand, duplicating healthcare contacts (by increasing the number of potential access routes into the system) and providing advice that is not safe or clinically appropriate. For example, an evaluation of the NHS111 telephone service at four pilot sites and three control sites found that in its first year the service was not successful in reducing 999 emergency calls or in shifting patients from emergency to urgent care⁵. A recent study of 23 symptom checker algorithms providing diagnostic and triage advice that would form the

basis of a 'digital 111' platform found deficiencies in both their diagnostic and triage capabilities (based on patient vignettes)⁶.

In 2017, NHS England carried out pilot evaluations of different systems in four regions of England. The evaluations aimed to assess whether digital/online triage was acceptable to users and connected them to appropriate clinical care⁷. The full report of the evaluations was not yet published at the time of writing. The objective of this systematic review was to inform further development of the proposed digital platform by summarising and critiquing the previous research in this area, both from the UK and overseas. The overall research question was: for people seeking guidance about an urgent health problem, what is the effect of digital and online services designed to assess symptoms and signpost patients to appropriate services (compared with non-digital services or no comparator) on important clinical and health service outcomes? Outcomes include safety; clinical and cost-effectiveness; diagnostic and triage accuracy; impact on service use; patient/carer satisfaction; compliance with advice received; and outcomes related to equity and inclusion.

Methods

The review protocol was registered with PROSPERO (registration number CRD42018093564) and is available from the project website (https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/164717/).

Literature search and screening

Initial scoping searches revealed that a highly sensitive search strategy, as typically conducted for systematic reviews, retrieved a disproportionately high number of references on GP decision-making and triage as demonstrated by examination of sample search results (e.g. first 100). We therefore devised a three stage retrieval strategy as an acceptable alternative to comprehensive topic-based searching. This involved:

 Targeted searches of precise high specificity terms in seven databases (MEDLINE, EMBASE, the Cochrane Library, CINAHL, HMIC (Health Management Information Consortium), Web of Science and ACM Digital Library). These searches were not restricted by language or date. The search strategies used for this part of the review are presented in Appendix 1.

- 2. Phrase searching for names of known symptom checkers using a list compiled from Semigran 2015 and other sources
- 3. Citation searches and reference checking of key included studies and reviews, complemented by contact with service providers (directly and via websites).

The main literature search was completed in April 2018 and follow-up searches in May 2018. Search results were stored in a reference management system (EndNote) and imported into EPPI-Reviewer software for screening, data extraction and quality assessment. The search results were screened against the inclusion criteria by one reviewer, with a 10% random sample screened by a second reviewer. Uncertainties were resolved by discussion among the review team.

Inclusion and exclusion criteria

Population: General population seeking information online or digitally to address an urgent health problem, including adults and children and issues arising from both acute and long-term chronic illness

Intervention: Any online or digital service designed to assess symptoms, provide health advice and direct patients to appropriate services. Services that only provide health advice were excluded, as were those that offer treatment, e.g. online CBT services.

Comparator: The 'gold standard' comparator is current practice of telephone assessment (e.g. NHS111) or face to face assessment (e.g. general practice, urgent care centre or ED). However, studies with other relevant comparators (e.g. comparative performance in tests or simulations) or with no comparator were included if they addressed the research questions.

Outcomes: The main outcomes of interest were safety (e.g. any evidence of adverse events arising from following or ignoring advice from online/digital services); clinical effectiveness; costs/cost-effectiveness; accuracy; impact on service use; compliance with advice received; patient/carer satisfaction; and equity and inclusion. 'Accuracy covered 1) ability to provide a correct diagnosis and 2) ability to distinguish between high and low acuity/urgency problems (and hence direct patients to appropriate services).

Study design: We did not restrict inclusion by study design (and included relevant audits or service evaluations in addition to formal research studies) but included studies had to evaluate (quantitatively or qualitatively) some aspect of an online/digital service **Other**: Studies from any developed country healthcare system were eligible for inclusion **Excluded**: Purely descriptive studies, conceptual papers, projections of possible future developments and studies conducted in low or middle income countries were excluded from the review.

Data extraction and quality/strength of evidence assessment

We extracted and tabulated key data from the included studies, including study design, population/setting, results and key limitations. Data extraction was performed by one reviewer, with a 10% random sample checked for accuracy and consistency.

To characterise the included digital and online systems as interventions, we identified studies reporting on a particular system and extracted data from all relevant studies using a modification of the TIDieR (Template for Intervention Description and Replication) checklist⁸ which we designated TIDieST (Template for Intervention Description for Systems for Triage). Further details may be found in the full report (Chambers et al., *Health Services & Delivery Research* 2019 (in press)).

Quality (risk of bias) assessment was undertaken for peer-reviewed full publications only (i.e. not grey literature publications (such as research reports, working papers, or reports produced by government departments, academics, business and industry) or conference abstracts). Randomised controlled trials were assessed using the Cochrane Collaboration risk of bias tool. For diagnostic accuracy type studies, we used the Cochrane Collaboration version of QUADAS⁹ and for other study designs we used the National Heart Lung and Blood Institute tool for observational cohort and cross-sectional studies (https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools, accessed 25th March 2019). Quality assessment was performed by one reviewer, with a random 10% sample checked for accuracy and consistency.

Assessment of the overall strength (quality and relevance) of evidence for each research question is part of the narrative synthesis. Overall strength of the evidence base for key

outcomes was assessed using an adaptation of the method described by Baxter et al.¹⁰ This involves classifying evidence as 'stronger', 'weaker', 'conflicting' or 'insufficient' based on study numbers and design. Specifically, "stronger evidence" represented generally consistent findings in multiple studies with a comparator group design or comparative diagnostic accuracy studies; "weaker evidence" represented generally consistent findings in one study with a comparator group design and several non-comparator studies or multiple non-comparator studies; "very limited evidence" represented an outcome reported by a single study; and finally, "inconsistent evidence" represented an outcome where fewer than 75% of studies agreed on the direction of effect. All studies in the review, including those that did not meet criteria for risk of bias assessment, were included in the strength of evidence assessment.

Evidence synthesis

We performed a narrative synthesis structured around the pre-specified research questions and outcomes. We did not perform any meta-analyses because the included studies varied widely in terms of design, methodology and outcomes.

Patient and public involvement (PPI)

The review was discussed at two meetings of an existing PPI group covering the programme from which the review was commissioned (Sheffield HS&DR Evidence Synthesis Centre). At the meetings there was discussion regarding the focus of the work, including a presentation on previous research on NHS111 telephone services to provide a context for understanding the current work. The meetings also included presentation and discussion of the findings of the review, in order to explore key messages for patients which could inform dissemination of the findings. Discussion during one meeting was structured using a SWOT (strengths, weaknesses, opportunities and threats) analysis approach, which revealed a number of potential concerns amongst patients (e.g. reliability and consistency; high costs of programming and development; whether patients would follow advice given; and threats to equity) as well as potential perceived benefits (e.g. improved access to care at all hours; value to those who might feel embarrassed discussing their problem with a health professional). Involvement of the advisory group was beneficial in highlighting some issues that had also emerged from the systematic review, and enabled the reviewers to structure the review findings taking this into account. For example, the group's uncertainty about the likely

impact of 'digital 111' was reflected in the review findings and recommendations for ongoing evaluation and further research. The review report also reflects the group's relatively cautious attitude (while recognising the need to update the way services are accessed) which contrasts with the strong belief in some quarters that 'digital 111' will help to ensure that patients receive appropriate care more quickly while reducing 'inappropriate' visits to EDs and GP appointments.

Results

Results of literature search

Twenty-seven studies (29 publications) were included in the review. Figure 1 presents the flow of studies through the selection process. Inter-rater agreement on initial study selection was moderate (Kappa = 0.582). This reflects a degree of learning by the review team: our initial sift of the search results consciously favoured inclusivity and items found not to meet the inclusion criteria on detailed examination were subsequently discarded.

Figure 1: PRISMA flow diagram

Characteristics of included studies

Seventeen studies (Table 1) evaluated symptom checkers as a self-contained intervention, of which eight covered a limited range of symptoms, e.g. respiratory¹¹⁻¹³ or gastrointestinal^{14 15} symptoms which we considered to be 'urgent'. The remaining studies in this group evaluated symptom checkers covering a wider range of common urgent care symptoms. Studies either evaluated a single system¹⁶⁻¹⁹ or multiple systems^{6 20}. We found only one study of a symptom checker specifically intended for assessment of children's symptoms, a development of the SORT (Strategy for Off-Site Rapid Triage) system for influenza-like illness²¹ Two reports with some overlap of content evaluated the 'babylon check' app^{16 22}. Studies were conducted in the USA, UK or other European countries.

Five studies (four from the UK)^{7 23-26} evaluated symptom checkers as part of a broader self-assessment and consultation system (often referred to as electronic consultation or e-consultation). Study characteristics are summarised in Table 2. In this type of system, the role of symptom checkers is to help patients decide whether their symptoms require a consultation with a doctor or other health professional or can be dealt with by self-care. If a consultation is required, details of the symptoms and a request for an appointment or call-back can be submitted electronically. This type of study is important because it considers the service within the broader context of the urgent and emergency care system. A limitation is that some studies focused mainly on the 'downstream' elements of the pathway, e.g. consultation with GPs, and provided limited data on the symptom checker element of the system.

A final group of five studies examined patient and/or public attitudes to online self-diagnosis in the context of urgent care²⁷⁻³¹. See the full report for further details (Chambers et al. *Health Services & Delivery Research* 2019 (in press)).

Table 1: Studies of symptom checkers as a self-contained intervention

D. C	G. 1 1 ·		j	D 1 (* / 1
Reference	Study design	System type	Comparator _{\rightarrow}	Population/sample
Babylon	 Uncontrolled 	• Digital	• Health professiona	General population
Health 2017 ²²	observational	Smartphone app	performance on real-world	Participants in the London
UK	No control group but		data 3	pilot evaluation of 'digital 111'
	some comparison with	/	• Other	services
	NHS111 telephone		NHS111 data for 12 <u>\breee</u>	
	data		months from Februa	
		1 0	2017	
Berry 2016 ¹⁴	Simulation	• Online	• None	Specific condition(s)
USA	Evaluation of	17 symptom checkers	• None http://bmjope	Gastrointestinal symptoms
	symptom checker		þ://k	
	performance on		mji	
	clinical vignettes	· C1	pper	
Berry 2017 ³²	Controlled	• Online	Health professionals	Specific condition(s)
USA	observational	Three online symptom checkers (WebMD, iTriage and	performance on real-world	Patients with a cough
		FreeMD)	data	presenting to an internal
			on /	medicine clinic
Berry 2017 ¹⁵	• Controlled	• Online	• Health professiona E.	Specific condition(s)
USA	observational	Three online symptom checkers (WebMD, iTriage,	performance on real-world	Abdominal pain
		FreeMD)	data	
Kellermann	Simulation	• Online	• Health professiona	Specific condition(s)
2010 11	The developed	SORT was available on 2 interactive websites	performance on real world	Influenza symptoms
USA	algorithm was tested		data &	
	against past patient		The algorithm was tested	
	records		against clinicians'	
			decision on past patient	
			records.	
	1	1	. <u> </u>	1

Little 2016 ¹²	Experimental	• Online	• Other	• Specific condition(s)
UK	Randomised	'Internet Doctor' website	Usual GP care with out	Respiratory infections and
	controlled trial (RCT)		access to the Interne₽	associated symptoms
	, , ,		Doctor website ≥	
Luger et al.	Simulation	• Online	• Other	General population
2014 33	Described as	Google and WebMD	Comparing two integnet	Older adults (50 years or
USA	"human-computer		health tools.	older)
	interaction study"		Do	
	using think-aloud		wn	
	protocols.		oad	
Marco-Ruiz	• Qualitative	• Online	Downloaded from http://bmjopen.bmj.com/ on None	General population
et al. 2017 ³⁴	Qualitative element	Erdusyk	rom m	Internet tool users
Norway	• Other	10/2	http	
	1. Online evaluation)://b	
	by users (problem		<u>nj</u> o	
	detection) 2. Think	• Online Erdusyk	pen	
	aloud technique by		.bm	
	smaller sample of	'612	j.co	
	participants		m/ c	
	(usability)		on A	
Middleton	Simulation	• Digital	• Health professiona €	General population
2016^{16}		'babylon check' automatic triage system	performance on ,o	
UK			test/simulation	
			Twelve 'clinicians' 🧸	
			(doctors) and 17 nurses	
Nagykaldi	Uncontrolled	• Online	• None	Specific condition(s)
2010 35	observational	Customised practice website including a bilingual	st. Protected by copyr	Influenza
USA		influenza self-triage module, a downloadable influenza	ecte	
		toolkit and electronic messaging capability. A	d b	
		bilingual seasonal influenza telephone hotline was	ر ا	

		available as an alternative.)277	
Nijland 2016 ¹⁹ Netherlands	• Uncontrolled observational Retrospective	Online Web-based triage system (http://www.dokterdokter.nl)	• None 3 on 1 Aug	General population
	analysis of 15 months' data		August 20	
Poote 2014	• Uncontrolled	• Online	• Health professiona	General population
17	observational	Prototype self-assessment triage system	performance on real-world	Students attending a
UK			data	University Student Health
			GPs triage rating was	Centre with new acute
			compared with rating from	symptoms
		700	the self-assessment system	
Price 2013	Uncontrolled	• Online	• Health professional	Specific condition(s)
21	observational	A web-based decision support tool - Strategy for Off-	performance on real-world	Influenza in children
USA		site Rapid Triage (SORT) for Kids designed to help	data 💆	
		parents and adult caregivers decide whether a child	The sensitivity of the	
		with possible influenza symptoms needs to visit the	algorithm was compa <mark>z</mark> ed	
		emergency department for immediate care.	with a gold standard \S	
			evidence form child's	
			medical records that hey	
			received 1 or more of ₹	
			ED-specific interventions.	
Semigran	 Experimental 	• Multiple	• Other	General population
2015 6	Described as an audit	23 symptom checkers were evaluated. Symptom	Vignettes had a diagnosis	Where a single class of
N/A	study	checkers available as apps (via the App Store and	and triage attached to	illness was examined by the
		Google Play) were identified through searching for	them and these were.	symptom checker, the
		"symptom checker" and "medical diagnosis" and	compared against the	symptom checker was
		screened the first 240 results. Symptom checkers	symptom checker ad ce.	excluded from the study.
		available online were identified through searching	ا م	
		Google and Google Scholar for "symptom checker"	by со	

			6	
		and "medical diagnosis" and screened the first 300 results.	27743	
Semigran	Experimental	• Multiple	• Health professiona	General population
2016 20	Comparison of	"Human Dx is a web-and app based platform"	· –	Of the 45 condition vignettes
USA	physician and		test/simulation	- there were 15 low, 15
	symptom checker		performance on test/simulation Clincial vignettes - &	medium and 15 huigh acuity
	diagnoses based on		comparison of 23 symptom	vignettes - there were 26
	clinical vignettes		checkers with physican	common and 19 uncommon
		Or	diagnosis for 45 vignettes	condition vignettes
Sole 2006 18	Uncontrolled	• Online	• Health professiona	General population
USA	observational	(0.4/7.17/1.16.1)	1 6 1 1 1 1 1 1 1	General population
USA	Descriptive	A web-based triage system (24// webinea)	data data	
	comparative study		Data was evaluated #rom	
	comparative study	1	students who had used the	
		· 01	web based triage and then	
			requested an appointment	
		A web-based triage system (24// WebMed)	via email (so triage data	
			NWAS AVAILABLE TOT	
			comparison).	
Yardley	Experimental	• Online	• Other	Specific condition(s)
2010^{13}	Exploratory	'Internet Doctor' website	Self-care information	Minor respiratory symptoms,
UK	randomised trial		provided as a static web	e.g. cough, sore throat, fever,
			page with no sympto E	runny nose
			checker or triage ad cice	
			st. F	
			^o rote	
			ecte	
			ρ	
			\ c <u>o</u>	
			Protected by copyright	
			ght.	15

Table 2: Studies of symptom checkers as part of an electronic consultation system

Table 2: Stu	ıdies of symptom ch	BMJ Open eckers as part of an electronic consultation	6/bmjopen-2018-027743 on 1 system	
Reference	Study design	System type	Aug	Population/sample
Carter 2018	• Uncontrolled	• Online	Comparator § S	• General population
23	observational	webGP(subsequently known as eConsult)	Investigate patient experience by	General practices in NHS
UK	Mixed-methods	webol (subsequently known as econsall)	surveying patients who had us&l	Northern, Eastern and
	evaluation	· O ₄₅	webGP and comparing their	Western Devon Clinical
			experience with controls (patients	Commissioning Group's
			who had received a face-to-face	area
		700	consultation during the same time	
		Ch	period) matched for age and gender	
Cowie	• Uncontrolled	• Online	• None	General population
2018 ²⁴	observational	eConsult, accessed via GP surgery websites.	n)jop	Patients registered with
UK	6-month evaluation	Service provides self-care assessment and	en.b	participating GP practices
	at 11 GP practices	advice, including symptom checkers; triage and	omj	
	in Scotland	signposting to alternative services; access to	com	
		NHS24 (phone service); and e-consults	on	
		allowing submission of details by e-mail)	• None • None • None	
Madan	Uncontrolled	• Online	• None	General population
2014 ²⁵	observational	webGP (subsequently known as eConsult)	200	
UK	Report of 6-month		24 b	
) H I G	pilot study	26111	9	
NHS	• Uncontrolled	• Multiple	• None	General population
England ⁷	observational	Pilots featured NHS Pathways (Web-based;	Authors stated it was not appropriate	
UK	Analysis of data	West Yorkshire); Sense.ly ('voice-activated	to compare pilot sites because of	
	from four pilot	avatar'; West Midlands); Espert 24 (Web-	differences in starting date,	
	studies together	based; Suffolk) and babylon (app; North	'footprints' covered, method of uptake	
	with data from other	Central London)	and underlying population	

			8-0	
	sources		277	
Nijland	• Other	• Online	• None &	 General population
2009^{26}	Online survey	Responses of interest relate to 'indirect e-)n 1	Patients with Internet access
Netherlands		consultation' (consulting a GP via secure e-mail	Au	but no experience of e-
		with intervention of a Web-based triage system)	gus	consultation

Results by outcome

Safety

None of the six included studies that reported on safety outcomes identified any problems or differences in outcomes between symptom checkers and health professionals. Most of the studies compared system performance with that of health professionals using real or simulated data. The only study with no comparison group was the 6-month pilot study of webGP²⁵, which reported 'no major incidents'.

Limitations of the studies included not being based on real patient data¹⁶; covering only a limited range of conditions^{11 21}; and sampling a young healthy population (students) not representative of the general population of users of the urgent care system¹⁷. Studies of e-consultation systems did not generally collect data on those respondents who decided not to seek an appointment, limiting their ability to assess any impact on safety for this group. Overall, the evidence should be interpreted cautiously as indicating no evidence of a detrimental impact on safety rather than evidence of no detrimental effect.

Clinical effectiveness

Only two studies reported on clinical effectiveness outcomes, making it difficult to draw any firm conclusions. In the study by Little et al., those who used the Internet Doctor website experienced longer illness duration and more days of illness rated moderately bad or worse than the usual care group¹² The pilot study of the webGP system²⁵ reported that several patients received advice to seek treatment for serious symptoms that might otherwise have been ignored. However, no details or quantitative data were provided.

Costs/cost-effectiveness

Two included studies provided limited data on possible cost savings. Based on 6 months of pilot data, Madan²⁵estimated savings of £11,000 annually for an average general practice (6,500 patients) compared with current practice. The report also suggested a saving to commissioners equivalent to £414,000 annually for a CCG (Clinical Commissioning Group, responsible for specifying and purchasing most health services in the NHS in England) covering 250,000 patients. These savings were specifically related to self-reported diversion of patients from GP appointments to self-care and from urgent care to e-consultation. Using

similar methodology, the manufacturers of the 'babylon check' app claimed average savings of over £10/triage compared with NHS111 by telephone, based on a higher proportion of patients being recommended to self-care²².

Diagnostic accuracy

Eight studies reported at least some data on the diagnostic accuracy of symptom checkers. In spite of the diverse methods and comparisons in the included studies, almost all agreed that the diagnostic accuracy of symptom checkers was poor in absolute terms (e.g. in evaluating 'vignettes' designed to test knowledge of specific conditions, where the correct diagnosis was already known by definition) or relative to that of health professionals. In the most comprehensive evaluation, Semigran et al. evaluated 23 symptom checkers across 770 standardised patient evaluations⁶. Overall the correct diagnosis was made in 34% of cases (95% CI 31%-37%), although performance varied widely between symptom checkers, high and low acuity conditions and common and rare conditions. When the same authors compared the 23 symptom checkers with physicians using 43 vignettes, physicians were more likely to list the correct diagnosis first (out of three differential diagnoses) (72.1% vs. 34% p<0.001) as well as among the top three diagnoses (84.3% vs. 51.2% p<0.001)²⁰.

The only exception to the rule was an evaluation carried out at a student health centre¹⁸. Using data from 59 participants who used the 24/7 WebMed system and who were subsequently treated at the health centre, the study found good agreement between chief complaint, 24/7 WebMed classification and provider diagnosis (kappa values 0f 0.89 to 0.94). This study differed from the others in using data from students rather than a general population sample. In addition, the students' complaints were generally common and uncomplicated, a scenario in which symptom checkers performed relatively well in the study by Semigran et al.²⁰.

Accuracy of disposition (triage and signposting to appropriate services)

Six included studies reported on this outcome, all except one of which¹⁵ evaluated a 'general purpose' symptom checker. As with diagnostic accuracy, diverse methodologies and outcome measures were used.

The results overall presented a mixed picture but most studies indicated that symptom checkers were inferior and/or more cautious in their triage advice compared with doctors or other health professionals. In their review of 23 symptom checkers, Semigran et al. found that the systems provided appropriate triage advice in 57% (95% CI 52% to 61%) of cases⁶. Performance varied across the systems evaluated, correct triage ranging from 33% to 78%. The NHS England pilot evaluation of four systems⁷ found that agreement with clinical experts varied from 30% to 95%, although the number of responses also varied, reducing the comparability of the results.

For abdominal pain, Berry et al. evaluated three symptom checkers and found that 33% of diagnoses were at the same level of urgency as physician diagnoses (emergency, non-emergency or self-care); 39% were diagnosed as more serious and 30% less serious than the physician's judgement¹⁵. A similar level of agreement between algorithm and clinician (39%) was reported by Poote et al.¹⁷, while the system evaluated by Nijland et al. advised patients to visit a doctor in 85% of cases, even when the symptoms were appropriate for self-care¹⁹.

The only studies to report clearly equal or superior accuracy of disposition using an automated system were the evaluations of Babylon check by the company that developed the system. Middleton et al. 16 reported that using patient vignettes, the app gave an accurate triage outcome in 88.2% of cases, compared with 75.5% for doctors and 73.5% for nurses (unaware of the 'correct' diagnosis for the vignettes). When vignettes were delivered by a medical professional rather than actors, the accuracy of Babylon check increased to over 90%. A later report looked at triage results obtained as part of the NHS England pilot evaluation, concluding that all of 74 referrals to urgent or emergency care were appropriate²².

Impact on service use/diversion

Eight studies reported on this outcome, although one of them¹¹ merely stated that it was not possible to assess the effect of the intervention (a web-based influenza triage system) on patients' use of health services.

The pilot evaluation of the webGP system reported that 18% of users planned to book an appointment but chose not to do so²⁵ In addition, 14% of users reported that they would have

attended a walk-in centre or other urgent care service if they had not had access to the webGP system.

The NHS England pilot evaluation of four online/digital systems in different regions of England⁷ compared the recommendations of the digital systems with those of the NHS111 telephone service over a similar time period (the first months of 2017). Compared with the telephone service, the online and digital services directed a slightly higher proportion of patients to self-care (18% vs. 14%) and a lower proportion to other primary care services such as GPs, dental and pharmacy (40 vs. 60%). The manufacturer's data on the 'babylon check' app collected as part of the NHS England evaluation indicated that patients were more likely to be triaged to self-care by the app compared with NHS111 by telephone (40 vs. 14%)²². This figure includes people who received information leaflets on self-care as well as those who were actively triaged. If the former group is excluded, the figures for the two services are similar (14% for NHS111 and 15.6% for 'babylon check'²².

In their study of self-assessment for students attending a university health centre, Poote et al. found that the prototype system they studied was able to identify a proportion of cases that doctors considered appropriate for self-care, suggesting a potential to reduce service use¹⁷. Similarly, Little et al's RCT of a web-based symptom checker designed to support self-care for respiratory symptoms¹² reported that patients in the intervention group had fewer contacts with doctors than the usual care control group despite having a longer duration of illness and more days with relatively severe symptoms. This was balanced by an increase in contacts with the NHS Direct telephone service (which preceded NHS 111) and it should be noted that the system under evaluation recommended people needing treatment to contact NHS Direct rather than go directly to a doctor. Finally, a study of young adults (students) found that intention to seek treatment for a hypothetical illness was stronger when the diagnosis was made with the aid of WebMD or Google than with no electronic aid³⁰.

Patient compliance with triage advice

Only two of the included studies reported specifically on patients' compliance (or intention to comply) with advice received. The NHS England pilot evaluation in four regions asked participants in two of those regions (Suffolk and London) what they intended to do based on the advice received⁷. No quantitative data were provided but the report stated that in the

Suffolk pilot, 'overall users would have followed the advice given'. However, those who were recommended to call 999 or attend an ED were more likely to seek advice from primary care or self-management. Similarly, in the London region there was generally good agreement between advice and intended action but patients recommended to call 999 or go to an ED indicated that they would seek advice from a GP. In a study of a web-based triage system in the Netherlands, 192 patients were asked about their intention to comply immediately after receiving advice from the system¹⁹. Thirty-five patients responded to a follow-up survey on actual compliance, of whom 20 (57%) reported that they had followed the advice. Compliance was correlated with intention to comply, which in turn was correlated with the patient's attitude towards the advice received.

Equity and inclusion

Fourteen studies investigated the outcome of equity and inclusion or compared users and non-users. One study ¹²reported that patients who were classed as less deprived were more likely to agree to use "Internet Doctor" than decline participation, although no relationship was found between deprivation and results in this study or between e-Consult use and deprivation in another study ²⁴. Association between e-consultation use and education levels was explored in a third study. Patients with low to medium levels of education tended to be motivated toward indirect e-consultation (which involves contact with a health professional via e-mail), mainly to reduce uncertainty ²⁶

Evidence from included studies suggests that users of e-consultation were more likely to be young⁷ ²³⁻²⁵, employed¹⁹ ²³ ²⁵ and female⁷ ¹⁹ ²⁴ ²⁵than non-users. One study also found a significantly larger use by white patients (78%) than other ethnicities²⁴.

Risk of bias assessment

We assessed risk of bias in the two included RCTs¹² 13 using the Cochrane risk of bias tool. Thirteen studies¹¹ 18 19 23 24 26-29 31 33-35 were assessed with the tool for cross-sectional and cohort studies and four (six publications⁶ 17 20 21 36 37) with the modified QUADAS tool. Seven grey literature reports and conference abstracts were not formally assessed for risk of bias⁷ 14-16 25 30 32. Identified limitations were extracted for all included studies.

Risk of bias results are presented in Appendix 2. With the possible exception of the two randomised trials, the included studies generally had at least a moderate risk of bias. However, the diverse designs and objectives of the studies made risk of bias difficult to assess in some cases with the available tools. Grey literature reports containing relevant data were included in the review but not formally assessed for risk of bias. Reports prepared by individuals with a commercial interest in a specific system and published without independent peer review^{16 25} should be treated with particular caution because of possible conflicts of interest.

Overall strength of evidence assessment/evidence map

The overall strength of evidence for key outcomes is summarised in Table 3. We found relatively strong evidence that the diagnostic accuracy of digital and online symptom checkers tends to be lower than that of health professionals; and that patients who have used these systems generally show high levels of satisfaction (mainly in non-comparative studies). Areas where evidence is lacking or inconsistent include clinical and cost-effectiveness, accuracy of disposition to appropriate services and patient compliance with advice received. For safety, we found no evidence of an increased risk with digital/online systems but the available evidence was weak.

Outcome	Relevant studies	Evidence statement	Strength of evidence	Comments of
Safety	=Kellermann 2010 ¹¹ =Little 2016 ¹² =Middleton 2016 ¹⁶ =Poote 2014 ¹⁷ =Price 2013 ²¹ Madan 2014 ²⁵	No evidence of a difference in risk between health professionals and symptom checkers	Weaker	Rating changed from stronger based on study numbers and design to weaker because of low numbers of adverse events reported
Clinical effectiveness	-Little 2016 ¹² ?Madan 2014 ²⁵	Insufficient evidence to draw any firm conclusions	Very limited	ad from
Costs/cost- effectiveness	+Babylon Health 2017 ²² +/-Cowie 2018 ²⁴ +Madan 2014 ²⁵	Insufficient evidence to draw any firm conclusions	Inconsistent	http://bmjop
Diagnostic accuracy	?Berry 2016 ¹⁴ -Berry 2017 ³² -Berry 2017 ¹⁵ -Price 2013 ²¹ ?Semigran 2015 ⁶ -Semigran 2016 ²⁰ =Sole 2006 ¹⁸	Symptom checkers appear inferior to health professionals in terms of diagnostic accuracy	Stronger	Mainly for specific conditions or pre-prepared vignettes 3.
Disposition accuracy	=Babylon Health 2017 ²² -Berry 2017 ¹⁵ =Middleton 2016 ¹⁶ ?Nijland 2010 ¹⁹ -Poote 2014 ¹⁷ +/-Semigran 2015 ⁶	Inconsistent findings on accuracy of disposition	Inconsistent	Performanc variable between different systems by guess. Protected by cop

Outcome	Relevant studies	Evidence statement	Strength of evidence	Comments 774
	+/-NHS England 2017 ⁷			on 1 /
Service	?Kellermann	Inconsistent findings on effects on service use	Inconsistent	- Qu
use/diversion	201011	-		ust
	+/-Little 2016 ¹²			ugust 2019. Downloaded from http://b
	+/- Poote 2014 ¹⁷			19.
	?Carter 2018 ²³			Do
	?Cowie 2018 ²⁴			l w
	+Madan 2014 ²⁵			oac
	+/- NHS	' ' ' ' ' ' ' ' ' '		<u>a</u>
	England 2017 ⁷	$\mathcal{O}_{\mathcal{O}}$		fro
	+Babylon Health			3 -
	2017 ²²			₫
	-Luger 2011 ³³			₹
Compliance	?Nijland ¹⁹	No comparative data on compliance	Very limited	mjoper
	?NHS England			Per
	20177			5
Patient/carer	?Nagykaldi	Most studies report high rates of patient	Weaker	Few studies with comparator data
satisfaction	2010^{35}	satisfaction with symptom checkers and e-	· 11.	om
	?Nijland ¹⁹	consultation systems generally		on on
	?Price 2013 ²¹			Ap
	+Yardley ¹³			71
	?Carter 2018 ²³			φ, ν
	?Cowie 2018 ²⁴			022
	?Madan 2014 ²⁵			(d 1
	?NHS England) gu
	2017 ⁷			April 9, 2024 by guest
C + 11 1 +	?Lanseng 2007 ²⁹		1	··· · · · · · · · · · · · · · · · · ·

Controlled studies in bold; = means no significant difference in outcomes; + means better outcome with symptom checker; +/- varying results within study; ? results difficult to interpret in comparative terms

Discussion

Main findings

The literature search identified 29 publications describing 27 studies that met the inclusion criteria. The overall strength of the evidence base varied between outcomes (Table 3), but in absolute terms the evidence is weak, being based largely on observational studies. A substantial component of grey literature of uncertain quality complicates the interpretation of the evidence. Interpretation of the evidence should also take into account risks of bias in individual studies. In addition, one included study evaluated 23 symptom checkers and only the overall findings are summarised in this review⁶.

We found little evidence to indicate whether or not digital and online symptom checkers are detrimental to patient safety. The studies that reported on the outcome were mostly short-term and involved relatively small samples and hence reported few or no adverse events. Some were limited to people with specific types of symptoms and others recruited from specific population groups not representative of typical users of urgent care services. This body of evidence should therefore be interpreted cautiously and not extrapolated to the possible impact of a nationally available digital urgent care service being used by millions of people annually.

The evidence on patient satisfaction with digital and online systems also had some limitations but these findings appear more likely to be generalisable. Study participants generally expressed high levels of satisfaction, albeit in uncontrolled studies. For example, in the NHS England pilot evaluation, 70–80% of users were satisfied with their experience at each of the pilot sites⁷. This evidence, together with the increasing reliance on digital technology in all areas of life, suggests that any national digital urgent care service may be popular and well-used, although different sections of the population may differ in their degree of engagement (see the discussion of equity and inclusion below).

Digital and online systems have yet to achieve a high level of accuracy in the diagnosis of specific conditions. This finding applies both to 'general purpose' symptom checkers and to those limited to particular conditions. Although the evidence was classified as relatively

strong, several caveats should be applied. Some of the included studies did not recruit representative populations and others were based on standardised vignettes rather than real-world data. In addition, studies that compared symptom checkers with health professionals tended to use the doctors' clinical diagnosis as the reference standard, which would bias the comparison in favour of the health professionals. Poor diagnostic accuracy could also have implications for patient safety, although the limited evidence on safety outcomes (small samples and small numbers of events) makes it difficult to draw any firm conclusions. If symptom checkers are generally risk averse, this could potentially mitigate any effects on safety.

Accuracy of signposting of patients to the most appropriate level of service is closely related to diagnostic accuracy, but results for this outcome were inconsistent between studies. In general, algorithm-based triage tended to be more risk-averse than that of health professionals, with 85% of respondents being advised to visit their doctor in one study¹⁹. While there is considerable uncertainty about the magnitude of the effect, a national digital urgent care service could result in considerable numbers of patients receiving inappropriate advice to visit the ED or request an urgent GP appointment. Middleton and colleagues¹⁶ claimed that the 'babylon check' app had a high degree of triage accuracy for vignettes compared with health professionals, but this non-peer-reviewed report requires further validation.

We also found inconsistent evidence on effects on service use. There was some indication that symptom checkers can influence the pattern of service use but the magnitude and direction of the effect varied between studies. Patients' reactions to online triage advice and whether they follow the advice or seek further help or information would have implications for service use but we found limited evidence for this outcome. Preliminary findings from the NHS England evaluation suggest that patients may be more likely to seek further advice for more urgent conditions⁷ but further confirmation is required.

Over half of the included studies considered equity and inclusion issues either directly or by comparing users and non-users of digital triage systems. Not surprisingly, studies revealed a clear consensus that younger and more highly educated people are more likely to use these services while older and less educated patients are more likely to prefer telephone or face-to-face contact. This could have implications for health equity if urgent care pathways prioritise

(or appear to prioritise) requests originating from digital sources. Problems have arisen in primary care because patients using e-consultation systems to request an appointment following online triage may be seen more quickly than those contacting the practice by telephone.

Strengths and limitations

This systematic review was undertaken on a short timescale using a relatively large team of experienced researchers, including both methodological and topic experts. We performed a rigorous search of the literature including reference checking and citation searching. Rather than a conventional highly sensitive search (which would have resulted in inefficiencies in the screening process), we combined an initial focused search with subsequent rounds of follow-up searching, including searches for named symptom checker systems. We assessed risk of bias in individual studies using a variety of appropriate checklists as well as summarising the overall strength of evidence for key outcomes (Table 3).

The heterogeneous and descriptive nature of the included studies meant that meta-analysis was not feasible for any of the outcomes of interest. Our narrative synthesis approach used a mixture of description and tabulation to summarise the evidence for each of the pre-specified outcomes of interest. This was a review of published (including non-peer-reviewed) literature and the coverage of systems is not exhaustive; for example, we did not extract data from websites. We also did not carry out any original analyses of raw data even where such data were available. The timing of the review meant that final results of NHS England's pilot evaluation were not available to us. We were able to make use of a draft report that was published online⁷ but we acknowledge that the findings of the final evaluation report, when available, will supersede those of the 2017 draft.

The review inclusion criteria were relatively broad and findings from symptom checker systems for specific conditions may not be applicable to more general systems and vice versa. We have also included studies of symptom checkers as part of electronic consultation systems in general practice, which again represents a slightly different setting from a general 'digital 111' service, and this should be kept in mind when interpreting the results.

A systematic review in such a topical area of research will require regular updating to keep track of new studies. For example, Verzantvoort et al.³⁸ published a study of self-triage using a smartphone app for out-of-hours primary care in the Netherlands shortly after our literature searches were completed. The app was rated highly for clarity and patient satisfaction. Sensitivity and specificity (using nurse telephone triage as reference standard) were 84% and 74% respectively, although diagnostic accuracy was only evaluated in a sample of participants (126/4456). Inclusion of this study would not have affected the main conclusions of our review.

Implications for service delivery and research

The implications of this systematic review for service delivery should be considered in the context that a decision has already been taken to introduce a 'digital 111' service and the service became available across England by December 2018. Achieving a high level of diagnostic accuracy will be key to the success of a 'digital 111' service. Failure to provide an accurate diagnosis may result in outcomes including patient dissatisfaction and unwillingness to use the service again; increased use of other urgent and emergency care services; and possible risks to patient safety (although the cautious approach characteristic of most existing systems may help to mitigate this).

The studies included in the review suggest a high level of uncertainty about the impact of 'digital 111' on the urgent care system and the wider healthcare system. Some of these uncertainties can be addressed by research and data collection but the health service may need to respond to short-term increases (or decreases) in demand and/or shifts from one part of the system to another. This may increase pressure on the system, at least in the short-term. In the longer-term, if usage of the 111 telephone service decreases as planned, there may be opportunities to reconfigure the workforce to support the integrated urgent care agenda.

Based on the areas of limited evidence identified by the review, priorities for research (in addition to ongoing collection of data to monitor usage and safety of the 'digital 111' service) include studies to compare the performance of different systems directly; rigorous economic evaluations based on real-world data; research to investigate the pathways followed by patients using the service; evaluation of systems designed for childhood illnesses; and

investigation of the possible role of behaviour change theory in the development and implementation of symptom checkers. Qualitative research to investigate perceptions of symptom checkers and barriers to their use by people who are less familiar with digital technology would also be of value.

Ethical approval

Ethical approval was not required for this work

Funding

This report presents independent research funded by the National Institute for Health Research (NIHR) Health Services & Delivery Research Programme (project number HSDR16/47/17). The funding programme approved the review protocol but had no role in the collection, analysis and interpretation of the data, the writing of this paper or the decision to submit the paper for publication. The views and opinions expressed are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

Data sharing

No new data have been created in the preparation of this report and therefore there is nothing available for access and further sharing. All queries should be submitted to the corresponding author.

References

- 1. NHS England. NHS 111 minimum data set 2018-19 2018 [Available from: <a href="https://www.england.nhs.uk/statistics/statistical-work-areas/nhs-111-minimum-data-set/statistical-work-areas-nhs-111-minimum-data-set-nhs-111-minimum-data-set-2018-19/accessed 29 October 2018.
- 2. NHS England. Five year forward view. Leeds: NHS England, 2014.
- 3. NHS England. Next steps on the NHS Five Year Forward View. Leeds: NHS England, 2017.
- 4. NHS England. Sustainability and transformation partnerships [Available from: https://www.england.nhs.uk/integratedcare/stps/ accessed March 25 2019.

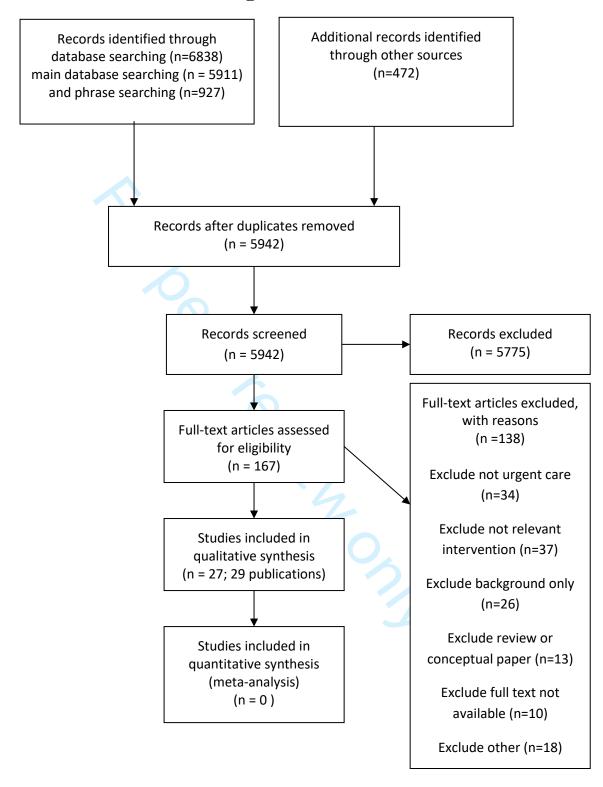
- 5. Turner J, O'Cathain A, Knowles E, et al. Impact of the urgent care telephone service NHS 111 pilot sites: a controlled before and after study. *BMJ Open* 2013;3(11):e003451. doi: 10.1136/bmjopen-2013-003451
- 6. Semigran HL, Linder JA, Gidengil C, et al. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015;351:h3480.
- 7. NHS England. NHS111 online evaluation. Leeds: NHS England, 2017.
- 8. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687. doi: 10.1136/bmj.g1687
- 9. Reitsma JB, Rutjes AWS, Whiting P, et al. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 100: The Cochrane Collaboration 2009.
- 10. Baxter S, Johnson M, Chambers D, et al. The effects of integrated care: a systematic review of UK and international evidence. *BMC Health Serv Res* 2018;18(1):350. doi: 10.1186/s12913-018-3161-3
- 11. Kellermann AL, Isakov AP, Parker R, et al. Web-based self-triage of influenza-like illness during the 2009 H1N1 influenza pandemic. *Annals of Emergency Medicine* 2010;56(3):288-94.e6.
- 12. Little P, Stuart B, Andreou P, et al. Primary care randomised controlled trial of a tailored interactive website for the self-management of respiratory infections (Internet Doctor). [Erratum appears in BMJ Open. 2017 Mar 21;7(3):e009769corr1; PMID: 28325861]. *BMJ Open* 2016;6(4):e009769.
- 13. Yardley L, Joseph J, Michie S, et al. Evaluation of a Web-based intervention providing tailored advice for self-management of minor respiratory symptoms: exploratory randomized controlled trial. *Journal of Medical Internet Research* 2010;12(4):e66.
- 14. Berry AC, Berry BB, Nakshabendi R, et al. Evaluation of Accuracy Between Online Symptom Checkers for Diagnosis of Gastrointestinal Symptoms from MKSAP Clinical Vignette Board Review Questions. *Gastroenterology* 2016;150(4):S849-S50. doi: 10.1016/s0016-5085(16)32869-4
- 15. Berry AC, Cash BD, Mulekar MS, et al. SYMPTOM CHECKERS VS. DOCTORS, THE ULTIMATE TEST: A PROSPECTIVE STUDY OF PATIENTS PRESENTING WITH ABDOMINAL PAIN. *Gastroenterology* 2017;152(5):S852-S53. doi: 10.1016/s0016-5085(17)32937-2
- 16. Middleton K, Butt M, Hammerla N, et al. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. London: Babylon Health, 2016.
- 17. Poote AE, French DP, Dale J, et al. A study of automated self-assessment in a primary care student health centre setting. *Journal of Telemedicine & Telecare* 2014;20(3):123-7.
- 18. Sole ML, Stuart PL, Deichen M. Web-based triage in a college health setting. *Journal of American College Health* 2006;54(5):289-94.
- 19. Nijland N, Cranen K, Boer H, et al. Patient use and compliance with medical advice delivered by a web-based triage system in primary care. *Journal of Telemedicine and Telecare* 2010;16(1):8-11. doi: 10.1258/jtt.2009.001004
- 20. Semigran HL, Levine DM, Nundy S, et al. Comparison of Physician and Computer Diagnostic Accuracy. *JAMA Intern Med* 2016;176(12):1860-61. doi: 10.1001/jamainternmed.2016.6001
- 21. Price RA, Fagbuyi D, Harris R, et al. Feasibility of web-based self-triage by parents of children with influenza-like illness: A cautionary tale. *JAMA Pediatrics* 2013;167(2):112-18.
- 22. Babylon Health. NHS111 powered by babylon: outcomes evaluation. London: Babylon Health, 2017.
- 23. Carter M, Fletcher E, Sansom A, et al. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. *BMJ Open* 2018;8(2):e018688.
- 24. Cowie J, Calveley E, Bowers G, et al. Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. *International Journal of Environmental Research and Public Health* 2018;15(896)
- 25. Madan A. WebGP: the Virtual general practice. London: Hurley Group, 2014.

- 26. Nijland N, van Gemert-Pijnen J, Boer H, et al. Increasing the use of e-consultation in primary care: Results of an online survey among non-users of e-consultation. *International Journal of Medical Informatics* 2009;78(10):688-703. doi: 10.1016/j.ijmedinf.2009.06.002
- 27. Backman AS, Lagerlund M, Svensson T, et al. Use of healthcare information and advice among non-urgent patients visiting emergency department or primary care. *Emergency Medicine Journal* 2012;29(12):1004-06.
- 28. Joury AU, Alshathri M, Alkhunaizi M, et al. Internet Websites for Chest Pain Symptoms Demonstrate Highly Variable Content and Quality. *Academic Emergency Medicine* 2016;23(10):1146-52.
- 29. Lanseng EJ, Andreassen TW. Electronic healthcare: a study of people's readiness and attitude toward performing self-diagnosis. *International Journal of Service Industry Management* 2007;18(3-4):394-417. doi: 10.1108/09564230710778155
- 30. Luger TM, Suls J. Online health information and intentions to seek healthcare. *Psychosomatic Medicine* 2011;73 (3):A59.
- 31. North F, Varkey P, Laing B, et al. Are e-health web users looking for different symptom information than callers to triage centers? *Telemedicine Journal & E-Health* 2011;17(1):19-24.
- 32. Berry AC, Berry NA, Wang B, et al. Symptom checkers versus doctors: A prospective, head-to-head comparison for GERD vs. Non-GERD Cough. *American Journal of Gastroenterology* 2017;112 (Supplement 1):S190. doi: http://dx.doi.org/10.1038/ajg.2017.299
- 33. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. *Journal of Medical Internet Research* 2014;16(1):e16.
- 34. Marco-Ruiz L, Bones E, de la Asuncion E, et al. Combining multivariate statistics and the thinkaloud protocol to assess Human-Computer Interaction barriers in symptom checkers. *Journal of Biomedical Informatics* 2017;74:104-22.
- 35. Nagykaldi Z, Calmbach W, Dealleaume L, et al. Facilitating patient self-management through telephony and web technologies in seasonal influenza. *Informatics in Primary Care* 2010;18(1):9-16.
- 36. Fraser HSF, Clamp S, Wilson CJ. Limitations of Study on Symptom Checkers. *JAMA Internal Medicine* 2017;177(5):740-41.
- 37. Mehrotra A, Semigran HL, Levine DM, et al. Limitations of Study on Symptom Checkers. *JAMA Internal Medicine* 2017;177(5):741-41.
- 38. Verzantvoort NCM, Teunis T, Verheij TJM, et al. Self-triage for acute primary care via a smartphone application: Practical, safe and efficient? *PLoS One* 2018;13(6):e0199284. doi: 10.1371/journal.pone.0199284 [published Online First: 2018/06/27]

Identification

Screening

PRISMA 2009 Flow Diagram



Appendix 1: Database search strategies

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present> Search Strategy:

- 1 (symptom checker or symptoms checker or symptom checkers or symptoms checkers).tw. (21)
- 2 ("self diagnosis" or "self referral" or "self triage" or "self assessment").tw. (10438)
- 3 TRIAGE/ (10017)
- 4 2 or 3 (20415)
- 5 (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone).tw. (658190)
- 6 4 and 5 (1568)
- 7 ("online diagnosis" or "web based triage" or "electronic triage" or etriage).tw. (42)
- 8 1 or 6 or 7 (1608)

Embase

1 (symptom checker or symptoms checker or symptom checkers or symptoms checkers).tw.

- 2 ("self diagnosis" or "self referral" or "self triage" or "self assessment").tw.
- 3 emergency health service/
- 4 2 or 3
- 5 (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone).tw.
- 6 4 and 5
- 7 ("online diagnosis" or "web based triage" or "electronic triage" or etriage).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
- 8 1 or 6 or 7

Cochrane Library

#1 symptom checker or symptoms checker or symptoms checkers:ti,ab,kw (Word variations have been searched)

#2 "self diagnosis" or "self referral" or "self triage" or "self assessment":ti,ab,kw (Word variations have been searched)

#3 MeSH descriptor: [Triage] explode all trees

#4 #2 or #3

#5 online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone:ti,ab,kw (Word variations have been searched)

#6 #4 and #5

#7 "online diagnosis" or "web based triage" or "electronic triage" or etriage:ti,ab,kw (Word variations have been searched)

#8 #1 or #6 or #7

CINAHL

S8 (S1 OR S6 OR S7)

S7 TI ("online diagnosis" or "web based triage" or "electronic triage" or etriage) OR AB ("online diagnosis" or "web based triage" or "electronic triage" or etriage) S6 S4 AND S5

S5 TI (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone) OR AB (online or on-line or web or electronic or automated or internet or digital or app or mobile or smartphone)

S4 (S2 OR S3)

S3 (MH "Triage")

S2 TI ("self diagnosis" or "self referral" or "self triage" or "self assessment") OR AB ("self diagnosis" or "self referral" or "self triage" or "self assessment")

S1 TI (symptom checker or symptoms checker or symptom checkers or symptoms checkers) OR AB (symptom checker or symptoms checker or symptoms checkers)

ACM digital library

WOS

#8 #7 OR #6 OR #1

#7 TS=("online diagnosis" OR "web based triage" OR "electronic triage" OR etriage)

#6 #5 AND #4

#5 TS=(online OR on-line OR web OR electronic OR automated OR internet OR digital OR app OR mobile OR smartphone)

#4 #3 OR #2

#3 TS=triage

#2 TS=("self diagnosis" or "self referral" or "self triage" or "self assessment")

#1 (symptom checker or symptoms checker or symptom checkers or symptoms checkers)

HMIC

- 1 (symptom checker OR symptoms checker OR symptom checkers OR symptoms checkers).ti,ab
- 2 ("self diagnosis" OR "self referral" OR "self triage" OR "self assessment").ti,ab
- 3 TRIAGE/
- 4 (2 OR 3)
- 5 (online OR on-line OR web OR electronic OR automated OR internet OR digital OR app OR mobile OR smartphone).ti,ab
- 6 (4 AND 5)
- 7 ("online diagnosis" OR "web based triage" OR "electronic triage" OR etriage).ti,ab 8 (1 OR 6 OR 7)

Appendix 2: Risk of bias tables

Risk of bias results for randomised trials

Short Title	Reference	Selection and performance bias	Detection and attrition biass	Reporting and other bias
Little (2016)	Study ID	Random sequence generation	Blinding of outcome $\vec{\varphi}$	Selective reporting
	• Reference	• Low risk	assessment* • Low risk □	• Unclear
	Little 2016 ¹²		• Low risk	
			Blinded assessment of primary	
		Allocation concealment	care records	Anything else, ideally
		• Low risk	care records d from htt	prespecified
		0.6	http	• Low risk
			Incomplete outcome data	
		Blinding of participants and	• Low risk	
		personnel*	- pen	
		• Unclear	ı.bm	
		10	y, co	
Yardley (2010)	Study ID	Random sequence generation	Blinding of outcome	Selective reporting
	• Reference	• Low risk	assessment* □	• Unclear
	<i>Yardley</i> 2010 ¹³		• Low risk Blinding of outcome assessment* • Unclear On April 9, 2024	
		Allocation concealment	.022	Anything else, ideally
		• Low risk	Incomplete outcome data	prespecified
			• Low risk	• Low risk
			st.	
		Blinding of participants and	Pro	
		personnel*	tect	
		• Low risk	ed t	
			• Low risk guest. Protected by capyright	
	1	1	фу	
			righ	

Risk of bias results for cohort/cross-sectional studies

Reference	Questions 1-4	Questions 5-7	Questions 8-10 ভূ
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Unclear
Backman A-S	The aims refer to "non-urgent" but the information is	• No	"Health care information use in the past"
et al. 2012 ³⁰	sought prior to visiting ED.		nloa
			3d ee
	$\mathcal{O}_{\mathcal{O}}$	6. Did the study	9. Were outcome assessors blinded?
	2. Was the study population clearly specified and	examine exposure	• Not applicable
	defined?	levels?	nttp
	• Yes	• Yes	//bn
	4	Health advice seeking	10. Were confounders adjusted for?
			• Yes 💆
	3. Was the participation rate at least 50%?		To some extent: participant and physician attributes
	• Yes	7. Were exposure	assessed for influence on the results.
	79%	measures clearly	0
		defined?	on April 9, 2024 by guest
		• Unclear	ori :
	4. Were all the subjects selected or recruited from	Measures are vague, e.g.	9, 2
	the same or similar populations?	"previous use" of	024
	• Yes	information Also,	ьу
	Primary care and ED attendees	discriminating between	gue
		types of information	st. F
			⁵ rot
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes Ö
Carter 2018 ²⁶		• No	Attitudes and experiences of practice staff and

			patients on web P.
	2 Was the study nanulation sleavly specified and	6 Did the study	∑ Q
	2. Was the study population clearly specified and defined?	6. Did the study examine exposure	9. Were outcome assessors blinded?
	• Yes	levels?	• Not applicable
	GPs, practice staff and their patients at 6 practices in	• Not applicable	Not applicable
	Devon	1 voi applicable	2018
	Beron		10. Were confognders adjusted for?
		7. Were exposure	• Not applicable
	3. Was the participation rate at least 50%?	measures clearly	loac
	• No	defined?	ēd.
	Postal survey only had response rate of 35.1% but also	 Not applicable 	rom
	GPs judgement of webGP requests and 5GPs and 5		httt
	administrators were interviewed.	4-)://b
			m _{jo}
			De
	4. Were all the subjects selected or recruited from		bmj.
	the same or similar populations? • Yes		.00
	GPs, practice staff and their patients at 6 practices in		0,
	Devon		> P
	Devon		• Not applicable loaded from http://bmjopen.bmj.com/ on April 9,
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes $\frac{4}{5}$
Cowie 2018 ²⁷		• No	• Yes by gue
	2 W		s :
	2. Was the study population clearly specified and defined?	6. Did the study	9. Were outcome assessors blinded? • No
	• Yes	examine exposure	• No tected
	103	levels?	d by
		• No	10. Were confounders adjusted for?
		• No	10. Were confo@nders adjusted for?

		BMJ Open	6/bmjopen-2018-027743 on 1 August 2019. Dow
	2 W41	1	20 8- 1. v
	3. Was the participation rate at least 50%?No		• Yes 27743
	No for patient surveys	7. Were exposure measures clearly defined?	on 1 Augu
	4. Were all the subjects selected or recruited from the same or similar populations?	Not applicable	st 2019.
	• Yes		Down
• Reference	1. Was the research question clearly stated? • Yes	5. Was a sample size justification provided?	8. Were outcome measures clearly defined? • Yes
Joury et al. 2016 US ³¹		• No	Scores used for geadability, popularity, content and quality
	2. Was the study population clearly specified and defined?	6. Did the study	p://bmj
	Not applicable	examine exposure levels? • Not applicable	9. Were outcome assessors blinded? • Not applicable
	3. Was the participation rate at least 50%?Not applicable		10. Were confounders adjusted for?
	4. Were all the subjects selected or recruited from	7. Were exposure measures clearly defined?	• Unclear of the second of the
	the same or similar populations? • Not applicable	Not applicable	• Unclear April 9, 2024 by guest
• Reference	Was the research question clearly stated? Unclear	5. Was a sample size justification provided?	8. Were outcome measures clearly defined? • Not applicables
Kellermann 2010 11	Cheleat	• Not applicable	TNOT applications of the control of
	2. Was the study population clearly specified and		9. Were outcone assessors blinded?

	defined?	6. Did the study	• Not applicable
	• Unclear	examine exposure	43 .
	Patients with influenza-like illness in US that accessed	levels?	on a
	one of 2 websites http://www.flu.gov and www.H1N2	 Not applicable 	10. Were confounders adjusted for?
	ResponseCenter.com		• Not applicable
			↑
		7. Were exposure	019
	3. Was the participation rate at least 50%?	measures clearly	Do
	• Not applicable	defined?	Na N
		Not applicable	oac
	'	**	e d
	4. Were all the subjects selected or recruited from		2019. Downloaded from http://bmjopen.bm
	the same or similar populations?		h t
	• Unclear		ф://
	Only counted web hits, no demographic data available		<u>b</u>
	on patients. No data on usage of algorithm by	81.	op _e
	clinicians or call centers.		n. b
		10,	ற்.c
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Yes Use of TRI 9. Were outcome assessors blinded?
Lanseng &		• No	Use of TRI
Andreassen			9
2007	2. Was the study population clearly specified and		20:
Norway ³²	defined?	6. Did the study	9. Were outcome assessors blinded?
1101 way	• Yes	examine exposure	
	165	levels?	• No guess
		• No	.÷ □
	3. Was the participation rate at least 50%?	Readiness	10. Were conformders adjusted for?
	• Unclear	110000000	• Unclear
	One ion		
		7. Were exposure	by copyric
		7. Were exposure	l 꽃

BMJ Open

	4. Were all the subjects selected or recruited from	measures clearly	277
	the same or similar populations?	defined?	027743 on
	• Yes	Not applicable	0n 1
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	-	Y
	• Yes	justification provided? • No	• Yes 201
Luger et al. 2014 ²³		• NO	9. Do
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable
	• Yes	examine exposure	ed fr
	\sim	levels?	rom
	0,6	• No	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?		• Unclear
	• Unclear		m _{jo}
		7. Were exposure	• Unclear //bmjopen.bmj.com/ on Apri
		measures clearly	ı.bm
	4. Were all the subjects selected or recruited from	defined?) j. CC
	the same or similar populations?	 Not applicable) m
	• Yes		on ,
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
 Reference 	• Yes	justification provided?	• Not applicable
Marco-Ruiz et		• No	y by
al. 2017 Norway ²⁴	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
rvoi way	defined?	6. Did the study	
	• No	examine exposure	• Not applicable of the control of t
	110	levels?	ctec
		• No	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?	110	
	or reason pur nerpunon rate at least 50 /0.		• Unclear $\frac{6}{5}$

		BMJ Open	6/bmjopen-2018-0
	L. V.	T	
	• Yes		774
	53%	7. Were exposure	27743 on 1 August 2019.
		measures clearly	1,1
	4. Were all the subjects selected or recruited from	defined?	, lbn,
	the same or similar populations?	• Not applicable	ust 2
	• Unclear	Tvot applicable	2019
	Onercui		9. D
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes
Nagykaldi		• Not applicable	Web hits on customised practice website influenza
2010 25	$\mathcal{N}_{\mathcal{O}}$	The state of the s	self-managemera webpages. Downloads of self-
	2. Was the study population clearly specified and		management influeza toolkit. Completion of Iflueza
	defined?	6. Did the study	self-triage modite sessions. Volume of calls to
	• Yes	examine exposure	telephone hotlings. Qualitative feedback from patient
	Study population was patients from 12 primary care	levels?	on statisfaction with and utility of self-management
	practices in US.	 Not applicable 	websites and telephone hotline. Qualitative feedback
		101	from clinicians ground their involvement and their
		\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	perceptionsof patient self-management techniques.
	3. Was the participation rate at least 50%?	7. Were exposure	on /
	Not applicable	measures clearly	γpril
		defined?	9. Were outcome assessors blinded?
		 Not applicable 	• Not applicable
	4. Were all the subjects selected or recruited from		t by
	the same or similar populations?		gue
	• Yes		10. Were confounders adjusted for?
	All participants were patients from 12 primary care		• Not applicable of 2
	practices that accessed customised practice website or		ecte
	telephone helpline		ਲੂ ਰ
			<u> </u>
			ected by copyright

	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
• Reference	• Yes	justification provided?	• Yes $\frac{7}{4}$
Nijland 2009 ²⁹		• No	on
			1 A
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• No $\frac{\aleph}{8}$
	• Yes	examine exposure	• No 2019
		levels?	. Dc
		Not applicable	10. Were confonders adjusted for?
	3. Was the participation rate at least 50%?		• Yes og
	• Unclear		Methods not ver clearly reported but appears to be
	$\mathcal{N}_{\mathcal{O}}$	7. Were exposure	multiple regresson
	CO.	measures clearly	htt
	4. Were all the subjects selected or recruited from	defined?	p://k
	the same or similar populations?	Not applicable	om _j .
	• Yes	01	nttp://bmjoper
			n.bn
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
 Reference 	• Yes	justification provided?	• Yes
Nijland 2016 ¹⁹		• No	on A
			V pril
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• No 80
	• Yes	examine exposure	• No 2024 by
		levels?	Q
		• Not applicable	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?		• Unclear
	• No		• Unclear Protected by copyrig
	Low participation rate in survey relative to users of	7. Were exposure	te d
	triage system (though unclear how many were invited	measures clearly	by c
	to participate)	defined?) Op:

6/bmjopen-2018-

		T ==	
		Not applicable	2774
	4. Were all the subjects selected or recruited from		<u>ω</u> ο
	•		1
	the same or similar populations?		Aug
	• Yes		-027743 on 1 August 2
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Yes 💆
North et. al. 2011 ³⁴		Not applicable	• Yes Download
	2. Was the study population clearly specified and		9. Were outcome assessors blinded?
	defined?	6. Did the study	• Not applicable
	• Yes	examine exposure	htt
		levels?	p://k
		• Yes	10. Were confounders adjusted for?
	3. Was the participation rate at least 50%?	Self-exposure	• Unclear
	Not applicable	10.	Some discussion of potential confounders.
		7. Were exposure	com/ on April 9,
	4. Were all the subjects selected or recruited from	measures clearly	On .
	the same or similar populations?	defined?	Apri
	Not applicable	Not applicable	202
	1. Was the research question clearly stated?	5. Was a sample size	8. Were outcome measures clearly defined?
Reference	• Yes	justification provided?	• Not applicable
Sole 2006 ¹⁸	"The primary purpose of this study was to identify and	• No	est.
	describe the demographic profile of students who used		Pro
	the newly implemented Web-based triage system. A		9. Were outcone assessors blinded?
	secondary purpose was to compare Web-based triage	6. Did the study	• Not applicable
	diagnoses to the diagnoses made in clinic for a subset	examine exposure	y o
			copyright

of students who requested appointments"	levels?	10. Were confounders adjusted for?
	• Yes	• Not applicable
		on 1
2. Was the study population clearly specified and		Au
defined?	7. Were exposure	gus
• Yes	measures clearly	August 2019
Students who used the web based triage over a four	defined?	919.
month implementation period (1290 students). Then of	• Yes	Do
those students, those who requested an appointment via		w _n
email (143 students), then of those 59 who attended the		oad
health centre after requesting an email appointment.		ed T
3. Was the participation rate at least 50%?		Downloaded from http://bmjopen.bmj.com
• Not applicable		/bm
Two applicable	O_1	Jope Personal Property of the
		n.b
4. Were all the subjects selected or recruited from	10,	m _{j.c}
the same or similar populations?		, which is a second of the sec
• Yes		on
		Apri.

Risk of bias results for diagnostic studies

Reference	Questions 1 to 4	Questions 5 to 8	Questions 9 to 11 g
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical Information?
•	• No	verification	• Yes g
Reference	Study participants were all patients registered at a student	avoided?	ted
Poote	health centre in England attending with new acute	• Not applicable?	by a
			10. Were uninterpretable results reported?

2014 17	symptoms. If the self-assessment triage system was only		• Not applicable 777 43
	for students to be representative the study population		743
	would have needed to include range of student health	6. Was the	on .
	centres in different areas. If the system was for any UK	reference standard	11. Were withdrawals from the study explained?
	general practices the study population would have needed	independent of the	• Yes G
	to include patients of all ages, ethnicity, gender etc from a	index test?	st 20
	range GP practices in different areas.	• Unclear	019.
		Patients took the	Do
	· O.	assessment from self-	wnl
	2. Acceptable reference standard?	triage through to	oad
	• Yes	their GP	ed f
		consultation.	ro m
			h _{tt}
	3. Acceptable delay between tests?		p://k
	• Yes	7. Index test results	omjo
		blinded?	р er
		• No).bn
	4. Partial verification avoided?	Patients took the	ن. در
	• Yes	assessment from self-	/mv
	All patients that completed self-triage also had a GP	triage through to	On S
	consultation where the GP rated the urgency of their	their GP	Apri.
	consultation.	consultation.	J.9,
			202
			4 by
		8. Reference) gu
		standard results	gust 2019. Downloaded from http://bmjopen.bmj.com/ on April 9, 2024 by guest. Protected
		blinded?	Pro
		• Yes	ytec
			_
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?

•	• No	verification	• Yes
			• Yes 27743
Reference	SORT was only trialled in 2 Emergency Departments in	avoided?	<u>ω</u> <u>ο</u>
Price 2013	US, a larger range would be needed for a representative	• Not applicable?	
20	spectrum. Also, patients were from ED not home so		10. Were uninterpretable results reported?
	potentially sicker patients in the sample.		• Not applicable
		6. Was the	• Not applicable
		reference standard	19.
	2. Acceptable reference standard?	independent of the	11. Were withdraws from the study explained?
	• Yes	index test?	• No No
	Sensitivity of SORT for kids algorithm in identifying the	• Yes	oa d.
	need for ED care was based on an explicit gold standard:		ed f
	documented evidence that the child received 1 or more of		rom
	5 ED-specific interventions.	7. Index test results	n htt
		blinded?	Inloaded from http://bmjopen.bmj.com/ on April 9
		• Yes	<u> </u>
	3. Acceptable delay between tests?	Θ_{1}	оре
	• Yes		n.b
		8. Reference	ற <u>்</u> .
		standard results	Ö
	4. Partial verification avoided?	blinded?	on
	• Yes	• Yes	Ap
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?
•	• Unclear	verification	• Yes $\overset{4}{\sigma}$
Reference	There were 45 standardised patient vignettes which were	avoided?	This is the clinical information that would be supplied by
Semigran	divided into three levels of triage urgency and included	• Not applicable?	the patient which may or may not differ from the
2015 4	more and less common conditions. It is not clear how		information given by \$\frac{1}{2}\$he vignette.
	closely this replicates the spectrum of conditions that		[#548 Semigran 2015, pdf] Page 8: ion of the true clinical
	people use symptom checkers for.	6. Was the	accuracy of sympton checkers.33 Some standardized
	r · · · · · · · · · · · · · · · · · · ·	reference standard	patient vignettes constained specifc clinical language (fo
		independent of the	op Tri
			.₩

	2. Acceptable reference standard?	index test?	example, mouth ulcers, tonsils with exudate), and actual
	• Yes	• Yes	patients with the sante condition might struggle with the
	[#548 Semigran 2015.pdf] Page 2: The source for each		words to use to describe their symptoms or use diferent
	vignette also provided the associated correct diagnosis.		terms. Therefore, ou analysis represents an indirect
		7. Index test results	assess- ment of how well symptom checkers would perform
		blinded?	with actual patients 8
	3. Acceptable delay between tests?	• Yes	99.
	Not applicable		Dog
			10. Were uninterpretable results reported?
		8. Reference	• Yes
	4. Partial verification avoided?	standard results	[#548 Semigran 201 spdf] Page 3: ns for diagnosis and
	Not applicable	blinded?	triage was high (CoRen's κ 0.90). In some cases we could
	(0,4	• Yes	not evaluate a vignette because some symptom checkers
		6	focus only on children or on adults or the symptom checker
			did not list or ask for the key symp- tom in the vignette. To
			avoid penalizing the se symptom checkers, we referred to
			standardized patient gignettes that successfully yielded an
		'01.	output as "standardized patient evaluations."
			m,
			n n n n n n n n n n n n n n n n n n n
			11. Were withdraw ব্লাs from the study explained?
			• Not applicable 9
Study ID	1. Representative spectrum?	5. Differential	9. Relevant clinical information?
•	• Unclear	verification	• Yes ©
Reference	There were 45 standardised patient vignettes which were	avoided?	The physicians and the symptom checkers used the same
Reference		• Not applicable?	vignettes
Semigran	divided into three levels of triage urgency and included	• Not applicable:	vigneties
Semigran	divided into three levels of triage urgency and included more and less common conditions. It is not clear how	- Not applicable:	vignenes To
		6. Was the	10. Were uninterpretable results reported?

BMJ Open

6/bmjopen-2018

j.com/ on April 9, 2024 by guest. Protected by copyright.

		Ċ
	reference standard	• Not applicable
	independent of the	43
2. Acceptable reference standard?	index test?	on `
• Yes	 Not applicable 	11. Were withdrawals from the study explained?
		• No gr
		It is unclear why the physicians chose to solve the specific
3. Acceptable delay between tests?	7. Index test results	vignettes $\vec{\omega}$
Not applicable	blinded?	Do
	• Yes	w _n !
		oad
4. Partial verification avoided?		ided f
• No	8. Reference	from
There was a total of 234 physicians involved in the study	standard results	
and of the 45 vignettes, each was solved by at least 20	blinded?	http://bmjop
physicians but it is not clear why they chose the specific	• Yes	jmj
vignettes to solve.	CV:	open.b
		i i



PRISMA 2009 Checklist

		01	
Section/topic	#	Checklist item 277 43	Reported on page #
TITLE		0 2	
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT	· · · · · · · · · · · · · · · · · · ·	gust :	
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION		h loac	
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, in reference, comparisons, outcomes, and study design (PICOS).	5
METHODS		p://br	
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplion tensor and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I²) for each meta-ranalysis - http://bmjopen.bmj.com/site/about/guidelines.xhtml	N/A

BMJ Open



PRISMA 2009 Checklist

1		Page 1 of 2	
Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7
RESULTS	•		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reach stage, ideally with a flow diagram.	9
7 Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOs, follow-up period) and provide the citations.	10-16
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Appendix 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summare data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	17-22
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	22-23
DISCUSSION		70 71:	
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	26-27
33 Limitations 34	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	28
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implication of the research.	28-29
FUNDING		· ec	
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	29

43 doi:10.1371/journal.pmed1000097

The PRISMA for Abstracts Checklist

	BMJ Open Bjope			
BMJ Open Spen 2018-2018-2018 The PRISMA for Abstracts Checklist				
TITLE	CHECKLIST ITEM 3 9 1	REPORTED ON PAGE #		
1. Title:	Identify the report as a systematic review, meta-analysis, or both. Agust 2019	1 (also in 'Design')		
BACKGROUND	2019			
2. Objectives:	The research question including components such as participants, interventions, comparators, and outcomes.	3 (Objectives)		
METHODS	ŭ d ed			
3. Eligibility criteria:	Study and report characteristics used as criteria for inclusion.	3 (Eligibility criteria)		
4. Information sources:	Key databases searched and search dates.	3 (Data sources)		
5. Risk of bias:	Study and report characteristics used as criteria for inclusion. Key databases searched and search dates. Methods of assessing risk of bias.	3 (DE and synthesis)		
RESULTS	om/			
6. Included studies:	Number and type of included studies and participants and relevant characteristics of stydies.	3 (Results)		
7. Synthesis of results:	Results for main outcomes (benefits and harms), preferably indicating the number of studies and participants for each. If meta-analysis was done, include summary measures and confidence intervals.	3 (Results)		
8. Description of the effect:	Direction of the effect (i.e. which group is favoured) and size of the effect in terms meaningful to clinicians and patients.	3 (Results)		
DISCUSSION	st. Pr			
9. Strengths and Limitations of evidence:	Brief summary of strengths and limitations of evidence (e.g. inconsistency, imprecisions indirectness, or risk of bias, other supporting or conflicting evidence)	3 (Results)		
10. Interpretation:	General interpretation of the results and important implications	3 (Conclusions)		

		Q	
OTHER		18-02	
11. Funding:	Primary source of funding for the review.	77743	In text
12. Registration:	Registration number and registry name.	on 1	3

August 2019. Downloaded from http://bmjopen.bmj.com/ on April 9, 2024 by guest. Protected by copyright