Supplementary material BMJ Health Care Inform

Table 2		Country	Participants	Method	Most importent findings	Strengths in the citing	Weakness in the citing	Citing conclusions
Safety of patient-facing digital symptom heckers. Fraser H, Colera E, Wong D. ancet. 2018 Nov 24;392(10161):2263- 2264. doi: 10.1016/S0140- UZBATIR/20419-8. Emit-2018 Nov-6. Patier Review: the Babbylon	2018	UK / Australia	None	Expert opinion / correspondence	Symptom checkers have great potential to improve diagracia, quality of care, and health system of performance worldwish. However, systems far a proxyl designed for this rightoral critical evaluation can put patients at trisk and filterly increase the load on health systems. Evaluation guidelines specific to gryptom checkers have three benefits. First, they would provide system creaters with a fixed to circline, alread of firms, on which they will be assessed. Second, they would allow external cheeners to issuess the comprehensives and quality of evaluation, discouraging system creaters for minding the importance of their results. First, they would facilitate policy makers in determining a minimum level of anxietacs results and first valuations.	Academic writing using references	Evidence considered on level of expert opinion Evidence considered on level of expert opinion	It is not possible to determine how well the Babylon Diagn and Triage System would perform on a broader randomiz- set of cases or with data entered by patients instead of doctors. Babylon's study does not offer convincing evide that the Babylon Diagnostic and Triage System can perfort better than doctors in any resistis situation, and there is a possibility that thingh perform significantly worse. Evalua- of soundom checkers should follow a multistane process. The reviewed study are considered a very preliminary and
hatbot(Internet). The Guide to Health formatics 3rd Edition. 2018 (pixed 19 May 29). Available from: ps://coiera.com/2018/06/29/paper- view-the-baby/on-chatbot/.		(from reference list)		expert opinion	vigentes exploring other dispositions would likely have resulted in a much poorer performance. This sets displaying on what it she fower not have full seame for the result of more first persentation of information was in the GCEC format, which is artificial and not now patients might present. So there was no real settings in the GCEC format, which is artificial and not now patients might present. So there was no real settings and consultation and failestings like that would be needed to manage an entire of patient presentation. A feature of the consultation model would have been to draw a random subset of cases and present them to both GPs and Babylon.			artificial test of a Bayesian reasoner on cases for which it already been trained. In machine harming the would be roughly equivalent to in-sample reporting of performance the data used to develop he algorithm. Good practice is report out of sample performance on previously unseen cases. The results are confounded by artificial conditions and us few and non-independent assessors. There is lack of ola the way data are analyzed and there are numerous risks.
azzaki S, Baker A, Perov Y, iddleton K, Baxter J, Mullarkey D, et A comparative study of artificial elligence and human doctors for the urpose of triage and diagnosis. Xv1100610698 Cs Stat [Internet]. 118 Jun 27 (cited 2019 May 15):		UK/U.S. (From reference list)		Accurate outcome analysis	A prospective validation study of the accuracy and safety of an All powered Triage and Disgrostic System was performed unique an experimental partiagin designed to simulate resides consultances. It was found that the Bashyion Triage and Diagnostic System was able to identify the condition modeled by a landical signed with accuracy comparable to human doctors in times of proceion and restall, it was also found that the triage advices incommenced by the Bashyion Triage and Diagnostic System was safety also found that the triage advices incommenced by the Bashyion Triage and Diagnostic System was safety subject, with only minimal reduction is approximates, so for event, the All Systems was able to safely subject, with only minimal reduction is approximates, in other words, the All Systems was able to safely subject, with only minimal reduction is approximates.	validation study using an experimental paradigm designed to simulate realistic consultations.	Evaluated only clinical cases that were based on a single underlying condition. Similar cases was used to train the algorithm.	Artificial intelligence powered symptom checkers have it potential to provide diagnostic and triage advice with a la accuracy and safety approaching that of human doctors Such systems may hold the promise of reduced costs improved access to healthcare voorfulde, but realising requires greater levels of confidence from the medical community and the wider public. Rey to this confidence to confidence
rro. Facilitating User Symptom Checking a Personalised Chathot-Oriented dogue System cosh S, Bhatis A, ad Health Technol Inform. 18:252:51-56	2018	Australia	30 test-case vignettes. Evaluation of chatbot triage accuracy	Accurate outcome analysis	Sympton extraction from over proj. involves employing algorithms, used to recognite production of the project o	Thorough and apparently transparent description of chatbot function and development.	The study evaluation is based on only 30 clinical scenarios vignets, and are not evaluated by real patients in a natural patients in gapes, Should be evaluated. On one of a much larger number real patient department of the patient of the program.	An automated medical conversational pattern powered basering algorithm that provides personalized assessing based on symptoms is described. The boffs symptom recognition and condition assessment performance cou- greatly improved by adding support for more medical features, such as location, adverse events, and medical entities.
yand Dr. Google: the evidence on nsumer-facing digital tools for gnossis. lenson ML, Baldwin JL, Zipperer L, igh H. ugnosis (Berl). 2018 Sep 25	2018	U.S. (from reference list)	None	Systematic literature review	Evaluations were highly variable in methodology and conclusions, with about half describing approximations and assuming actual professors. An experiment of the professor is useful and effectiveness, ashough the susthlesses of the evidence was limited by a through the concerning country and experiment of the professor is useful and by a through a page are regoligly everying, optom and stratefunded evaluations are secretable in helm decisions by diricisors, patients, polymorates and other stateholders. The use of artificial intelligence (Al) – has strateful and the professor in the professor in the professor is consistent (Al) and a professor in the professor in the professor in the professor is a professor in experiment of the professor in the professor is consistent (Al) and a professor in the described scale and the decisions by directive and the professor in	The study does not manage to identify high quality studies	A systemic and transparent approach	Overall, the current evidence base on DTC, interactive diagnostic appe is spanse in scope, uneven in the inform provided, and inconculsave with respect to safety and effectiveness, with no studies of clinical risks and benef involving real-voicd consumer use. Athough some studie axamined rigorously determined the sensitivity and spe of app-generated diagnoses, methodologies varied considerably. Given that DTC diagnostic apps are rapic excitors.
valuation of a Diptal Consultation and, all Grare Advice for in Primary Care: Multi-Methods Study, owie J. Caveley E, Bowers G, owers J 1.1 Environ Res Public Health. 2018 ay 2:15(6)	2018	UK	patient population, 80546 patients	Multi-method: 5 month observational study. Quantitative data, qualitative data and health economic analysis	scurace numbit be asmitisative immoved with Al. Researchers should validate and standardise. This soldy provide an inspirit to the experience and challenges of imprimenting destroots cell- imasagement and convolutation both in a primary can setting across diverse GP settings. This also happinglished sense of immodities content with any selfect the witnesses to adopt or relate. Ghostul or as consultation, that verificated was not decreased and, in general, patients that used of consult were felt to benefit own the service. This soldy also sought to identify appeals control the imprimentation of closus that were seen to support or Inside the institution and integration of the system. It was written that facts with foliabled the imprimentation of closus the representation where the control of the provident that facts with foliabled the imprimentation of closus the representation provident that facts with foliabled the presentation of closus the representation provident that the control of the provident that facts with the control of the control of the control of the control of the provident that the control of the control of the control of the control of the provident that the control of the control of the control of the provident that the control of the control of the control of control of the control of the control of the control of control of the control of the control of control of the control of the control of control of cont	The evaluation used a multi-methods design incorporating both quantitative data (log data from eConsult use, patient survey data) and qualitative data (interviews with GP seaff and free lax design existing the consultation of digital systems and has been used throughout the literature	The level of effectiveness was difficult to assess given born unchers, only aminoty of patients (6.5%) completed evaluation. The evaluation was only conducted over a fermitip priord, it was not possible to assess the long-term sustainability of eConsuit processes adopted.	section. Tonce froueth and protoco an evaluation are set in Through the evaluation of deforust across Socioland, the has been hown to have the potential to spriftcastly in the provision of privacy can service as after the pro- tocol protocol or privacy, can service as after settle the provision of privacy can service as after settle identified key recommendations for successful implement of a digital advice and consultation tool.
asbilly, acceptability and colored abstractive to activeness of an origine alternative to active face consultation in general, active a mixed-methods study of those in section practices; rater M, Fetcher E, Sansom A, arren FC, Campbell JL. AJ Open. 2018 Feb	2018		provided consultations data; 20 GPs completed case reports (regarding 61 e- consults); 81 patients completed	Multi method: Mixed-methods evaluation, including data extraction from practice databases, general practitioner (GP) completion of case	practice; by volumes of accorable leading to some furtishes and nability to assess effectiveness, and WeeD [*] Graphic during the evaluation was smit, showing not discernible impact of practice verificated. The completeness of cross sectional data on consultation workload varied between practices. CPS judged 1910 (1794) and WedD [*] Grequests to explore a face-the face or telephone consultation, introducing vettod [*] appeared to be associated with inition in responsibility and workload between practices stiff and consultant was consultantly urguest and most leading with the properties of the properties. WebG [*] appeared to mostly upong and most leading which to be employed into face be face reproducted. WebG [*] appeared broadly acceptable to patients regarding meliness and qualify lexiperience of care provided. Similar profession was presented by all respondents. Both groups appeared equally familiar with other practice orders services; consultants were somewhalm nore likely to be used than. From constructured staff interviews, tappeared that, while targety acceptable within practice, introducing e- terostructured staff interviews. Lappeared that, while targety acceptable within practice, introducing e- terostructured staff interviews. Lappeared that, while targety acceptable within practice, introducing e- terostructured and interviews. Lappeared that, while targety acceptable within practice, introducing e-	a range of list sizes and deprivation The two patient surveys provided important information about patient views and enabled comparisons between groups.	The evaluation was limited to a small group of practices participating fine webCP plot implementation in jact one CCG area, with a proceduration with British population. Uptake of webCP varied, with very few patients in most practices making use of the new system the small number of interviews does not allow the authors to assume that saturation of emergent themes has been achieved.	This small study has demonstrated that, by applying a of complementary methods, it is possible to learn value of complementary methods, all to possible to learn value lessons about the impact of imnovative approaches to managing workshold in general practice. Online constitution provide an effective and convenient alternative for some groups of patients (such as those in employment) with urgent problems, who do not require face-to-face contains the practice. It may also offer an acceptable means by patients can consult their clinicians about sensitive topic which may offerwise remails unarticulated.
e of a primary care online issultation system, by whom, when I why: evaluation of a pilot, tervational study in 36 general edices in South West England, wards HB, Marques E, Hollingworth Horwood J, Farr M, Bernard E, isbury C, Northstone K.	2017	UK	296828 patients	renords, natient Multi-method, 15- month observational study, quantitative, qualitative data and health economic analysis	crossable, but or obtential for a disverse instructions, with one service conception, assistent. The mentionalities for Preferren in one support upon that of other conception systems could be eveloped to improve from the property of the	Multi-method study including a large number of patients and an independent evaluation of a commercial primary care online consultation system. Use of this methodology is considered appropriate for evaluation of digital systems and has been used throughout the literature	Website analytics data showed that the use of actual e-correlutations was very low. Over the 38 participating practices, there was a mean of 2.00 e- consultations per 1000 patients per morith. Another weakness is the observational nature of the study design that limits the authors to a descriptive evaluation.	Use of e-consultations was very low, particularly at wee Unless this can be improved, the impact of e-consultat systems on reducing staff workload and improving wait times is likely to be negligible. It is also possible that use consultations may be associated with increased costs i workload in primary care.
periences with online consultation, stems in primary care: case study of e anty adopter site, sey M, Shaw S, Swinglehurst D. J Gen Pract. 2017	2017	UK	interviews with a sample of seven stakeholders (three development/op	Mixed method case study. Qualitative and a sample of six documents (including a pilot report, training presentations, and	mediators of the success (or not) of technology implementation. Uptake by patients remixed loxe, there was little evidence that efficiency gains were related. The research suggests that — all seats for GPs and administrators — the assumed potential of Tele-Doc for increased efficiency is difficult to achieve. Developers and childran invested considerable them and resource for Tele-Doc's intelligence Developers and childran invested considerable than a difficult to a fine the Tele-Doc's intelligence may reflect patients' inelatance or inability to commit to a specific condition at the submitted for the consultation, or a poor to behave the near not of plantiff profession and the algorithm (social principles) and the profession of the committee of plantiff profession and the algorithm (social principles) and the profession of the committee of plantiff profession and the algorithm (social principles) and the profession of the committee of the profession of the plantiff profession and the algorithm (social principles) and the profession of the profession of the plantiff profession and the algorithm (social principles) and the profession of the profession of the profession of the plantiff profession and the algorithm (social principles) and the profession of the profession of the plantiff profession and the algorithm (social principles) and the profession of the profe	Qualitative results (interviews) that uncover numerous aspect of gains, challenges and problems introducing an online triage tool "Teledoc". As case study researchers, the authors prioritise opportunity to learn over concerns about typically and particularization over generalisation. Single case studies, like this one, can be valuable	To large amount based on personal opinions from relability few persons. This study was smal, undertaken as an MSc project, and focused on staff not patients. It involved one atypical practice — an early adopter, closely related to the software developers, and interested in commercial opportunity. Participants were likely to be heavily invested in making Tele-Doc work.	Tele-Doc offers a new kind of consultation that is still be worked out in practice. It may offer convenience for paid with discrete, single problems, and a velcome variation (Pe's workload. Tele-Doc's potential for a
lect of an evidence-based website. healthcare usages an interrupted is-series study, cockman WA, Borten TN, de Waal W, Dronthen T, Sneele LJ, Nielen M, Chavannes NH.	2016	The Netherlands	general practitioners 18.1 million times during the study period	Observational interrupted time series (ITS) analysis was used. ITS is conside-red a strong design when randomization is not possible and can thus be used to study causal effects using an observational	After baunch of the website, the trend in consultation rate decreased with 1,800 consultations (100) quantities more in 2001. This corresponds to 12% declare in consultance of years after braunch of the production of the consultance of the consultance of the consultance of the consultance of the analyses showed as specific declare for consultations by prime and were appricant for all other subspicups, except for supragrating approximations to prime and were appricant for all other subspicups. Except for supragrating approximation in the first consultance into before release of the verbolish, the declare after release of the website and the absence of changes in consultation rate in the releasement groups supports a causal relation between the intervention and the observed defects.	Long-term follow-up (6 years) in a large number of healthcare consultations (18.1 million). An interrupted time series analysis is the strongest research design when randomization is not possible. A change in trend of the outcome before and after the intervention and additional use of a	Other events might have affected healthcare usage during the study proint, but were carefully considered and analyzed in sensitivity analyses which showed shifter results as the primary analysis. A limitation is that, owing to nation-wide implementation of the website, no reference group; could be created which was not exposed to the website at all during the study period.	The effect of providing evidence-based orinis health internation in healthcare usage on a population level was international to the providence of the control of the an evidence based health website, nationally private usage decreased by 1745. This effect was most provint phone consultations and present in all subgroups of se- parations of the providence of the providence providence of the providence of the providence that effects the proprior seed immagning reduce healthcare usage in times of increasing healthcare codes.
emparison of Physician and, emputer Disgnostic Accuracy, migran HL, Levine DM, Nundy S, shrotra A. MA Intern Med. 2016 Dec		U.S. (from reference list)	checkers solving 1105 clinical vignettes	Direct comparison of diagnostic accuracy - physician vs. computer algorithms	Across physicians, they were more likely to list the correct diagnosis first for high-acuty vigneties (vs low-acuty) vigneties) and for uncommon vigneties (vs common vigneties), to contents, symptom checkers were more likely to list the correct diagnosis first for low-acuty vigneties and common vigneties	Direct comparison of human vs. digital symptom checkers performance which a clinical relevant approach. The use och clinical vignets makes the study highly reproducible.	Investigation on vignets are potentially biased compared to using real patients who use common everyday language. Key initiations included the use of clinical vigneties, which likely do not reflect the complexity of real-world patients and clid not include physical examination or test results. Physicians and year of the proposed proposed of US physicians and therefore may differ in dismostria, encursus. Does not.	In what we believe to be the first direct comparison of diagnostic accuracy, physicians vastly outperformed computer algorithms in diagnostic accuracy, (84.39% vs. correct diagnostis in the top 3 listed). Despite physicians superior performance, they provided the incorrect diagn in about 15% of cases, primits to prior estimates (10%-for physician diagnostic error. The project compared unknown of the project com
abation of symptom checkers for disagnosis and triage; audit study- ingran HL, Linder JA, Gidengil C, shrotra A. AJ. 2015 Jul		U.S. (from reference list)	patient vignettes to audit 23 symptom checker	Statistic analysis of diagnostic accuracy and triage advice. Algirithm vs. Goldstandard	Physicians should be aware that an increasing number of their patients are using new internet based books such as symptom choices and that their diagnoss and trigos and chop patients receive may often be incurcials. Symptom choices may, however, be of value if the alternative is not seeking any advisor proportion based seeking any advisor of the alternative is not seeking any advisor incurrent based seeking any advisor and any advisor of the alternative is not seeking any advisor proportion based seeking and any advisor of the alternative is not any advisor decisions about their health. The 2 symptom choices provided the correct diagnosis list in 34% (50% confidence interval 31% to 27%) of standardised paster evaluations, seek the provided for the appropriate high parkines in 25% (25% of the 25%	The use och clinical vignets makes the study highly reproducible.	therefore most offer in dismonder, accuracy. Does not The symptom checkers was evaluated as a whole, individual high/low performance was therefore not possible to identify. Vigets does not reflect the complexity or language of real world patients.	Trage advice trom symptom checkers is generally risk averse, encouraging users to seek care for conditions self care is reasonable.
itizing diagnosis: a review of mobile locations in the diagnosis: process, el A. Lupton D. gnosis (Berl). 2015 Jun 1		New Zealand /Australia (from reference list)		Systematic content and descriptions analysis of 131 health apps	The analysis found that while these pags provide access to medical information previously visidily returned valuable to be user and offered a convenient diagnostic cost for practioners, many lated to describe the evidence base underprinning or any other credential supporting, their design and use. These apps pointedly with the ordigance operation, encoding redisease concepts and by professional relations. proceedings of the ordinary of th	Systematic approach, large number of apps investigated.	Mainly qualitative extraction of data. Possibility for bias through authors subjective interpretation	The diagnosis app raises not only the spector of shifting of submitty in the doubter patient encounter, but along up about a gardas concealed behind the algorithmic applic about qualify centre and ceredentials. They approach diagnosis from a nocessarily binary system of informat efficiently from the control of the contr
w to develop web-based decision, oport interventions for patients: a ceess map. yn G, Kreuwel I, Durand MA, Sivell Joseph-Williams N, Evans R, wards A, thent Educ Couns. 2011 tient use and compliance with	2011			Expert opinion. Based on experiences of developing and evaluating web- based decision support interventions. Mixed-Method	General principles are outlined: The content specification is fundamental and should go beyond the hadfored emphasis on scientific evidence to establish patients' perspectives on the salmost and hadfored emphasis on scientific evidence to establish patients' perspectives on the salmost and the salmost emphasis of the salmost emphasis emphas	Gives insight in "hand on experience" with the Topic and to large extent relate to current evidence	A none systematic expert opinion to some extent based on the authors own experience	making pathways of dispross appos – their algorithmic. This article provides a process may of how to develop based decision support inherventions for patients. The guiding principle need to be one of custorizing hid delivery, based on close consultation with the target us and an herative development process until the interven deemed accessible and useful. In conclusion, web-based trisse can contribute to a me
itent use and compliance with adical advice delivered by a web- sed triage system in primary care, land N, Cranen K, Boer H, van imert-Pijnen JE, Seydel ER. Telemed Telecare. 2010 preasing the use of e-consultation in	2010	Netherlands	entered, 3812 patients went through triage. Online survey 192 patients	Retrospective observational study + prospective online survey	So there was an inchalance between the central date of web-based trags and the intended use of it. As self amangement is the focus of electronic care, an in-depth evaluation of patients remost for automotival and their readiness to make decisions about their health care is important. Over a period of 13 months, 13.13 different people began using the web-based tribage system and 3812 patients went right travolu- tion and the process to the end. The most frequent complaints were common cold symptoms, such as 10 months of the process of the end of the	retrospective part	rates and invariability of the generated advice. The dropout rates were high, especially before entering a complaint (50%) and just before the system delivered advice (40%). Prospective survey: Small number of patients went through the full followup in the prospective study (58) and actually complained (20). Participants recruited through banners on frequently	efficient primary care system, because it facilitates the gatekeeper function. This implies that empowering pair means of web-based triage requires interventions to in and motivate users; which is not common in the implementation of health care technologies. Special streeting should be post to patient who can be
mary care: results of an online vey among non-users of e- nsultation. and N, van Gemert-Pijnen JE, Boer Steehouder MF, Seydel ER. J Med Inform. 2009 Oct		Netherlands (from reference list)		study, Patients enroll through banners on patient organizations websites	consultation, e-consultation not being provided by a CP and the preference to see a doctor. Education and examination of user expectations can provide a solution for these barriers, for both patients and caregivers alike. Platents are depended on a GPs ² provision of e-consultation. Therefore, it is important to a whose caregivers are in the faulta burefled of e-consultation, to encergances and implementation into the consultation is consequence and implementation in the consultation of the consultation is consequence and implementation in the consultation of the consultation. The study revealed that certain patient groups, such as see educated patients, develop patients and chronic users of medication were especially modified to	designed questions exploring the issue	visited websites of 28 well-trusted patient organizations, lawing possibility for seletion bias. Results based on participants self reported preferences	the most from e-consultation, while also facing the great chance of being excluded from the service. As health o- continues to evolve towards a more patient centered approach, Authors expect that patient expectations and demands will be a major force in driving the use of elect communication.
cision support system for medical ge. deghi S, Barzi A, Zarrin-Khameh N.	2001	U.S.	Not described	Preliminary clinical reviews, case studies	An intelligent triage system most be able to handle uncertainty and gabs in data as well subjective descriptions and perceptions of symptoms since data is filled in by patients. Also for the system in order to work efficiently its more important that correct disposition is made than the correct diagnoses is	Some degree of insight into the technical challenges designing intelligent triage systems,	The article is mostly a synopsis meant for planing future studies and a description of a planned triage system. There is no presentation of real results or studie population. The combining of the studie is	Bayesian networks is an efficient way of performing tri it brings consistency and reproducibility of results and reduces human or systemic errors.

Gottliebsen K, Petersson G. BMJ Health Care Inform 2020; 27:e100114. doi: 10.1136/bmjhci-2019-100114