

Data Supplement

Supplementary Tables

Age-range	1 month to 89 years
Male / female	43.0% / 57.0%
Body-system/specialism classification	
Haematology	1.5%
Cardiovascular	7.0%
Central Nervous System - non injury	6.0%
Dental	2.0%
Endocrine	3.5%
Ear, Nose, Throat (ENT)	3.5%
Female reproductive system	7.5%
Gastrointestinal	15.0%
Immune system	2.0%
Infection	3.5%
Lower respiratory	8.0%
Male reproductive system	2.5%
Mental health	5.0%
Musculoskeletal - injury	3.5%
Musculoskeletal - non-injury	7.5%
Oncology	0.5%
Ophthalmology	3.5%

Skin	8.0%
Upper respiratory	5.0%
Urinary tract	4.5%

Supplementary Table 1
Summary of the 200 clinical vignettes: age range, sex and condition type. See also **Supplementary Table 2** for a more detailed description of each vignette and **Supplementary Table 3** for the structure of example vignettes.

#	Included in main analysis	Sex M/F	Age	System/specialism	Primary complaint	Advice level	Main-diagnosis
1	YES	F	30 years	Female reproductive system	Abdominal pain	1	ectopic pregnancy
2	NO	F	6 years	Lower Respiratory	Shortness of breath	2	chest infection
3	YES	F	9 months	Lower Respiratory	Shortness of breath	1	bronchiolitis
4	YES	M	66 years	Cardiovascular	Chest pain	1	unstable angina
5	YES	M	35 years	MSK - Injury	Back pain	4	contusion lumbar spine
6	YES	M	23 years	Mental Health	Low mood	3	depression
7	YES	M	26 years	Gastrointestinal	Vomiting	3	viral gastroenteritis
8	YES	F	40 years	CNS - Non injury	Headache	1	brain haemorrhage
9	YES	M	40 years	Skin	Swollen foot	4	cellulitis of foot
10	YES	F	28 years	Dental	Toothache	6	dental abscess
11	YES	M	1 years	Upper respiratory	Lumps in neck	4	tonsillitis
12	NO	F	26 years	Dental	Pain in face	4	dental abscess
13	YES	F	6 years	Gastrointestinal	Itchy bottom	6	pinworm infection
14	YES	F	5 months	Gastrointestinal	Crying baby	5	cow's milk protein intolerance
15	YES	F	33 years	MSK - Injury	Abdominal pain	4	musculoskeletal back pain
16	YES	F	20 years	Immune system	Dizziness	1	allergic reaction
17	YES	M	14 years	Upper respiratory	Sore throat	6	sinusitis
18	YES	M	3 months	Upper respiratory	Crying baby	4	common cold
19	NO	F	27 years	Female reproductive system	Vaginal bleeding	2	early onset of labour
20	YES	F	86 years	Gastrointestinal	Vomiting brown liquid	1	perforated gastric ulcer
21	YES	M	57 years	Gastrointestinal	Diarrhoea	6	gastroenteritis
22	YES	F	29 years	Gastrointestinal	Abdominal pain	2	appendicitis
23	YES	M	2 years	CNS - Non injury	Fever	1	bacterial meningitis

24	NO	M	59 years	ENT	Earache	4	otitis externa
25	NO	F	1 years	Infection - General	Rash	6	varicella zoster
26	NO	M	10 years	Upper respiratory	Cough	6	upper respiratory tract infection
27	YES	M	9 months	Skin	Rash	6	eczema
28	NO	F	64 years	Lower Respiratory	Cough	6	acute bronchitis
29	YES	F	17 years	Lower Respiratory	Muscle aches	4	influenza
30	YES	M	14 years	Gastrointestinal	Diarrhoea	4	appendicitis
31	YES	F	20 years	Female reproductive system	Vaginal discharge	4	threatened miscarriage
32	YES	F	42 years	Mental Health	Low mood	4	postnatal depression
33	YES	F	2 years	ENT	Sticky eye	4	otitis media
34	YES	F	57 years	Mental Health	Aggressive behaviour	3	paranoid psychosis
35	YES	F	72 years	MSK - Injury	Back pain	2	wedge fracture thoracic spine
36	YES	F	89 years	Mental Health	Confusion	3	acute confusional state
37	YES	F	49 years	Mental Health	Suicidal thoughts	3	post traumatic stress disorder
38	NO	F	1 years	Lower Respiratory	Breathing difficulties	3	bronchiolitis
39	YES	M	89 years	Lower Respiratory	Hands shaking	3	influenza
40	YES	F	2 years	Upper respiratory	Earache	6	common cold
41	NO	F	5 years	CNS - Non injury	Fever	1	bacterial meningitis
42	YES	M	68 years	Gastrointestinal	Vomiting blood	2	acute gastritis/gastric erosions
43	YES	F	41 years	Gastrointestinal	Abdominal pain	2	acute cholecystitis
44	NO	F	52 years	Skin	Rash	4	shingles
45	YES	F	7 years	CNS - Non injury	Fever	1	meningococcal septicaemia
46	YES	M	3 years	Infection - general	Blisters on hands	6	hand foot and mouth disease
47	NO	M	30 years	MSK - non injury	Pain in big toe	4	gout
48	NO	F	47 years	MSK - Injury	Back pain	6	muscle strain
49	YES	F	27 years	Upper respiratory	Sore throat	2	quinsy (peritonsillar abscess)

50	NO	M	2 years	ENT	Earache	4	acute otitis media
51	YES	M	62 years	Mental Health	Can't sleep	5	depression
52	YES	M	58 years	Cardiovascular	Chest pain	5	angina
53	NO	F	37 years	Lower Respiratory	Cough	6	acute bronchitis
54	YES	F	67 years	ENT	Dizziness	6	benign positional vertigo
55	YES	F	32 years	Cardiovascular	Shortness of breath	2	pulmonary embolism
56	YES	M	77 years	CNS - Non injury	Drooping of left side of face	1	stroke
57	NO	F	82 years	Cardiovascular	Shortness of breath	5	heart failure (i.e. congestive cardiac failure)
58	YES	F	32 years	Cardiovascular	Irregular heartbeat	5	ventricular extrasystoles
59	YES	M	28 years	CNS - Non injury	Shaking of hands	5	benign essential tremor
60	YES	F	58 years	Lower Respiratory	Cough	5	asthma
61	NO	F	27 years	CNS - Non injury	Headache	5	migraine
62	YES	M	57 years	Cardiovascular	Pain in calves	5	peripheral vascular disease
63	NO	F	41 years	CNS - Non injury	Headache	1	subarachnoid haemorrhage
64	NO	M	32 years	Cardiovascular	Fast pulse	4	paroxysmal tachyarrhythmia
65	YES	F	1 years	Lower Respiratory	Cough	4	viral bronchitis
66	YES	M	13 years	Lower Respiratory	Wheezy	5	asthma
67	YES	F	3 years	Lower Respiratory	Cough	3	pneumonia
68	YES	F	32 years	MSK - non injury	Tingling in hands	5	carpal tunnel syndrome
69	NO	M	42 years	Cardiovascular	Chest pain	1	myocardial infarction
70	NO	M	10 years	CNS - Non injury	Headache	2	intracranial haematoma
71	YES	F	27 years	MSK - non injury	Colour change in fingers	5	Raynaud's phenomenon of the fingers
72	NO	F	26 years	Cardiovascular	Collapse	5	vasovagal attack
73	YES	M	73 years	Lower Respiratory	Cough	4	lung cancer
74	YES	F	8 years	CNS - Injury	Headache	6	minor head injury

75	YES	M	28 years	MSK - Injury	Chest pain	4	pleurisy
76	NO	F	77 years	Gastrointestinal	Abdominal pain	1	intestinal obstruction
77	YES	F	63 years	Gastrointestinal	Blood on toilet paper	4	colorectal cancer
78	NO	F	37 years	Gastrointestinal	Abdominal pain	5	irritable bowel syndrome
79	YES	M	8 years	Gastrointestinal	Abdominal pain	2	appendicitis
80	YES	F	17 years	Gastrointestinal	Vomiting	6	viral gastroenteritis
81	NO	M	1 months	Gastrointestinal	Vomiting	3	pyloric stenosis
82	NO	M	38 years	Gastrointestinal	Yellow eyes	2	viral hepatitis
83	YES	M	67 years	Gastrointestinal	Difficulty swallowing	4	oesophageal cancer
84	YES	F	68 years	Skin	Redness of leg	4	cellulitis
85	YES	M	37 years	Ophthalmology	Swollen eye	2	periorbital/orbital cellulitis
86	YES	F	63 years	MSK - non injury	Shoulder pain	5	frozen shoulder
87	YES	M	63 years	MSK - non injury	Stiff neck	6	acute torticollis
88	YES	F	82 years	MSK - non injury	Hip pain	5	osteoarthritis of hip
89	YES	M	15 years	MSK - non injury	Knee pain	5	Osgood-Shlatters disease
90	YES	F	35 years	MSK - non injury	Numbness in fingers	5	carpal tunnel syndrome
91	YES	M	50 years	MSK - non injury	Knee pain	4	infected infra-patellar bursa
92	YES	F	50 years	MSK - non injury	Wrist pain	5	tenosynovitis
93	NO	M	67 years	ENT	Discharge from ear	4	otitis externa
94	YES	M	50 years	ENT	Ringing in ear	5	tinnitus
95	YES	F	58 years	Upper respiratory	Hoarse voice	4	laryngitis secondary to acid reflux
96	YES	F	23 years	Mental Health	Hearing voices	3	acute psychosis
97	NO	F	13 years	Mental Health	Heart pounding	5	anxiety disorder
98	NO	F	47 years	Endocrine	Passing lots of urine	4	type 2 diabetes
99	YES	M	9 years	Endocrine	Vomiting	1	diabetic ketoacidosis
100	YES	F	30 years	Endocrine	Sweating	4	Graves' disease

101	NO	F	59 years	Endocrine	Tiredness	5	hypothyroidism
102	YES	M	41 years	Endocrine	Can't get erection	5	pituitary tumour (prolactinoma)
103	YES	F	42 years	Endocrine	Swelling under jaw	5	salivary gland stone
104	YES	F	39 years	Endocrine	Weight gain	4	adrenal adenoma causing Cushing's syndrome
105	YES	M	61 years	Infection - general	Fever	2	malaria
106	YES	F	37 years	Urinary Tract	Painful urination	4	urinary tract infection
107	NO	M	45 years	Ophthalmology	Red eyes	6	viral conjunctivitis
108	YES	F	33 years	Female reproductive system	Abdominal pain	2	ectopic pregnancy
109	YES	F	6 years	Urinary Tract	Abdominal pain	3	urinary tract infection
110	NO	M	26 years	Urinary Tract	Pain in side	2	renal stones
111	YES	M	33 years	Ophthalmology	Red painful eye	3	corneal ulcer
112	YES	M	15 years	Male reproductive system	Testicular pain	2	testicular torsion
113	YES	F	42 years	Female reproductive system	Vaginal discharge	6	vaginal thrush
114	YES	M	57 years	Urinary Tract	Blood in urine	4	urinary tract infection
115	YES	F	14 years	Female reproductive system	Heavy periods	5	menorrhagia
116	YES	F	26 years	Female reproductive system	Vaginal bleeding	6	threatened miscarriage
117	YES	M	79 years	Male reproductive system	Passing urine at night	5	benign prostatic hyperplasia
118	YES	M	27 years	Ophthalmology	Dry eyes	6	dry eye syndrome
119	YES	M	2 years	Ophthalmology	Sticky red eye	6	bacterial conjunctivitis
120	YES	M	19 years	Male reproductive system	Testicular lump	4	hydrocoele
121	YES	F	45 years	Female reproductive system	Vaginal bleeding	4	vaginal atrophy
122	YES	F	62 years	Female reproductive system	Bloating	4	ovarian pathology (cancer)

123	YES	F	56 years	Female reproductive system	Hot flush	5	menopause
124	YES	M	7 years	Gastrointestinal	Constipation	6	constipation
125	NO	F	13 years	Immune system	Sneezing	5	allergic rhinitis
126	YES	M	29 years	Cardiovascular	Palpitations	5	ectopics secondary to caffeine intake
127	YES	F	19 years	Mental Health	Low mood	5	depression
128	YES	M	38 years	Lower Respiratory	Phlegm with blood	4	carcinoma of lung
129	YES	M	32 years	Infection - general	High fever	2	malaria
130	YES	M	20 years	Skin	Rash	4	eczema
131	YES	F	34 years	Skin	Pimples on face	5	acne
132	YES	F	3 years	Infection - general	Rash	6	varicella zoster
133	YES	M	27 years	Immune system	Itchy rash	6	acute urticaria
134	YES	M	33 years	Skin	White toenail	5	fungal nail infection
135	YES	F	37 years	Blood	Tiredness	5	iron deficiency anaemia
136	YES	F	11 years	Blood	Lump in neck	4	lymphoma
137	NO	M	43 years	Blood	Bruising	4	clotting disorder
138	NO	F	33 years	Cardiovascular	Chest pain	3	pulmonary embolism
139	YES	M	40 years	Gastrointestinal	Stomach pain	3	diverticulitis
140	YES	M	50 years	Gastrointestinal	Stomach pain	1	diverticular abscess
141	YES	M	64 years	Male reproductive system	Difficulty passing urine	4	benign prostatic hyperplasia
142	YES	F	64 years	Oncology	Lump in breast	4	breast cancer
143	NO	F	38 years	CNS - Non injury	Headache	5	tension headache
144	NO	F	25 years	Female reproductive system	Tummy bloating before period	5	premenstrual syndrome
145	NO	F	35 years	Gastrointestinal	Abdominal pain	5	irritable bowel syndrome
146	YES	F	36 years	Skin	Hair falling out	5	alopecia
147	YES	F	39 years	Skin	Flaky patches of skin	5	psoriasis

148	YES	F	3 years	Urinary Tract	Crying on weeing	3	urinary tract infection
149	YES	F	3 years	Lower Respiratory	Cough	3	pneumonia
150	NO	M	3 years	Male reproductive system	End of penis red	4	balanitis
151	YES	F	12 years	Skin	Rash	4	scabies
152	YES	F	32 years	MSK - non injury	Heel pain	5	plantar fasciitis
153	YES	M	30 years	MSK - non injury	Shin pain	6	shin splints
154	YES	M	40 years	CNS - Non injury	Back pain	1	cauda equina syndrome
155	YES	F	30 years	Female reproductive system	Irregular periods	5	polycystic ovary syndrome
156	NO	F	12 years	MSK - non injury	Hip pain	4	slipped upper femoral epiphysis
157	YES	M	21 years	Lower Respiratory	Shortness of breath	1	spontaneous simple pneumothorax
158	YES	F	70 years	Lower Respiratory	Shortness of breath	4	asbestosis
159	YES	F	33 years	Cardiovascular	Chest pain	2	pulmonary embolism
160	YES	F	47 years	Cardiovascular	Lump in lower leg	4	thrombophlebitis
161	NO	F	4 years	Upper respiratory	Runny nose	6	upper respiratory tract infection
162	YES	M	10 years	Skin	Rash on chin	4	impetigo
163	YES	M	40 years	Urinary Tract	Pain in left side	2	renal calculus
164	NO	M	62 years	Skin	Cracked skin between toes	6	athlete's foot
165	YES	M	52 years	Gastrointestinal	Blood on toilet paper	5	internal haemorrhoids
166	YES	M	39 years	ENT	Blocked nose	5	nasal polyps
167	NO	F	21 years	MSK - injury	Swollen ankle	4	ankle sprain
168	YES	F	11 years	MSK - non injury	Back ache	5	idiopathic scoliosis
169	YES	M	18 years	MSK - non injury	Knee pain	2	dislocated patella
170	YES	M	24 years	Skin	Rash on hands and arms	4	contact dermatitis
171	NO	F	2 years	Ophthalmology	Sticky eyes	6	viral conjunctivitis

172	YES	M	16 years	Immune system	Runny nose	6	allergic rhinitis
173	YES	M	68 years	Urinary tract	Blood in urine	4	bladder cancer
174	YES	F	36 years	Skin	Rash under breasts	4	candidal intertrigo
175	YES	F	32 years	Female reproductive system	Bleeding after sex	4	cervical cancer
176	YES	F	27 years	CNS - Non injury	Headaches	5	cluster headaches
177	NO	M	23 years	Gastrointestinal	Diarrhoea	4	coeliac disease
178	YES	M	6 months	Skin	Flaky scalp	6	cradle cap
179	YES	M	10 months	Upper respiratory	Difficulty breathing	3	croup
180	NO	F	28 years	Female reproductive system	Abdominal pain	5	endometriosis
181	YES	F	23 years	Female reproductive system	Blisters on vulva	4	genital herpes
182	YES	F	10 months	Gastrointestinal	Crying baby	3	constipation
183	YES	M	2 years	Upper respiratory	Cough	6	common cold
184	YES	F	5 years	Infection - general	Rash	4	rubella
185	NO	M	50 years	Dental	Toothache	4	dental abscess
186	NO	M	36 years	Ophthalmology	Blurred vision	2	corneal ulcer
187	NO	F	6 months	Gastrointestinal	Diarrhoea	3	viral gastroenteritis
188	YES	M	27 years	Gastrointestinal	Abdominal pain	1	acute pancreatitis
189	NO	M	2 months	Urinary Tract	Vomiting	3	urinary tract infection
190	YES	F	7 months	Gastrointestinal	Vomiting	2	viral gastroenteritis
191	YES	M	9 years	Dental	Toothache	5	dental caries
192	YES	F	33 years	Mental Health	Worried about pregnancy	4	anxiety
193	YES	M	24 years	Infection - general	Fever	4	glandular fever
194	YES	F	4 months	Gastrointestinal	Diarrhoea	4	viral gastroenteritis
195	YES	F	15 years	MSK - injury	Back pain	5	mechanical low back pain
196	YES	F	22 years	Urinary Tract	Pain on passing urine	3	urinary tract infection
197	YES	F	10 months	Gastrointestinal	Vomiting	4	viral gastroenteritis

198	NO	F	4 years	Cardiovascular	Fever	3	Kawasaki disease
199	YES	F	60 years	Skin	Crusted patch of skin	4	basal cell carcinoma
200	YES	F	56 years	Gastrointestinal	Pain in groin	4	femoral hernia

Supplementary Table 2

Individual details of the 200 clinical-vignettes: age range, sex, condition type (system/specialism), level of urgency advice, and main diagnosis. See also **Supplementary Table 1** for a summary.

Category	Example vignette #79 (see Supplementary Table 2)	Example vignette #86 (see Supplementary Table 2)
Sex	male	female
Age	8 years old	63 years old
Previous medical history	None of note.	None of note.
Primary complaint	abdominal pain	shoulder pain
Additional information on primary complaint and current symptoms	abdominal pain and fever. Temp 39.5C. Started to get pain in the middle of his abdomen last night before he went to bed (6 hours ago). Had some infant paracetamol suspension and went to sleep. Now he has woken up and is crying with pain. Feels feverish. Says his tummy hurts on the right side now. Severity 8/10. Tummy hurts to touch. Feels hot. Looks flushed. Constant pain - no radiation.	painful shoulder. Unable to move it for 24 hours. Onset of pain in right shoulder about 6 months ago. Gradually got worse. In the past week it has been much worse and now unable to move the shoulder. Right shoulder appears higher than the other. Background ache. Severe pain if she tries to move the shoulder.
If asked	location of pain: Right lower abdomen Character: constant Diarrhoea: NO constipation - NO blood in stools - NO mucus - NO urinary or testicular symptoms - NO vomiting - NO dehydration - NO fever - YES dry Mouth - NO	history of injury to the shoulder: NO redness of the skin over the shoulder: NO painful to touch: NO deformity: NO numbness, tingling or pain in the arm: NO weakness of the arm: NO – just can't lift it up because of pain. does the shoulder feel as if it is coming out of socket?: NO
Progression	getting worse	getting worse
Levels of urgency advice	see primary care within 4 hours	see primary care non-urgent
Main diagnosis	appendicitis	frozen shoulder
Differential diagnoses (not ordered)	urinary tract infection mesenteric adenitis inguinal hernia testicular torsion intestinal malrotation ureteric colic	osteoarthritis of shoulder bursitis of shoulder shoulder impingement syndrome dislocated shoulder

Body-system/ specialism classification	gastrointestinal	musculoskeletal - non-injury
--	------------------	------------------------------

Supplementary Table 3

Two example vignettes showing the vignette structure. The items in grey shading are the vignette gold-standard answer and vignette body-system/specialism classification.

Name of symptom assessment app	Provider	Platform addressed in this study	Regional availability	Questions on pre-existing factors/ attributes	Number of questions
Ada	Ada Health GmbH, Berlin, Germany	App	Available everywhere.	-Gender -Date of birth -Pregnant -Smoker -High blood pressure -Diabetes	Variable
Babylon	Babylon Healthcare Services Ltd., London, UK	App	Services provided in compliance with UK law and regulation-- outside of the UK, user must check if lawful. Not all services available outside of UK UK version tested in this study.	-Gender -Date of birth	Variable
Buoy	Buoy Health, Inc., Boston, MA, USA	Website (Note - the term 'app' is loosely used in this study as a generic term for symptom assessment tools - strictly Buoy is not an app, but instead a website, though it can be put onto phone homescreens	For US residents US version tested in this study.	-Gender -Age (between 2-120 years)	Variable
K Health	K Health Inc., New York, NY, USA	App	Services available outside the US & Israel but are solely directed to those in the US and/or Israel. US version tested in this study.	-Gender -Age -Chronic conditions relevant to symptoms	Around 20 questions but variable
Mediktor	Teckel Medical S.L., Barcelona, Spain	App	Available everywhere	-Gender -Age	Variable
Symptomate	Infermedica, Wroclaw, Poland	App	Available everywhere	-Gender -Age (Minimum age 18) -Overweight -Smoker	19 after region where live

				-Recently injured -High cholesterol -Hypertension - Diabetes	
Your.MD	YOUR.MD AS Oslo, Norway and subsidiary: Your.MD Ltd. London, UK	App	Available everywhere	-Gender -Year of birth (-Minimum age 16) -Asthma -Diabetes -Smoker	Variable
WebMD	WebMD, LLC Atlanta, GA, USA	App (website also available)	Available everywhere	-Gender -Age -Past medical history -Medications	4 (excluding age and gender)

Supplementary Table 4

Summary of the 8 apps evaluated in this study

Conditions should be considered a reasonable match when:
- the suggested condition is the same as the gold standard (GS).
- alternative names for the same condition are used.
- the suggested condition is a more precise description of the gold standard condition.
- the condition is an umbrella term including the other condition.
- it is reasonable to assume two different doctors might use the two different descriptions to label the same conditions.
- one condition causes the other (directly and explicitly)
- the suggested condition conveys the nature of the GS condition, is reasonably related to the GS condition (but is less precise) - similarity is so clear that in primary care medicine practice, the conditions would be considered a near match.
- the suggested condition is both highly related to the GS condition and shares symptoms to a high degree with the GS condition.

Supplementary Table 5

Matching criteria that were considered by panel physicians in deciding if condition-suggestions (from GPs or from apps) match the gold-standard diagnoses. Note - the matching criteria apply equally to all apps and to the tested-GPs.

Levels of urgency advice	all vignettes for K Health (200)	150 vignettes with source verified data for K Health	all vignettes for Your.MD (200)	150 vignettes with source verified data for Your.MD
no advice given (%) [not recorded]	32.0 [12.0]	33.3	32.5	38.7
potentially unsafe (%)	10.5	14.0	5.0	4.7
safe but underconservative (%)	18.0	23.3	5.0	4.0
exactly matched (%)	17.0	20.0	22.0	23.3
overconservative but suitable (%)	6.0	6.7	14.0	15.3
overconservative (%)	4.5	2.7	21.5	14.0
statistical testing (%) Fisher's exact test for significant difference between the urgency advice allocation	<i>p</i> -value	0.59 (no significant difference)	<i>p</i> -value	0.69 (no significant difference)
% of safe advice (where safe is defined as maximum one level less conservative than gold-standard, expressed per vignette provided with advice).	81.3*	79.0	92.6	92.3

Supplementary Table 6

Urgency advice subanalysis for K Health and Your.MD. One app-entry-Dr (#4) did not record complete and source data verifiable data for K Health and Your.MD. For K Health, 12.0% of urgency advice was not recorded by app-entry-Dr (#4) and no screenshot was saved. For Your.MD, all urgency advice was recorded but in 25% of vignettes, screenshots were not saved for source data verification. A subanalysis was conducted to show the K Health and Your.MD performance in the fully verified 150 vignettes and to allow comparison to their performance in the full set of 200 vignettes. There are some differences in Your.MD's performance in the 200-vignette and in the 150-vignette analyses, but these are not statistically significant. It is not possible to determine what proportion of the differences present was due to unrecorded or unverified data, and what proportion was due to random differences between the 150 and 200 vignettes. * - in the case of K Health, % of safe advice (where safe is defined as maximum one level less conservative than gold-standard, expressed per vignette with available advice).

Condition-suggestion performance	all vignettes for WebMD (200)	150 vignettes with source verified data for WebMD	all vignettes for Your.MD (200)	150 vignettes with source verified data for Your.MD
M1 (%)	21.0	25.3	21.5	18.7
M3 (%)	35.5	38.0	23.5	20.0
M5 (%)	43.5	45.3	23.5	20.0
COMP. (%)	38.9	42.5	14.3	12.0
RELE. (%)	33.4	34.8	40.2	35.1
statistical testing (%) Fisher's exact test for significant difference between the urgency advice allocation	<i>p</i> -value	0.88 (no significant difference)	<i>p</i> -value	0.88 (no significant difference)

Supplementary Table 7

Condition-suggestion subanalysis for WebMD and Your.MD. One app-entry-Dr (#4) did not record complete source data verification screenshots for WebMD or for Your.MD and therefore verification of entered data was not possible for 50 out of 200 of the randomly assigned vignettes for these apps. A subanalysis was conducted to show the performance of WebMD and Your.MD in the fully verified 150 vignettes and to allow comparison to their performance in the full set of 200 vignettes. There are some differences in performance between the 200-vignette and the 150-vignette analyses, but these are not statistically significant. It is not possible to determine what proportion of small differences present were due data entry error in the 50-vignettes, and what proportion was due to random differences between the 150 and 200 vignettes.

Scenarios for which there was no urgency advice							
	Under the age limit ¹ % (specified age limit)	Pregnant %	Symptom Not Found ² %	No condition-suggestions ³ %	Not Recorded ⁴ %	Mental health ⁵ %	Total for which no urgency advice (i.e. 100% - breadth of coverage)
Ada	0	0	0	0	0	0	0
Babylon	1.0 (16+)*	3.0	5.0	0	0	0	9.0
Buoy	9.5 (13+)	0	0.5	0	0	0	10.0
K Health	31.0 (18+)	0	1.0	0	12.0	0	44.0
Mediktor	2.5 (1+)	0	0	13.5	0	1.5	17.5
Symptomate	31.0 (18+)	0	0.5	0	0	0	31.5
Your.MD	23.0 (16+)	0.5	2.0	5.5	0	0	31.0

Supplementary Table 8

Coverage of each app and reasons why apps did not provide condition-suggestions for some vignettes.

* - Babylon provides levels of urgency advice to these patients such as “go to the nearest ED”, but it does not provide condition-suggestions for patients under 16 years of age.

¹ - The app did not provide a condition-suggestion for those under its specified age limit.

² - A matching symptom or appropriately matching symptom to the presenting complaint could not be found in the app, so the vignette could not be entered.

³ - No condition-suggestions were provided by the app, and, as a result of this, the app did not provide urgency advice.

⁴ - One app-entry-Dr (#4) did not fully record the levels of urgency advice, and there were no corresponding source data verification screenshots for this subset of data. See **Supplementary Table 6** for a subanalysis of the 150 vignettes with complete source-data-verified data for K Health on levels of urgency advice.

^s - These were mental health scenarios where no level of urgency advice was provided. Note: for the other apps, where question flow was stopped due to mental health, the level of urgency advice of 'seek emergency care' was none-the-less provided.

Scenarios for which there were no condition-suggestions								
	Under the age limit ¹ % (specified age limit)	Pregnant %	Symptom Not Found ² %	Condition Severity ³ %	Generic ⁴ %	Clear Statement No Results ⁵ %	Mental health ⁶ %	Total for which no condition suggestions (i.e. 100% - breadth of coverage)
Ada	0	0	0	0.5	0	0	0.5	1.0
Babylon	25.5 (16+)	2.5	4.0	8.5	6.0	0.5	1.5	48.5
Buoy	8.5 (13+)	0	1.0	0.5	0	0	1.5	11.5
K Health	23.5 (18+)	0	2.0	0	0	0	0	25.5
Mediktor	4.0 (1+)	0	1.0	1.0	0	11.5	2.0	19.5
Symptomate	31.5 (18+)	1.5	0	0	0	5.0	0.5	38.5
Your.MD	23 (16+)	1.0	0.5	0.5	4.0	5.5	1.0	35.5
WebMD	5.0	0	0.5	0	0	1.5	0	7.0

Supplementary Table 9

The breadth of condition-suggestion coverage is shown for each app along with the reasons why apps did not provide condition-suggestions for some vignettes.

¹ - The app does not provide a condition-suggestion for users under a specified age limit.

² - A directly or appropriately matching symptom to the presenting complaint could not be found in the app, so the vignette could not be entered.

³ - The app did not to give condition-suggestions for very serious symptoms - e.g. the app stated only 'Condition causing severe [*symptom*]'

⁴ - The app gave a generic answer rather than a condition, e.g. "further assessment is needed"

⁵- The app provided a clear statement that no condition-suggestion results were found for the vignette. The reason why the app failed to give a condition-suggestion for these vignettes is uncertain, but generally these vignettes relate to minor conditions, and in most cases, it seems that the app does not have a matching condition modelled.

⁶ - These were mental health scenarios where no conditions were provided at all

Classification of reason	App	Number of vignettes for which this was the reason for no condition-suggestion
Mental Health “If she is struggling with suicidal thoughts, please stay calm, don’t leave her alone and make sure to look for professional help right now...(Ada)” “I’m sorry you’re having thoughts of self-harm or suicide. I’m stopping your assessment as this is beyond my capability...(Babylon)” “Some of the symptoms you reported might need to be checked out by a GP within the next 6 hours...(Babylon)” “When experiencing thoughts of suicide, she should go immediately to the ER...(Buoy)” “Seeing or hearing things is a sign of a possible mental health issue. She should...(Buoy)” “Mental or medical condition causing hallucinations. (Buoy)” “Go to the hospital right away. It is very likely you need medical care. (Mediktor)” “Thank you for telling me, Guest. As you know, I am not a doctor, but there are people that can help you with these thoughts...(Your.MD)” “[Symptom] are worrying symptoms. You should see a doctor within 48 hours...(Your.MD)”	Ada	1
	Babylon	3
	Buoy	3
	Mediktor	4
	Symptomate	1
	Your.MD	2
Generic “Your symptoms require further medical assessment. (Babylon).” “This needs looking into...(Babylon)” “Range of possible causes...(Babylon)” “Symptoms can be managed at home...(Babylon - note condition not named)” “Ok guest, here is some information you might find helpful: [symptom/related condition information] (Your.MD)”	Babylon	12
	Your.MD	8
Emergency/ very severe “Call the local emergency number for a life-threatening medical emergency (Ada)” “Severe chest pain...[Symptom] might be a sign of something serious...(Buoy)” “Condition causing...[serious symptom] (Buoy)” “[Symptoms] are worrying symptoms. You should see a doctor within 48 hours (Your.MD)” “Symptoms should not be ignored (Babylon)” “Could be signs of something serious...(Babylon)”	Ada	1
	Buoy	18
	Your.MD	1
	Babylon	16

Clear statement of no results “No results (Symptomate)” “I wasn’t able to find any causes for your symptoms...(Your.MD)” “No prediagnosis was reached. Please retry the assessment describing it differently or check with a health professional (Mediktor)” “No conditions found, try adding more symptoms (WebMD)” “The symptoms you describe are quite complex...(Babylon)”	Symptomate	10
	Your.MD	11
	Mediktor	23
	WebMD)	3
	Babylon	1

Supplementary Table 10
Detailed analysis of why apps did not provide condition-suggestions for some vignettes

	GP mean			App mean			Ada			Babylon		
Vignette group	All (200)	NHS (64)	Non-NHS (136)	All (200)	NHS (64)	Non-NHS (136)	All (200)	NHS (64)	Non-NHS (136)	All (200)	NHS (64)	Non-NHS (136)
Condition-suggestion performance												
M1 (%)	71.2±5.6	62.7±7.8	77.7±2.8	26.1±8.9	23.2±6.4	27.5±10.4	48.5 (41.7-55.4)	39.1 (28.1-51.3)	52.9 (44.6-61.1)	21.5 (16.4-27.7)	18.8 (11.1-30.0)	22.8 (16.5-30.5)
M3 (%)	82±5.2	75.2±5.9	86.9±5.3	38.0±13.5	35.9±9.5	39.0±15.6	70.5 (63.8-76.4)	57.8 (45.6-69.1)	76.5 (68.7-82.8)	32.0 (25.9-38.8)	29.7 (19.9-41.8)	33.1 (25.7-41.4)
M5 (%)	82.8±4.7	76.6±5.5	87.4±5.0	40.8±15.2	39.6±12.6	41.3±16.8	77.5 (71.2-82.7)	67.2 (55.0-77.4)	82.4 (5.1-7.8)	32.5 (26.4-39.3)	29.7 (19.9-41.8)	33.8 (26.4-42.1)
COMP. (%)	45.9±5.0	48.0±6	44.9±4.9	26.1±9.1	26.5±9.7	25.9±9.0	44.0 (40.7-47.4)	43.5 (36.6-50.4)	44.2 (40.4-47.9)	18.7 (15.4-21.8)	15.1 (9.2-20.5)	20.3 (16.4-24.1)
RELE. (%)	69±9.0	64.9±6.9	70.3±11.2	36.2±7.5	32.5±6.8	37.9±7.9	45.9 (42.4-49.6)	38.4 (32.3-44.5)	49.5 (45.1-53.8)	28.3 (23.5-32.9)	22.4 (14.0-30.2)	31.0 (25.0-36.8)
Urgency-advice performance												
no advice given (%) [not recorded]	0.9±0.6	2.0±1.5	0.5±1.0	20.6±15.8	19.6±16.1	21.1±15.9	0.0	0.0	0.0	9.0	9.4	8.8
potentially unsafe (%)	2.9±2.5	3.3±3.4	2.4±1.9	7.6±5.8	10.5±6.1	6.2±5.7	3.0	6.3	1.5	4.5	7.8	2.9
safe but underconservative (%)	10.8±5.9	7.1±3.5	13.0±7.2	12.9±5.0	10.3±5.6	14.1±6.6	12.0	7.8	14.0	13.5	6.3	16.9
exactly matched (%)	57.6±8.0	48.2±11.3	63.5±8.3	25.9±10.2	24.6±7.9	26.6±11.5	33.0	32.8	33.1	40.0	34.4	42.6
overconservative but suitable (%)	15.2±6.3	17.6±7.1	14.6±10.1	16.3±7.2	13.8±4.7	17.4±8.7	24.5	20.3	26.5	16.5	14.1	17.6
overconservative (%)	12.6±7.7	21.0±10.4	5.4±7.7	16.7±8.3	21.2±9.5	14.6±8.5	27.5	32.8	25.0	16.5	28.1	11.0
% of safe advice	97.0±2.5	96.6±3.5	97.6±1.9	90.1±7.4	86.5±7.9	91.8±7.2	97.0	93.8	98.5	95.1	91.4	96.8

	Buoy			K Health			Mediktor			Symptomate		
Vignette group	All (200)	NHS (64)	Non-NHS (136)	All (200)	NHS (64)	Non-NHS (136)	All (200)	NHS (64)	Non-NHS (136)	All (200)	NHS (64)	Non-NHS (136)
Condition-suggestion performance												
M1 (%)	28.0 (22.2-34.6)	18.8 (11.1-30.0)	32.4 (25.1-40.6)	26.0 (20.4-32.5)	23.4 (14.7-35.1)	27.2 (20.4-35.2)	23.5 (18.2-29.8)	25.0 (16.0-36.8)	22.8 (16.5-30.5)	19.0 (14.2-25)	18.8 (11.1-30.0)	19.1 (13.4-26.5)
M3 (%)	43.0 (36.3-49.9)	35.9 (25.3-48.2)	46.3 (38.2-54.7)	36.0 (29.7-42.9)	39.1 (28.1-51.3)	34.6 (27.1-42.9)	36.0 (29.7-42.9)	39.1 (28.1-51.3)	34.6 (27.1-42.9)	27.5 (21.8-34.1)	28.1 (18.6-40.1)	27.2 (20.4-35.2)
M5 (%)	43.0 (36.3-49.9)	35.9 (25.3-48.2)	46.3 (38.2-54.7)	39.0 (32.5-45.9)	42.2 (30.9-54.4)	37.5 (29.8-45.9)	36.5 (30.1-43.4)	39.1 (28.1-51.3)	35.3 (27.8-43.6)	30.5 (24.5-37.2)	29.7 (19.9-41.8)	30.9 (23.7-39.1)
COMP. (%)	27.9 (24.5-31.1)	28.8 (22.5-35.1)	27.4 (23.6-31.1)	27.5 (24-31)	30.7 (23.5-37.5)	26.1 (22.1-30)	28.1 (24.3-31.7)	29.9 (22.9-36.7)	27.2 (22.7-31.6)	16.3 (13.3-19.2)	15.9 (10.4-20.9)	16.5 (12.9-20.0)
RELE. (%)	47.5 (42.3-52.7)	42.7 (33.6-51.6)	49.8 (43.4-56.2)	38.6 (33.7-43.6)	35.5 (27.1-43.7)	40.1 (33.7-46.2)	33.3 (28.9-37.5)	29.6 (22.6-36.3)	35 (29.4-40.3)	28.7 (23.6-33.8)	26.9 (17.8-35.5)	29.6 (23-35.7)
Urgency-advice performance												
no advice given (%) [not recorded]	10.0	10.9	9.6	44.0	46.9	42.6	17.5	12.5	19.9	31.5	32.8	30.9
potentially unsafe (%)	18.0	21.9	16.2	10.5	12.5	9.6	10.5	14.1	8.8	1.5	4.7	0.0
safe but underconservative (%)	20.0	12.5	23.5	18.0	12.5	20.6	12.0	18.8	8.8	9.5	12.5	8.1
exactly matched (%)	34.0	28.1	36.8	17.0	15.6	17.6	24.0	21.9	25.0	11.5	14.1	10.3
overconservative but suitable (%)	8.5	10.9	7.4	6.0	6.3	5.9	23.0	18.8	25.0	21.5	12.5	25.7
overconservative (%)	9.5	15.6	6.6	4.5	6.3	3.7	13.0	14.1	12.5	24.5	23.4	25.0
% of safe advice	80.0	75.4	82.1	81.3	76.5	83.3	87.3	83.9	89.0	97.8	93.0	100.0

	WebMD			Your.MD		
Vignette group	All (200)	NHS (64)	Non-NHS (136)	All (200)	NHS (64)	Non-NHS (136)
Condition-suggestion performance						
M1 (%)	21.0 (15.9-27.2)	20.3 (12.3-31.7)	21.3 (15.3-28.9)	21.5 (16.4-27.7)	21.9 (13.5-33.4)	21.3 (15.3-28.9)
M3 (%)	35.5 (29.2-42.3)	32.8 (22.6-45)	36.8 (29.1-45.1)	23.5 (18.2-29.8)	25.0 (16.0-36.8)	22.8 (16.5-30.5)
M5 (%)	43.5 (36.8-50.4)	48.4 (36.6-60.4)	41.2 (33.3-49.6)	23.5 (18.2-29.8)	25.0 (16.0-36.8)	22.8 (16.5-30.5)
COMP. (%)	32.2 (28.4-35.8)	33.2 (26.7-39.7)	31.7 (27.2-36.1)	13.9 (11.5-16.2)	14.9 (10.7-19.1)	13.4 (10.5-16.1)
RELE. (%)	27.3 (24.0-30.5)	25.7 (20.3-30.9)	28.1 (23.9-32.1)	39.7 (33.3-45.9)	38.7 (27.7-49.5)	40.1 (32.5-47.8)
Urgency-advice performance						
no advice given (%) [not recorded]	n.d.	n.d.	n.d.	32.5	25.0	36.0
potentially unsafe (%)	n.d.	n.d.	n.d.	5.0	6.3	4.4
safe but underconservative (%)	n.d.	n.d.	n.d.	5.0	1.6	6.6
exactly matched (%)	n.d.	n.d.	n.d.	22.0	25.0	20.6
overconservative but suitable (%)	n.d.	n.d.	n.d.	14.0	14.1	14.0
overconservative (%)	n.d.	n.d.	n.d.	21.5	28.1	18.4
% of safe advice	n.d.	n.d.	n.d.	92.6	91.7	93.1

Supplementary Table 11

Condition-suggestion and urgency-advice performance for all apps for the complete set of 200 vignettes, and subanalyses for the 32.0% of vignettes sourced from NHS Direct and the 68.0% of vignettes created by the vignette creation team (95% confidence intervals are shown in brackets). The mean app and mean GP performance (\pm S.D.) are also shown for each set of vignettes. n.d. – not determined (this applies to WebMD which was not included in the urgency-advice comparison).

Appendix S1:

Comparison of the vignettes to UK health statistics

We explored the demographic properties and frequency of health conditions of the vignettes to investigate how our results would generalize beyond the artificially generated vignette patient, population and spectrum of conditions (see main manuscript section on vignette creation for details). The age and sex distribution of the vignettes was compared to the latest available data UK population [1]. The age and sex distribution of the chosen vignettes corresponds broadly to that of the UK population (see **Supplementary Fig. 2**). Vignettes with young adults as subjects are proportionately more frequent than young adults in the UK population, but those groups constitute the current main user group for digital symptom assessment tools and there is therefore good justification for their frequent inclusion in vignettes. Vignettes with young children as subjects are also proportionately more frequent than young children in the UK population, but young children are a high health dependency population, where the use of a symptom assessment application by the caregiver for advice is a highly relevant use case, and a use case in which condition suggestion and urgency advice accuracy is highly important. Over-representation of this demographic in the vignettes is therefore justified. Importantly, the vignettes sample the whole range of ages found in the UK population.

To compare the frequency of health conditions found in the vignettes to that in the English population, we compared the relative frequency of condition groups of the vignettes to the prevalence of condition groups reported by Quality and Outcomes Framework (QOF) for the NHS in England (Quality and Outcomes Framework) [2]. The QOF reports the prevalence of conditions belonging to six groups: (i) cardiovascular conditions (atrial fibrillation, cardiovascular disease primary prevention, coronary heart disease, heart failure, hypertension, peripheral arterial disease, as well as stroke and transient ischaemic attack); (ii) respiratory conditions (asthma, COPD); (iii) lifestyle related conditions (obesity, smoking); (iv) high dependency and long term care conditions (cancer, chronic kidney disease, diabetes mellitus, palliative care); (v) mental health and neurological conditions (dementia, depression, epilepsy, learning disabilities, mental health); and, (vi) musculoskeletal conditions (osteoporosis, rheumatoid arthritis). The condition

classification system adopted by the QOF differs substantially to the classification system used to subdivide the vignettes (see **Supplementary Table 1**), and some of the QOF groups were not explicitly included in specific vignettes as they overlap other QOF groups e.g. lifestyle related conditions such as obesity and smoking related conditions were not represented by individual vignette main diagnoses, but these were relevant to the clinical history provided in some vignettes. Therefore, in order to compare the condition coverage of the vignettes to the QOF data, we adapted the condition groups of the QOF and matched them to the vignette condition groups (as described in **Supplementary Table 1**). The distribution of the condition categories among the vignettes shows the vignette creation goal, of inclusion of a diverse range of conditions in the vignettes spanning the conditions seen by GPs, was achieved (see **Supplementary Fig. 3**). With the exception of cardiovascular conditions, the proportion of condition categories found among the vignettes and among the English NHS population are broadly similar. Although there is a mismatch between the relative frequency of cardiovascular conditions in the vignettes and in the English NHS population, it should be noted that there are a broad range of conditions included this QOF category (e.g. hypertension, and primary prevention for cardiovascular disease). It should also be noted that the proportions reported by the QOF are not necessarily comparable to an *ideal* proportion of vignettes - the QOF reports condition prevalence based on all patients who fulfil the requirements for the condition list in the respective year of the report, whereas new occurrences of a disease are more likely to lead to the use of a digital symptom assessment application. More generally, population morbidity data collected at doctors' practices based on confirmed diagnoses will always lack the unknown proportion of un- or misdiagnosed conditions which may be picked up by digital symptom assessment applications. It is therefore not necessarily optimal to their intended evaluation goal, for vignettes used to assess symptom assessment applications to have a perfect match in condition frequency to the QOF prevalence data.

References

1. Overview of the UK population—Office for National Statistics. (2019). Retrieved October 15, 2020, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/august2019>.
2. Quality and Outcomes Framework. (2019-20). NHS Digital. Retrieved October 15, 2020, from <https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data/2019-20>.

Supplementary Figures

Supplementary Figure 1

Mapping between the native levels of urgency advice provided by each app and tested-GP and the study defined common levels of urgency advice (as shown in main manuscript **Table 1**).

Supplementary Figure 2

Distribution of age and sex in the study vignettes (bars) compared to the demography of the UK population (shaded area), expressed as proportion of the vignettes or the total population, respectively.

Supplementary Figure 3

Proportion of cases per condition category reported in NHS prevalence data versus vignette conditions.