

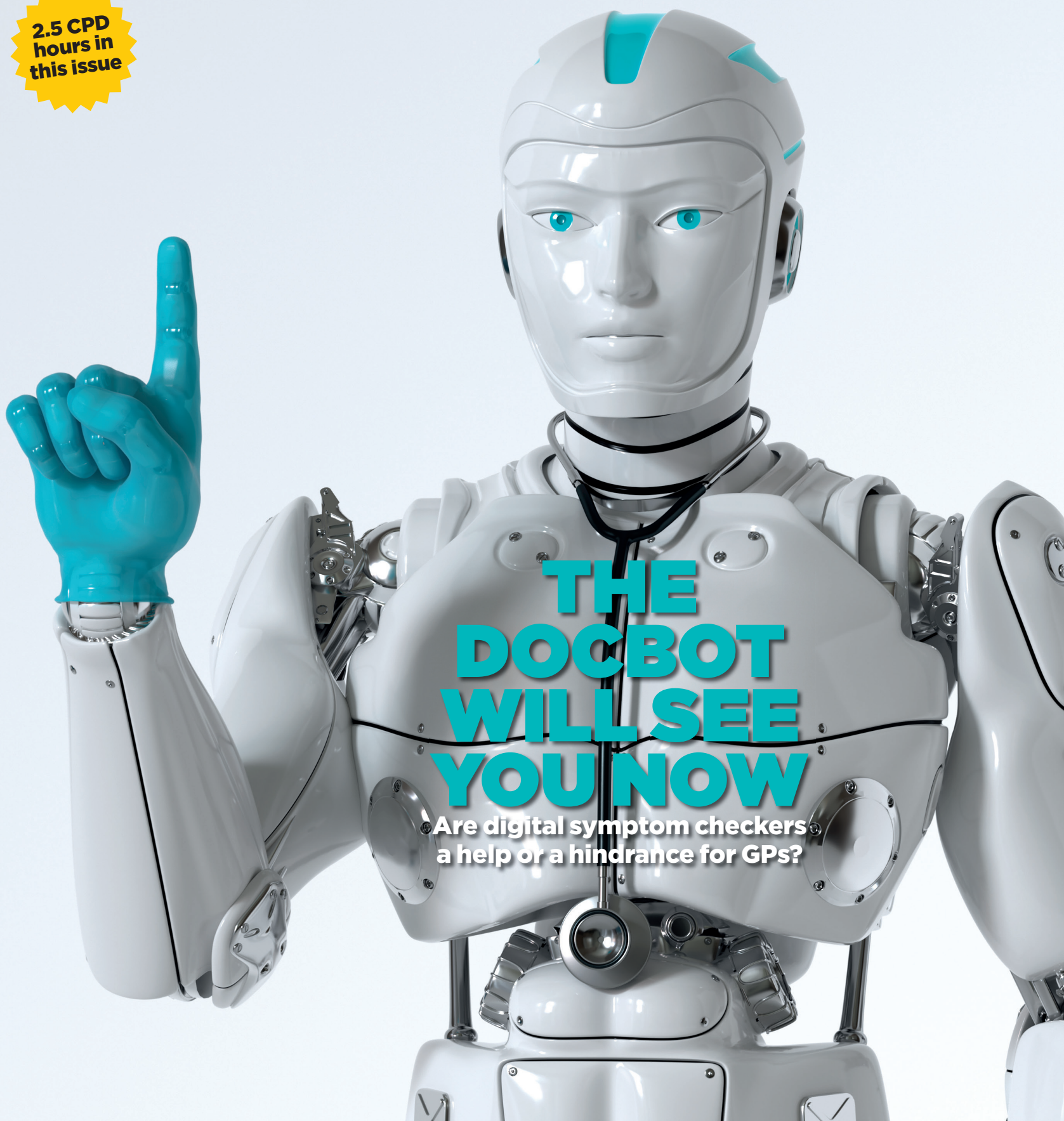
# PULSE

Supporting GPs since 1960  
September 2019 | [www.pulsetoday.co.uk](http://www.pulsetoday.co.uk)



**Pulse Power 50**  
Love it or hate it,  
the annual list of  
most influential  
GPs ► **page 24**

**2.5 CPD  
hours in  
this issue**



## THE DOCBOT WILL SEE YOU NOW

Are digital symptom checkers  
a help or a hindrance for GPs?



# The docbot will see you now

The NHS, the health secretary and a number of private companies are promoting AI as the gatekeeper of the future. *Léa Legraïen* and *Emma Wilkinson* ask whether we are putting too much faith in tech

**A**lexa, what are the early signs of a stroke?’ GPs may no longer be the first port of call for patients looking to understand their ailments. ‘Dr Google’ is already well established in patients’ minds, and now they have a host of apps using artificial intelligence (AI), allowing them to input symptoms and receive a suggested diagnosis or advice without the need for human interaction.

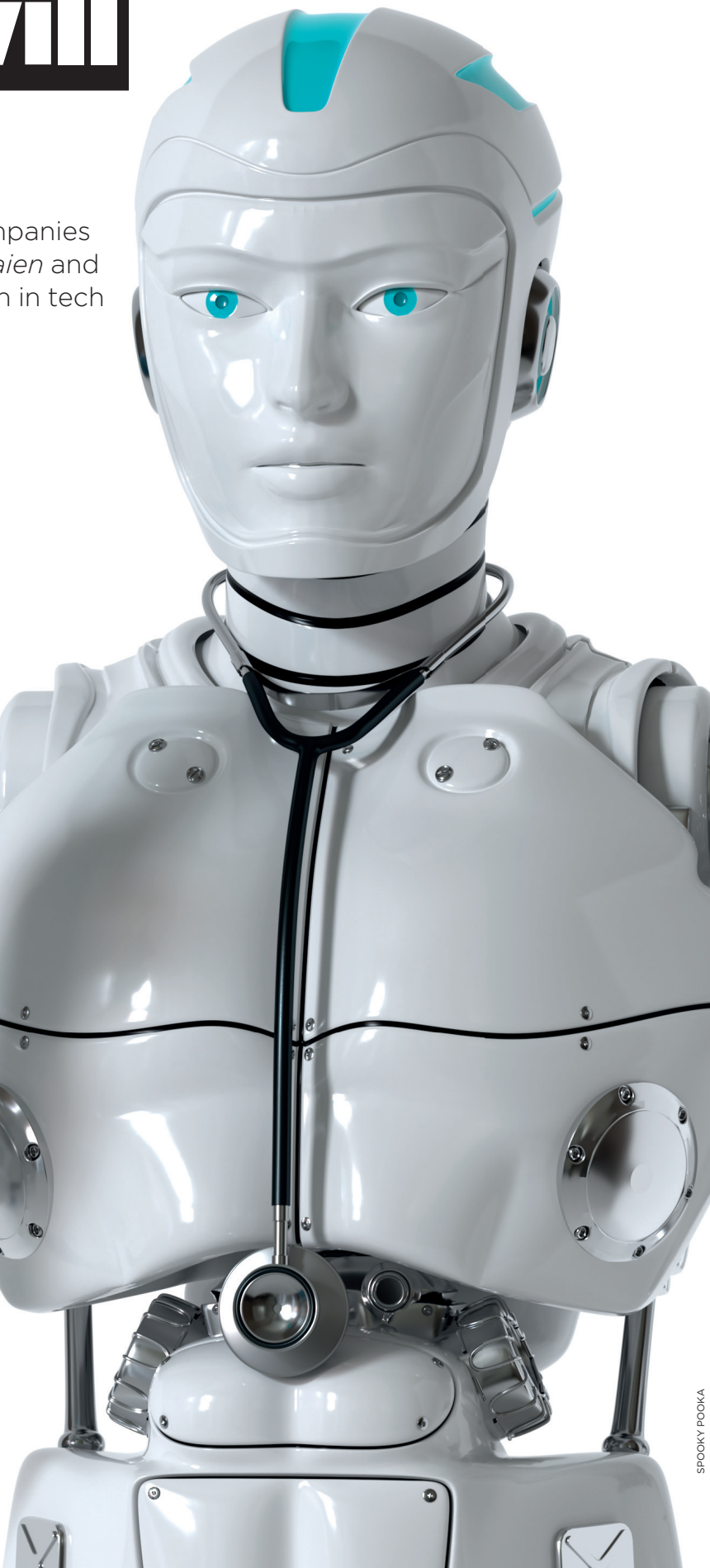
And policymakers are on board. Matt Hancock is the most tech-friendly health secretary ever, NHS England chief executive Simon Stevens wants England to lead the world in AI, and the prime minister last month announced £250m for a national AI lab to help cut waiting times and detect diseases earlier. Amazon even agreed a partnership with NHS England in July to allow people to access health information via its voice-activated assistant Alexa.

Little surprise then that private developers see now as a good time to develop AI to guide patients through their various ailments. Babylon last month announced a £450m R&D investment, partly for AI technology to manage chronic conditions, while the likes of Ada and Your.MD also offer patients the chance to check symptoms. This is on top of the NHS App’s own symptom checker.

Yet the evidence in support of algorithms – and AI – is still lacking, and a Pulse analysis has shown potential drawbacks, such as overreaction to mild conditions and potentially unsafe advice.

Dr Rebecca Fisher, a GP and senior policy fellow at the Health Foundation says: ‘If patients are using symptom-checker apps, I would have two main worries. The first is that the app will give

a false positive, with the risk that the patient becomes anxious and also potentially unnecessary use of NHS resources. Even worse, there a risk of an app giving a false negative, meaning you



SPOOKY POOKA

## Taking a journey around the symptoms checkers

There are a range of apps out that allow patients to check their symptoms, and receive advice on what action to take and, for some apps, possible diagnoses. Pulse put these to the test to see whether they were offering appropriate advice.





### Method

We asked four experienced GPs to each test out four symptoms checkers. They were given instructions to go through the eight different scenarios listed below for each of the apps (ie, anxious 50yo male with a cough with bruised rib).

The GPs were told to input the following when asked for the initial symptoms: chest pain in the case of the cough with bruised rib;

toothache (dental abscess); rash (shingles); back pain (acute pyelonephritis). After doing so, they were asked to take the route that follows for the patients based on what they would say with their condition and their levels of concern: ‘anxious’ patients would fear/assume the worst with potential to maximise symptoms; ‘non-anxious’ patients were the opposite. There was an element of subjectivity following the initial input of symptoms, as there is with patients.

We then asked them to record the most urgent advice given in each scenario – the thinking being the patient would follow the most urgent advice to play safe. The advice varied among the four GP testers. With this in mind, we picked out the most common advice below and, where given, potential diagnoses.

	 Available to all patients, the symptoms checker is headed ‘111 online’. It says the patient can ‘find out when and where to get help’, and ‘be contacted by a nurse, if needed’.	 ‘Chat bot’ – It says it ‘provides health information only’ and not a diagnosis. It says the information provided ‘is based on risk factors and statistics, rather than a personalised assessment. The AI services are not a substitute for a doctor’.	 It says it ‘asks simple, relevant questions and compares your answers with thousands of similar cases to help you find possible explanations for your symptoms’.	 It offers patients the chance to ‘use our symptom checker to see what you might be suffering from and get guidance on what to do next’.
Anxious 50yo male with a cough with bruised rib	Speak to GP today	Phone 999	See GP that day – broken rib suggested	Call ambulance
Non-anxious 50yo male with a cough with bruised rib	Speak to GP today	Call ambulance/see GP urgently	Not urgent – common cold suggested – see GP/manage at home	Doctor review/should get better
Anxious 20yo female with a dental abscess	See your dentist today	Dental assessment	See doctor straight away – dental abscess suggested	See doctor – tooth decay suggested
Non-anxious 20yo female with a dental abscess	See your dentist today	Dental assessment	See doctor straight away – dental abscess suggested	See GP within 48 hours – tooth decay suggested
Anxious 75yo female with shingles	Speak to GP urgently	Speak to GP in next few hours	Seek medical advice straight away – shingles suggested	Meningitis – call ambulance
Non-anxious 75yo female with shingles	Speak to GP within 24 hours	Speak to GP in next few hours	Seek medical advice straight away – shingles suggested	Meningitis – call ambulance
Anxious 26yo female with acute pyelonephritis	Speak to GP urgently	Call 999/speak with doctor	Emergency care - acute pylonephritis /kidney stones suggested	Emergency response to be safe, alternatively cystitis or back sprain
Non-anxious 26yo female with acute pyelonephritis	Speak to GP urgently	Speak to GP in next few hours	Emergency care - acute pylonephritis /kidney stones suggested	Emergency response to be safe, alternatively cystitis or back sprain
What the providers say (The full responses can be found in the online version of this article)	“ 111 Online is designed to direct patients to the right care provider within the right timeframe, it is not designed to provide a diagnosis. 111 Online algorithms follow a robust process of clinical development to ensure that they follow the latest clinical evidence, and are reviewed by the independent National Clinical Governance Group, chaired by the RCGP.	“ Our symptom checker is still fairly new - it has been running for a couple of years (and has never missed a serious case) but its strength is we update it every two weeks and it never forgets or regresses. We’re also learning from how people use it. For example, we have found that if it reports cancer, patients are less likely to get checked, but if it says ‘potentially something serious’ they are far likelier to speak to their GP.	“ Over 60 expert doctors input into Ada’s platform and rigorously test outcomes against gold-standard diagnosis. We’re happy the tests confirmed almost 90% accuracy. For advice levels, Ada is able to take into account the full symptom picture, differential results and other red flags, as a good clinician would. It is designed to be applied alongside healthcare professionals as part of the full care pathway.	“ We believe user safety to be paramount for symptom checkers and made it top priority in our design. The [overall] outcomes of the tests conducted by Pulse were all safe even if they were not all accurate. We are confident we have a service that besides being safe can provide enormous value to the community as it can help people decide whether to see a doctor, with greatly positive impact on the health service and themselves.

## What our testers said

- The apps were successful at spotting serious conditions, such as a heart attack. They worked quickly, and were easy to use.
- However, our ‘anxious’ patients were told by some to call an ambulance for coughing with a bruised rib, while most patients with shingles were told to seek medical help within a few hours.
- Ada was the most accurate ‘diagnostic’ app, but while it correctly diagnosed a dental abscess, it advised seeing a doctor – possibly a failure to localise it to the UK. Your.MD made a similar suggestion.
- Your.MD suggested meningitis in many shingles scenarios, which would cause anxiety. For one tester, it suggested an anxious patient with acute pyelonephritis did not need to take action. For the other testers, it suggested cystitis or back sprain but advised they go to A&E just in case. Your.MD said it will review lower and upper UTIs after Pulse’s feedback.
- Most of the apps in all scenarios bar the dental one advised seeing a GP urgently or go to A&E/call 999. Many of the suggestions would increase patient anxiety, which would be detrimental in the long term.



## Is there evidence to support AI in healthcare?

● **A 2013 study by the University of Sheffield<sup>1</sup>** revealed that NHS 111 increases ambulance and urgent and emergency care use. It looked at 400,000 calls, including 277,163 triaged using NHS Pathways, and found emergency ambulance incidents rose by 2.9%. It estimated this could mean an additional 14,500 call-outs for a service attending 500,000 incidents a year. In addition, emergency and urgent care activity rose by between 5-12% per month.

The study concluded: 'The findings reflect the inherent characteristics of the NHS Pathways system such as the levels of caution and risk built into the assessment algorithms, particularly as it is designed to be used by non-clinical call handlers. There may be less flexibility to change decisions compared with assessments made by nurses and it is possible that a different call assessment system could produce different results.'

● **A 2015 evaluation by Harvard Medical School<sup>2</sup>** found 23 symptom checkers for self-diagnosis provided the correct diagnosis first in 34% of 45 standardised patient evaluations, listed the correct diagnosis within the top 20 diagnoses given in 58% and provided appropriate triage advice in

57% of cases. It said: 'Overall they had deficits in both diagnosis and triage accuracy. The risk-averse nature of symptom checkers' triage advice is a concern. In two-thirds of evaluations where medical attention was not necessary, we found symptom checkers encouraged care.'

● **A 2017 evaluation by NHS England<sup>4</sup>** found patients had a very good experience of triage and assessment tools including the digital version of NHS Pathways in West Yorkshire (web interface), Sense.ly system in West Midlands (voice-activated avatar), Expert 24 in Suffolk (web interface) and Babylon in London. As a result of their use, fewer people were directed to primary care services and more turned to self-management than from NHS 111.

● **A 2018 study by Babylon<sup>5</sup>** showed the company's triage and diagnostic system was able to identify patient conditions modelled by a clinical vignette with accuracy comparable with doctors', in terms of precision and recall, and was on average safer than doctors. The findings, based on the MRCGP examination, showed above-average pass marks. Yet the paper was not peer reviewed, and the research team included Babylon employees.

might not seek help you actually need.'

Dr Nick Mann, a London-based GP with an interest in AI, says he is already seeing this sort of impact: 'People will come in with headache and be convinced they've had a brain bleed whereas I know, talking to them, they haven't.'

'I've had a lot of requests in the past couple of years, which I never used to have, from people wanting investigations for symptoms they have diagnosed on Google, which are inappropriate.'

With this in mind, Pulse tested some of the available symptoms checkers. We found the apps were successful in offering appropriate advice in the case of a heart attack (see table, page 7), but problems also emerged. In one case, a 26-year-old female with acute pyelonephritis was told her condition would clear up on its own.

Dr Roger Henderson, a sessional GP who is also medical director of Liva Healthcare, a digital healthcare company that supports the management of patients with diabetes and who tested the apps for Pulse, says: 'In this tiny snapshot there are worrying features where everyday complaints were marked as emergencies and potentially severe ones were underplayed.'

'Symptom checkers use a linear algorithm approach and depend on the information provided to them, rather

than being able to follow the more nuanced process that GPs use. It is this black-and-white computer reasoning that causes problems, since diagnosis tends to be shades of grey in the real world.'

He says the fact that symptoms checkers encourage people to include all symptoms to give the fullest possible picture can lead to anxiety: 'If you give a patient a range of diagnoses ranging from minor to very serious, it is natural to focus on the serious even if this is incorrect, causing worry and anxiety.'

Lincolnshire GP Dr Phillip Williams, who also tested the apps for Pulse, agrees patients don't always present as textbook cases. 'Often real patients don't present with the symptoms we think they should. As these apps become more sophisticated, they may flag key symptoms which aren't on our radar. For example, we're taught motor neurone disease presents with fasciculations, whereas, in real life, a common first symptom is fatigue.'

The shortage of relevant research is a problem for many GPs (see box, above). Dr Benjamin Brown, a senior academic GP and health informatician in Manchester, says: 'The NHS should only bring in routine care systems that have an evidence base. In the case of model-driven triage, the models may be too conservative. I have anecdotally heard

that one of the well-known providers modified its algorithms over concerns about patient safety, which resulted in it sending many more patients to A&E.'

Perhaps the highest-profile patient-facing algorithm is NHS Pathways, used by NHS 111. A 2013 study found NHS 111 increased emergency and urgent care activity by 5-12% each month, while emergency ambulance incidents rose by 2.9%.<sup>1</sup> NHS England has introduced more clinicians into the call centres but, according to 616 GPs surveyed by Pulse, an average GP still receives around six inappropriate referrals from NHS 111 a month – totalling more than three million a year. Anecdotally, GPs say they are still seeing patients referred to them for dental problems. And last month, a coroner said the lack of flexibility within the algorithm should be addressed following the death of a 17-year-old boy, whom the coroner said may not have understood what he was being asked.

Harry Longman, founder of Askmygp – an online triage and consultation tool for GPs – says: 'We don't use any AI or algorithms to triage automatically, we have tried that and found it doesn't work. Many questions were irrelevant or difficult for patients, and the resulting output was not that helpful for clinicians.'

The Medicines and Healthcare products Regulation Agency says if an app is intended to influence treatment or results in a diagnosis or prognosis including future disease risk then it is a device and should obtain a CE mark before use. New EU rules, taking effect next year, will introduce more stringent requirements for device manufacturers.

But, as Professor Brendan Delaney, chair in medical informatics and decision-making at Imperial College London, puts it: 'The letter of the regulation is fine, but it relies on developers to self-certificate and register - which is OK, provided entry to the market place is actually policed and purchasers insist on CE marking.'

There are positives. AI is being developed to help target patients for screening, and help doctors make decisions – uses few would argue with.

And the Topol Review<sup>2</sup> commissioned by Mr Hancock to explore how the healthcare workforce will 'deliver the digital future', concluded that 'early benefits of AI and robotics will include the automation of mundane repetitive tasks that require little human cognitive power, improved robot-assisted surgery and the optimisation of logistics.' This would allow the workforce to focus on 'interaction and care'.

However, the first signs are that AI will, at best, increase GP workload. It might be time for the Mr Hancock to review his championing of this new technology.

“

**My main concerns would be about apps giving false positives or negatives**  
Dr Rebecca Fisher



”

### References

- 1 Turner J et al. Impact of the urgent care telephone service NHS 111 pilot sites: a controlled before and after study. *BMJ Open* 2013;3:e003451
- 2 Topol E. Preparing the healthcare workforce to deliver the digital future. Health Education England. February, 2019. [topol.hee.nhs.uk/](http://topol.hee.nhs.uk/)
- 3 Semigran H. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015;351:h3480
- 4 NHS England online evaluation, December 2017.
- 5 Razzaki S et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. June, 2018