



## VIEWS AND REVIEWS

### NO HOLDS BARRED

# Margaret McCartney: AI in medicine must be rigorously tested

Margaret McCartney *general practitioner*

Glasgow

Artificial intelligence (AI) could be a great thing in medicine. It could make healthcare safer and faster. It could make medicine more satisfying to practise and less unpleasant to receive.

But we must test a hypothesis before we roll it out to the public. The huge datasets being collected to use AI in diagnostic radiography, for example, are indicative of the amount and quality of work needed to make reliable, safe tools. AI must be subjected to the same criteria and testing we'd expect with other forms of evidence based medicine.

NHS 24 in Scotland now has a symptom checker app. While there's no published evaluation, it's been based on NHS 111 algorithms, has undergone user testing, and remains open to improvements (on raising a concern about the fact that a symptom of "fever" is managed without recourse to asking about other symptoms, I was told that this would be promptly reviewed). Such positive feedback loops are welcomed and should be normal.

Babylon, the healthcare company that offers private and NHS GP services, has created an "NHS 111 powered by Babylon" app. It's currently being piloted in north London, as I've discussed before,<sup>1</sup> and has been extended to offer a "new paediatric symptom checker for parents."

Babies are often terrifying. They become ill quickly. They get well quickly. One small thing—a rash, a temperature, a rapid heart rate—can tip the bayesian scales, requiring a blue light ambulance. The paediatric app version isn't currently available on the NHS, where under 17s are told to use the NHS 111 phone line, but it is available through the Babylon NHS and private service.

Who's in charge of ensuring that this app is safe and fit for purpose?

Knowing the staggering lack of publicly available robust testing that had accompanied the adult symptom checker app, I thought that perhaps Babylon might have done better with its paediatric one. What's Babylon's evidence? I don't know, for it replied with, "we won't be responding to your enquiry." The binary nature of the chatbot means that one thing that doesn't happen

is history taking, in the medical sense ("Shut up, your patient is telling you the diagnosis"). It has a series of yes/no questions and short multiple choices.

Who's in charge of ensuring that this app is safe and fit for purpose? The Medicines and Healthcare Products Regulatory Agency (MHRA) has said that it will ask Babylon to change the way it refers to the app as being "certified as a medical device with the MHRA." The MHRA says that, for class I devices such as this app, the manufacturer must register with the agency and self certify that the device meets the requirements of the regulations. The MHRA says that this process is purely administrative—the MHRA takes details of the types of devices manufactured, but it does not assess, certify, approve, or accredit devices as part of the CE (European Conformity) marking process.

Who else could act? The Care Quality Commission has inspected Babylon, but it made no mention of the reliability, or not, of the app that it uses to direct people to and from general practice consultations.<sup>2</sup> The General Medical Council regulates individual doctors, not clinical devices.

We have many regulators but little proactivity, even for an app which—despite the small print warning us that it "does not constitute medical advice, diagnosis, or treatment"—is being used as the front door into NHS care.

AI has great potential in healthcare. But this potential will not be realised, and harm may be caused, if we don't accept the need for robust testing before it's publicly launched and widely used. We have no clear regulator, no clear trial process, and no clear accountability trail. What could possibly go wrong?

Competing interests: See [www.bmj.com/about-bmj/freelance-contributors/margaret-mccartney](http://www.bmj.com/about-bmj/freelance-contributors/margaret-mccartney).

Provenance and peer review: Commissioned; not externally peer reviewed.

- 1 McCartney M. Margaret McCartney: Innovation without sufficient evidence is a disservice to all. *BMJ* 2017;358:j3980. doi:10.1136/bmj.j3980
- 2 Care Quality Commission. Babylon Healthcare Services Ltd: inspection report. 8 Dec 2017. [www.cqc.org.uk/sites/default/files/new\\_reports/AAAG6526.pdf](http://www.cqc.org.uk/sites/default/files/new_reports/AAAG6526.pdf).

Published by the BMJ Publishing Group Limited. For permission to use (where not already granted under a licence) please go to <http://group.bmj.com/group/rights-licensing/> permissions