VIEWS AND REVIEWS

NO HOLDS BARRED

Margaret McCartney: Innovation without sufficient evidence is a disservice to all

Margaret McCartney *general practitioner*

Glasgow

Breeding with the vigour of rabbits, a variety of private GP services are springing up and jostling for your attention. Why spend hours phoning your own GP when you can choose a private GP by looks and price, having immediate access online or by app without needing to leave your home? Recruitment for GPs in such companies is increasingly competitive, with advertisements promising hourly rates and perks unfamiliar to NHS employees.

Some of these companies have bigger ambitions. "Why couldn't Babylon be a patient's NHS GP?" asks Ali Parsa, CEO of Babylon—an IT company whose app, he says, is faster and more accurate than doctors in risk assessing cases.[1] The NHS in north London has gone further, commissioning his company to offer "symptom advice" through a smartphone app service—"NHS 111 powered by Babylon"[2]—although the NHS 111 telephone service still exists.

Babylon has several offerings intersecting medicine and IT. It provides private GP services by video link and has corporate contracts with Bupa, Sky, and Boots. Earlier this year the Advertising Standards Authority, after I'd contacted it, told Babylon to stop saying that it had the "world's best doctors" and the "world's most advanced AI [artificial intelligence]."

Good IT is something the NHS has struggled for years to provide. It's clear that NHS 111 has been beset by problems.[3] But how do we put better technology into the NHS? If we want to hand over clinical triage to smartphone apps we need to know that they work to a standard we can set and test independently. That should mean robust and independent trials. But we don't seem to have them.

> If we hand over clinical triage to smartphone apps they need to work to a standard we can set and test independently

The initial information sheet that Babylon co-wrote with the NHS contained a set of "frequently asked questions."[4] Under "Is it safe?" the answer was, "An independent study tested the app's symptom checker against nurses and junior doctors. It found that the app gave safe advice 100% of the time, whereas doctors gave safe advice 98% of the time . . . it also found that the app is more accurate than doctors or nurses."[4]

But this study was not "independent": it had six authors,[5] and five are current or past Babylon employees,[6-10] while the sixth is its owner (Parsa). The study itself—not indexed in PubMed—was not a real life trial of how humans use the app but a simulation using actors and invented scenarios. I would not regard it as a trial but a description of a development process, omitting essential details about the clinical scenarios tested. I can't find any other publicly available testing of this system. Although this claim on safety has now been removed, a pilot project such as this is, in my view, not adequate.

Babylon says that its partnership with NHS 111 is itself an independent pilot programme set up to evaluate the technology. "Before the pilot commenced, the local NHS conducted their own validation of the Babylon triage service against all their serious incidents over a recorded period, and found the service to be completely safe," the company says. Babylon also says that its symptom checking products have been shown to work well for UK patients, relieving pressure on healthcare systems.

It's not that I don't think technology has potential. It does. But we need high quality evidence, which should mean high quality trials. Is an automated app better than NHS 111, or are humans needed to over-ride problems in automation? When we drop the threshold to consultation we change the demographic such that false positives easily become more common, potentially leading to unnecessary emergency department or GP attendance—which, in turn, makes it harder for sick people to get attention. If this happens we will be creating more dilemmas, not solving them. Innovation without sufficient evidence is a disservice to all.

margaret@margaretmccartney.com

1 Donnolly L. Forget your GP, robots will "soon be able to diagnose more accurately than almost any doctor." *Daily Telegraph* 7 March 2017. www.telegraph.co.uk/technology/2017/03/07/robots-will-soon-able-diagnose-accurately-almost-doctor/.

2 Enfield Clinical Commissioning Group. A million Londoners to test new NHS 111 app. 6 Jan 2017. www.enfieldccg.nhs.uk/news/A-million-Londoners-to-test-new-NHS-111-app.htm.

3 Turner J, O'Cathain A, Knowles E, Nicholl J. Impact of the urgent care telephone service NHS 111 pilot sites: a controlled before and after study. *BMJ Open* 2013;358:e003451. doi:10.1136/bmjopen-2013-003451. pmid:24231457.

4 Haringey Clinical Commissioning Group. Frequently asked questions: NHS 111 online powered by babylon. 14 Feb 2017. www.lcwucc.com/LinkClick.aspx?fileticket=4loVHf26ndM%3D&tabid=77&mid=1589.

5 Middleton K, Butt M, Hammerla N, Hamblin S, Mehta K, Parsa A. Sorting out symptoms: design and evaluation of the "babylon check" automated triage system. https://arxiv.org/pdf/1606.02041.pdf.

6 Middleton K. Linkedin. https://uk.linkedin.com/in/katherine-middleton-54366aa4.

7 Butt M. Linkedin. https://uk.linkedin.com/in/mobasher-butt-77505b90.

8 Hammerla N. Linkedin. https://uk.linkedin.com/in/nhammerla.

9 Hamblin S. Remote.com. https://remote.com/stevenhamblin?ref=brandedme.

10 Mehta K. Linkedin. https://uk.linkedin.com/in/dr-karan-mehta-73272277.