©Stock.Adobe.com

Talking responsibly about **medicine** in the Fourth Industrial Revolution

By Alex Broadbent

A lot of what is currently being said about the future of medicine in the fourth industrial revolution (4IR) is irresponsible: it appears to be uttered without regard for whether it is true or false. The philosopher Harry Frankfurt argues that we should define and use the word "bullshit" as a technical term to cover speech of this sort (Frankfurt, 2005). In order to respect the reader's potential sensibilities, I will not follow his tempting suggestion, although I do in my recent book (Broadbent, 2019). (The interested reader may like to consult either Frankfurt's or my work and decide for his or her self whether the term is applicable in the present context). Regardless, irresponsible speech is something that medicine itself has been accused of, in the form of quackery, charlatanism, and so forth. David Wootton argues that, in effect, quackery was universal in the past (Wootton, 2006). He suggests that the entirety of medicine prior to 1865 was "bad", and suggests that doctors often knew it, or else simply didn't care, and thus knowingly or at least carelessly gave false hope for personal profit. I've defended medicine against this charge, notwithstanding its patchy track record and the existence of real quacks (Broadbent, 2018b, 2018a, 2019). Nonetheless, the high hopes we have for medicine and the difficulty of assessing its effectiveness render it an easy victim of negligent

talk, and I'm depressed to see this happening in some of the contemporary discourse about the 4IR.

It is sometimes acceptable to speak without regard for truth or falsity: to embellish for purposes of entertainment, as in after dinner speech, or to pass the time, as in the idle banter in the Uber, or when everyone knows what is going on and nothing hangs on it, as in the dean's word of welcome at a university event. But it's problematic when people might take it seriously, and especially so when it concerns medicine, about which people rarely doctor the truth for a joke. I want to discuss two categories of pronouncement on the future of medicine. One is almost always irresponsible, and I'll not hesitate to say so. This category comprises bold general pronouncements about the future of medicine. The other category is considered and careful; all the same, claims in this category are not true, and there are systematic mistakes that need to be highlighted. This category comprises serious research which nonetheless ignores limits on what one can learn from patterns in data alone, independent of what the data is about.

General Pronouncements on the Future of Medicine

General pronouncements on the future of anything are always to be treated with skepticism. That's not to say they can't be true. It's to say that we should be doubtful of their truth. That's because such claims have a terrible track record, and are subject to all kinds of well-documented biases, including financial bias (or bias for other personal gain—ego fulfilment, or whatever), confirmation bias, the base-rate fallacy (Kahneman, 2011; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1973, 1982), and plain old over-excitement. We are not driving flying cars or living on Mars or speaking to personal robot assistants—prospects that already seemed within reach in the 1960s.

How different is our epistemic position now from what it was in the 1960s? Our technology has advanced, but I've yet to see good evidence that our powers of prediction have kept pace. In fact, the last fifteen years have seen spectacularly unpredicted events: 9/11 in 2001, the crash of 2008, and in 2016 the doubly pundit-slamming Brexit vote and Trump victory. Our predictive inadequacies as humans were being studied as

> 66 Nonetheless, the high hopes we have for medicine and the difficulty of assessing its effectiveness render it an easy victim of negligent talk, and I'm depressed to see this happening in some of the contemporary discourse about the 4IR. 99

far back as 1973 by Amos Tversky and to-be Nobel winner Daniel Kahnneman (Kahneman & Tversky, 1973). Yet we continue to love our pundits even though we should know they are mostly bad.

What's more, expertise seems not to make much difference; academic pundits with doctorates who write books nearly 400 pages long are as likely to be wrong as the most gung-ho Fox News opinionistas, at least according to the body of empirical evidence amassed by psychologist Philip Tetlock (Tetlock, 2005). That's because taking a strong stance and sticking to it is a common factor in being both a pundit and a bad predictor—in fact a terrible predictor, worse than random. Pundits keep their jobs because of our hardwired psychological traits rather than their competence.

I've explored these general failings of general predictive competence elsewhere, and will do so in more detail in future. The present topic is medicine, and here, too, general predictive claims are usually incredible—and I'll explain why, after giving an example.

The general nature of this example is the "Magical Internal Doctor Hypothesis". It occurs to different authors in different forms. Klaus Schwab imagines "smart dust" consisting of tiny robots that could circulate in our blood stream, detecting and destroying pathogens before we even know it (Schwab, 2016). Noah Yuval Harari imagines something remarkably similar:

Within a few decades, Big Data algorithms informed by a constant stream of biometric data could monitor our health 24/7. ...by 2050, thanks to biometric sensors and Big Data algorithms, diseases may be diagnosed and treated long before they lead to pain or disability. (Yuval Harari, 2018, p. 49)

The Magical Internal Doctor Hypothesis might be true. Who am I to say it's not? Both the authors just mentioned pepper their work with qualifiers

to the effect that they are not saying that these scenarios will be actualized. They are not making specific predictions. These are just imaginary scenarios, used for illustrative purposes only (the actual contents of 4IR may differ). This is meant to excuse the complete lack of systematic marshalling of evidence, and of the systematic consideration of, and testing against, contrary hypotheses that are fundamental to science. Yet the pronoucements are often as close as can be to categorical predictions. They are written to sound plausible, as if the person writing believes them. Saying "by 2050", for example, suggests that there is some sort of basis for picking that date. If one says, "By next Tuesday, we might see a new president in charge", that implies that one has some reason to think that a change of power is imminent. So it won't do simply to say that the predictions are "mights" and "maybes". They are expressed as if they are based on sound reasoning from serious consideration of the evidence.

However, they are not. They are mere speculations. That is why the authors, and many others like them, periodically remind the reader that they are not making specific predictions. They know that they don't know what is going to happen. But it suits them to sound as if they do, without quite committing to anything. In the context of medicine, I regard this as irresponsible speech. These predictions are not innocuous, despite the qualifiers.

In the first place, even if the specific predictions are not supposed to be essential to any particular arguments made about how we must prepare ourselves for the future, it is obvious that these claims (or claims of this kind) are at least rhetorically essential to their respective books, and thus essential to the popularity and ensuing fame and wealth of the authors of these works in which they are found. In my view, it is intrinsically unethical to knowingly or carelessly allow people to believe falsehoods for personal profit. I believe that it matters what people believe.

In the second place, there is a danger of raising expectations about medicine, which is already a serious problem for clinical practice. People are already prone to expect modern medicine to painlessly cure everything, maybe with the exception of hereditary cancers (and even then, it's common to encounter the impression that a cure

> **❝**But modern medicine can't cure everything, and isn't painless. To promote excessive expectations of modern medicine places great pressure on medical practitioners, who have to find a way to puncture these inflated expectations. False hope creates acrimony in the consultation room and devastating disappointment outside it.**❞**

is round the corner.)

But modern medicine can't cure everything, and isn't painless. To promote excessive expectations of modern medicine places great pressure on medical practitioners, who have to find a way to puncture these inflated expectations. False hope creates acrimony in the consultation room and devastating disappointment outside it.

You might be wondering whether my stance towards medicine is unduly cynical. But I have the greatest respect for medicine and its achievements. But this shouldn't prevent a sober assessment of its achievements. Of the top ten causes of death in America in 1900, six are recognizable on the 1998 list (Rockett, 1999, p. 8). In the intervening century, antibiotics made a huge difference, and the four departures from the list are all infectious diseases. But their success was not the start of an upward trend, let alone an exponential curve (Stegenga, 2018). Viruses failed to fall to medical innovation, and remain basically intractable. Lifestyle-related diseases claim more lives than they used and suicide has made an appearance in the top ten—perhaps a symptom of the wider perplexity and disillusionment that has characterized the postmodern era (Tarnas, 1991). Mental illness generally is not proving especially tractable for modern medicine. Surgery is not the miracle it is sometimes thought to be; different surgical procedures have different success rates, and of course are hard to test empirically because of the lack of a plausible placebo, and so forth.

I'm not saying that modern medicine has enjoyed no successes. It definitely has. However, most of these successes were concentrated in a dramatic medical revolution running from around the end of the 19th century through the middle of the twentieth (Stegenga, 2018). Then the revolution stopped. Mere progress settled in: slow, painful progress. And the big picture remains

one of isolated successes amid a general sea of well-intentioned efforts to varying degrees of uselessness (Broadbent, 2019; Stegenga, 2018). This story is familiar from the pre-twentieth century history of medicine, and it's nothing to be ashamed of. It is what it is, which is considerably preferable to what it was. I'm no anti-medicine warrior. But the fact is that we're nowhere near what one might call a "completed medicine".

To hold out any hope that we are near a "completed medicine" in this situation is irresponsible, unless it is based on a proper assessment of the evidence. And artificial intelligence, machine learning, "Big Data algorithms", "biometric sensors", and so forth do not change the situation of medicine. Maybe some algorithmic panacea is just around the corner; but we have no more reason to think so in 2019 than we did in 1719, strange though that sounds to the inflated collective ego of the contemporary era. Even if that general claim is wrong, the Magical Internal Doctor Hypothesis that I've described is most emphatically not based on the discovery of any such reason. Rehearsing the familiar technobabble about algorithms and sensors and the latest advances in epigenetics available to us in 2019 simply does not cut the mustard, any more than the creation of the Principality of Liechtenstein three hundred years earlier gave reason to think that world peace was round the corner.

There are large-scale consequences of holding out false hope of imminent medical breakthroughs. Public sector funding is politically guided, and political guidance can be quite sensitive to irresponsible speech of the kind I'm describing. (In particular, Klaus Schwab appears to seek to influence policy.) As a consequence, money may be ploughed into one or another channel of research, sub-optimally. Maybe the development of deep learning tools for clinical application becomes all the rage, and swallows money that could be spent assessing the effectiveness of sugar taxes on reducing diabetes incidence. Maybe money is directed (even more) into genetically-oriented biomedical research, when it could be put into implementing known-effective health-related measures, like improving education (strongly correlated with longevity, among other obvious benefits), improving living conditions for the poorest, or even biomedical research directed at developing better cures for neglected tropical diseases.

Even if it had none of these consequences, however, it would be ethically problematic to be negligent in making bold factual claims for personal profit.

I've focused on one particular idea, the Magical Internal Doctor Hypothesis, but there are many other examples of similarly unfounded pronouncements. For example, a recent press article announced that robotics was the future of medicine. This is weird, because robotics is hardly a 4IR technology. The robotics revolution in the automotive industry brought Detroit to its knees in the 1990s (and in doing so incidentally spurred the rise of techno music). Even setting this historical point aside, however, it's obvious that there is more to medicine than surgery, and moreover that the abilities of surgeons to operate with greater precision and less invasion have in any case considerably increased over past decades. It's plausible that continued efforts to develop robotic aids for surgeons will continue this trend (although it would be nice to see a bit more thought given to whether there may also be reasons to doubt this). But it's a long leap from there to the future of medicine. Will robotics be increasingly developed for medical applications? Probably. Does that make robotics the future of medicine? No. Does it matter? Yes, because it might lead to research funding being sub-optimally deployed.

The limits of pure data

There are more serious attempts to predict or project the future of medicine, which do not amount to irresponsible speech. However, some of them are wrong, and for a specific reason, which I want to illustrate with an example.

The August 2019 issue of Nature reported on a tool built by Google Health for predicting acute

> 66 Will robotics be increasingly developed for medical applications? Probably. Does that make robotics the future of medicine? No. Does it matter? Yes, because it might lead to research funding being sub-optimally deployed. 99

> 66 However, it is far less clear that machine learning can satisfactorily justify the hypotheses they create. This is because merely working with data can only get us so far. We need to relate the data back to what it is about. We need to interpret these hypotheses. 99

onset kidney disease with impressive accuracy (Tomašev et al., 2019). It claimed that the tool was "clinically applicable". Aside from the specific significance for acute onset kidney disease, the claim to clinical applicability suggests that there might be a completely new way to understand, predict, and ultimately make health decisions of all kinds.

The usual process is a slow, painful accumulation of knowledge across multiple long and costly studies. Their results may be hard to synthesise, their transportability unknown, and after all that, we may not have a clear idea of how a system will respond to an intervention (Hernán & Robins, 2020; Pearl & Mackenzie, 2018). Instead of all that, we simply give a suitable dataset to Google (all data arising from hospital admissions), along with the questions we want answered ("Who is going to get acute onset kidney disease?"). Google then comes back with a plug-and-play programme that answers our questions. "Clinically applicable" means we can use this programme in a live clinical setting, and the fact that it learns as it goes means that its accuracy may be expected to improve over time.

It won't answer the questions that epidemiologists have traditionally asked. It won't tell us, for example, why certain patients are highly likely to develop acute onset kidney disease. But it will tell us that certain patients will develop acute onset kidney disease, with remarkable accuracy, 48 hours before it happens. And that, ultimately, is what the clinician wants—isn't it?

This is a familiar story: artificial intelligence radically outperforming old ways of doing things. It's also one that presents challenges very familiar to philosophers. Inductive inference in general is not susceptible to formalisation: that is, the form of the inference does not guarantee that it works. It matters what the inference is about. But

machine learning, however sophisticated, deals only in data that could, so far as the machine is concerned, be about anything at all. We therefore know that machine learning has in-principle limits. These limits correspond to the familiar problem of external validity or transportability, where we face the inferential challenge of understanding whether, when and how knowledge gained in relation to a studied population can be applied to a target population. Substitute "data set" for "population" and you have an expression of the exact same problem as it faces machine learning. The glitz of the new is apt to dazzle us to this old problem, however; hence the inappropriate use of "clinically applicable" in the above-mentioned article.

Machine learning can be an extremely important source of hypothesis generation, but there must be a distinction drawn between the context of discovery and the context of justification (Popper, 1959). This distinction isn't as influential as it was; it doesn't dominate the philosophy of science any more, and it isn't immune to criticism. Nevertheless, it has its uses, and this is one of them. Data-driven methods powered by machine learning are surely great for generating hypotheses. These may be highly creative, from our perspective—things we never would have thought of (as illustrated by chess computers, which conceive of plays that are, from the point of view of chess knowledge, novel and highly original). There is a separate question to be explored as to how this is so; the point is that it may be so.

However, it is far less clear that machine learning can satisfactorily justify the hypotheses they create. This is because merely working with data can only get us so far. We need to relate the data back to what it is about. We need to interpret these hypotheses. If we don't, then we cannot justify what we have hypothesised. This is because no inductive inference is ever justified by its form alone. That's in the nature of an inductive inference. Instead, it is facts about the things that the inference is about that provides the warrant. In this instance, it would be facts about acute kidney injury. What features of the incoming patients is the machine detecting, through the lens of the data? How do these give rise to, or arise from causes of, the onset of kidney failure?

Answering these questions is essential to

determining whether the machine has found a chance pattern in the data—any dataset large enough will contain all sorts of interesting-looking patterns that arise by chance alone—or if it corresponds to something real. This is important to our decision as to whether indeed to apply this approach in a clinical setting. But it's also important to the advance of medical knowledge. Exciting though machine learning is, nearly all our existing medical success stories have a basis in biomedical theory—even if their discovery was somewhat fluky in some cases, and even if some, such as anaesthesia, remain inadequately explained. None was discovered, let alone justified, by appeal to mere patterns in data.

A deep learning approach to de novo small molecule design was also published in a Nature group journal the following month (Zhavoronkov et al., 2019). This approach included empirical testing, with the "lead candidate" demonstrating "favourable pharmacokinetics in mice". This is hardly a randomised controlled trial, and I reserve my right to scepticism about the results. Nonetheless, the approach shows a laudable effort to understand the physical reality beneath the patterns in the data. This is how machine learning can really advance medicine.

Towards better predictions

I've heavily criticized some predictions for being irresponsible, and argued that others are well-researched but misguided. How can we do better?

Properly considering the evidence for a prediction means testing it against contrary hypotheses, and I've given a detailed account of how to do this in the context of public health predictions previously (Broadbent, 2011, 2013), which is readily extended to cover the categories of prediction about medicine covered here. There is a simple test you can apply: what could possibly go wrong?

Consider scenarios in which the prediction you are making comes out false. (For any machine-led discovery process, this will include the scenario in which the pattern discovered by the machine has no basis in reality, and is a data-fluke. That's why trying to understand the reality beneath the hypothesis is so important.) Search for the most plausible of such scenarios, the ones that are the likeliest, the ones that are compatible with your current evidence as far as possible, the ones that are best supported by the data. These are competitors to your favored hypothesis. Then set about looking for evidence that will discriminate between, on the one hand, your hypothesis, and on the other hand, the competitors. The more you iterate this process, the stronger your prediction becomes. It won't necessarily guarantee truth. But it certainly secures you from negligence, since you can take proper care over what you say, yet still come a cropper. And it has a place in protecting serious research, such as the research underlying the Nature paper discussed above, from over-reach, as when an entirely data-driven approach is pronounced clinically applicable without any consideration of the possible physical and biological explanations for the model's success, as if it were an oracle rather than a tool for scientific discovery. ∎

References

Broadbent, A. (2011). What could possibly go wrong? - A heuristic for predicting population health outcomes of interventions. Preventive Medicine, 53(4–5), 256–259.

Broadbent, A. (2013). Philosophy of Epidemiology. London and New York: Palgrave Macmillan.

Broadbent, A. (2018a). Intellectualizing Medicine: A Reply to Commentaries on "Prediction, Understanding, and Medicine." The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 43(3), 325–341. https://doi.org/10.1093/jmp/jhy002

Broadbent, A. (2018b). Prediction, Understanding, and Medicine. The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 43(3), 289–305. https://doi.org/10.1093/jmp/jhy003

Broadbent, A. (2019). Philosophy of Medicine. London and New York: Oxford University Press.

Frankfurt, H. G. (2005). On Bullshit. Princeton NJ: Princeton University Press.

Hernán, M. A., & Robins, J. M. (2020). Causal Inference: What If. Retrieved from https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

Kahneman, D. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). The simulation heuristic. Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237–257.

Kahneman, D., & Tversky, A. (1982). Evidential impact of base rates (D. Kahneman, P. Slovic, & A. Tversky, eds.). Cambridge: Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). The Book of Why. New York: Basic Books.

Popper, K. (1959). The Logic of Scientific Discovery. London: Routledge.

Rockett, R. H. (1999). Population and Health: An Introduction to Epidemiology. Population Bulletin, 54(4), 1–44.

Schwab, K. (2016). The Fourth Industrial Revolution. London: Penguin UK.

Stegenga, J. (2018). Medical Nihilism. Oxford: Oxford University Press.

Tarnas, R. (1991). The Passion of the Western Mind. New York: Ballantine Books.

Tetlock, P. (2005). Expert Political Judgement - How Good Is It? How Can We Know? Princeton and Oxford: Princeton University Press.

Tomašev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., … Mohamed, S. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. Nature, 572(7767), 116–119. https://doi.org/10.1038/s41586-019-1390-1

Wootton, D. (2006). Bad Medicine: Doctors doing harm since Hippocrates (electronic book). New York: Oxford University Press.

Yuval Harari, N. (2018). 21 Lessons for the 21st Century. New York and London: Spiegel & Grau, Jonathan Cape:

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., … Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology, 37(9), 1038–1040. https://doi.org/10.1038/s41587-019-0224-x