COMMENTARY *(see article on pp. 305–308)*

Skin Cancer Diagnosis: Shining Light into Dark Places

Jonathan Rees
Grant Chair of Dermatology, University of Edinburgh, Dermatology, Rm 4.018. Lauriston Building, Lauriston Place, Edinburgh, EH3 9HA, United Kingdom.
E-mail: reestheskin@me.com

Frank Davidoff, the celebrated US physician, has a telling metaphor for medical competence: he likens it to Dark Matter, the material physicists think makes up most of the Universe, and about which we know so little (1). Clinical competence would, in this analogy, be the material that literally holds and binds clinical practice – our professional lives – together. How come we know so little about it?

Literary fiction of course has plenty of bad doctors, but frequently they appear to be evil rather than merely clinically indifferent. Dr Watson, Sherlock Holmes' fictional friend and narrator, is of course a general practitioner, and whereas we are unlikely to admire his intellectual abilities, we assume he is 'sound' in his professional life. This lack of judgement or curiosity about the abilities of individual physicians is however widespread in the real world too. Within the UK national health service (NHS), doctors, it sometimes seems, are viewed as interchangeable units, who possess skills defined at a lower bound in terms of training and certification, and who can be deployed as though they were machine tools in an industrial factory. Many of us think otherwise, believing that there is considerable variation in clinical competence (2). Hidden in plain sight perhaps, but present.

Different types of evidence attest to this variation in performance. Most doctors when they or their immediate families are ill, choose who to consult with great care. Ever since the seminal work of Wennberg (3), we know that clinical decision-making by doctors is far from uniform or consistent, and that such differences account for much of the variance in health care costs. We do not understand all the reasons for clinical heterogeneity – financial biases are likely to play a role – but from the patient's perspective, medical advice is heavily dependent on who you consult. Finally, given what we know about expertise in general – in fields as diverse as musicianship, sport or academic ability – it would be bizarre if differences in ability were not the norm rather than the exception (4, 5).

It is the above arguments that make the paper from Ahnlide & Bjellerup (6) of Helsingborg Hospital and Lund University so welcome. The paper is a remarkably straightforward and lucid piece of work – as good a reason for publication as any – in which the authors have set out to measure diagnostic accuracy for skin cancer in a particular hospital practice. Key to their design is that the study was prospectively designed and that doctors were obliged to provide only one clinical diagnosis. Why is this important?

The problem that bedevils retrospective studies on this topic is that the process of recording a diagnosis (let alone making a diagnosis) is not as straightforward as people often think. Imagine you see a patient with a pigmented lesion: you deem it clinically suspicious, but most likely to be a benign naevus, rather than a melanoma. You are 60% certain it is a benign melanocytic naevus, but 40% certain it could be a melanoma. What do you record as your diagnosis? Given a choice of only one diagnosis or 'action point', do you write naevus, or say 'exclude melanoma'. Most of us, for good reason do not view errors in diagnosis as being symmetrical. If I scribble 'naevus' in the case notes, but the pathologist reports melanoma, my mental state differs from when the process is reversed – that is, when my clinical diagnosis of melanoma is contradicted by the pathologist saying it is a naevus. Our approach to diagnostic sensitivity and specificity in the clinical workflow is heavily context dependent, but from a researcher's perspective, this muddies the water. Forcing a uniform workflow process allows insights that would escape us if we just studied clinical case notes retrospectively.

Ahnlide & Bjellerup (6) interpret their results using two standard measures of performance: sensitivity and predictive value. To place their data in perspective however, we also need to think about specificity as well. Sensitivity of a test (here the clinician's diagnosis) can be defined as the ratio of true positives to the total number of positives defined with a gold standard (pathology). The latter denominator will of course include the genuine diagnoses that were missed by the physician (i.e. clinical false negatives). In one sense sensitivity is about making certain you 'rule in' a particular diagnosis. An easy way to do this of course is to call every pigmented lesion you see a melanoma: the downside of this is that your specificity goes down, that is, you are increasingly bad at 'ruling out melanomas'. But for a diagnosis that you cannot afford to miss, say, melanoma (rather than a basal cell carcinoma (BCC) for instance), a high sensitivity is a price surely worth paying for a low specificity. But notice the word 'price': it comes at a cost.

Predictive value reflects different aspects of accuracy. It is simply the ratio of the correct clinical diagnosis to all the cases you label as having the same clinical

diagnosis, as arbitrated by the pathologist. The denominator here includes the true positives identified by the clinician but also the false positives – those where the clinician has called it a melanoma but the lesion was a benign naevus on pathology (the reader will note I assume pathologists are infallible…). In one sense predictive value conforms to our everyday sense of 'how often we are right'. Its downside (as the authors remind us) is that it is heavily dependent on the prevalence of disease. An example that I use with clinical trainees highlights this. Imagine I sit in my office, and do not see any patients in the pigmented lesion clinic. Instead, knowing that say only 2% of referred patients have a melanoma, I diagnose every patient as 'not being melanoma'. I am right 98% of the time and boast about my 98% diagnostic accuracy, all without having to put my coffee aside to visit the clinic. Of course, if the case mix alters my claims are seen for what they are. If the clinic now contains 50% of patients with melanoma, my accuracy would have dropped to 50% – and yet my behaviour has not changed. It is this aspect of predictive value that so often limits attempts to compare primary care and secondary care physicians. A GP can claim expertise in melanoma diagnosis because no patient of his has died of melanoma in the last 10 years. However examination of the incidence of melanoma in the UK, tells us that even a blind clinician could make the same argument for his or her clinical acumen.

Ahnlide & Bjellerup's (6) findings are in one sense expected. We are better at diagnosing BCCs than squamous cell carcinomas, but that our sensitivity and predictive values for diagnosing melanomas are much lower (the inhabitants of Helsingborg in particular should remember however that this is a study based on a single diagnosis, not the decision whether to request a biopsy or not). What they also provide is interesting data on which way we makes errors between diagnostic categories.

I started this brief essay pointing out that we knew little about medical expertise (2). This situation is likely to change for at least two reasons. First, all that we know about the acquisition of any form of advanced skill, tells us that feedback on performance is critical. Without information on how we are doing, it is not possible to review progress and improve on our abilities. If you wish to perform – and in this sense medicine is about performance – you have to actively seek out ways to improve. Mere passive contemplation is not sufficient: data, and in particular structured data, is needed (5, 7). Second, doctors need to be increasingly aware that others too are interested in performance. David Margolis writing in the Archives of Dermatology almost 15 years ago recalled that whereas his early career ran parallel to an explosion of basic bioscience discovery, as far as clinical practice was concerned "no change has been as dramatic as the changing landscape of who is primarily responsible for patient care and who pays the bill" (8). Competence is inextricably linked to health care costs and patient outcomes. Ahnlide & Bjellerup's paper is worth reading: there is plenty more to come.

REFERENCES

1. Davidoff F. Focus on performance: The 21st Century revolution in medical education. Mens Sana Monogr 2008; 6: 29–40.
2. Mylopoulos M, Lohfeld L, Norman GR, Dhaliwal G, Eva KW. Renowned physicians' perceptions of expert diagnostic practice. Acad Med 2012; 87: 1413–1417.
3. Wennberg JE. Tracking medicine: A Researcher's quest to understand health care. Oxford: Oxford University Press; 2010.
4. The Cambridge handbook of expertise and expert performance. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, editors. Cambridge; New York: Cambridge University Press; 2006.
5. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. Med Educ 2007; 41: 1124–1130.
6. Ahnlide I, Bjellerup M. Accuracy of clinical skin tumour diagnosis in a dermatological setting. Acta Derm Venereol 2013; 93: 305–308.
7. Rees JL. Teaching and learning in dermatology: from Gutenberg to Zuckerberg via way of Von Hebra. Acta Derm Venereol 2013; 93: 13–22.
8. Margolis DJ. Do we have time for the change? Arch Dermatol 1998; 134: 1151.