

Bridging the gap between uncertainty, confidence and diagnostic accuracy: calibration is key

Laura Zwaan,¹ Wolf E Hautz^{2,3}

¹Institute of Medical Education Research Rotterdam (iMERR)/Erasmus MC, The Netherlands, Rotterdam

²Department of Emergency Medicine, Inselspital University Hospital Bern, University of Bern, Bern, Switzerland

³Centre for Educational Measurement, Faculty of Educational Sciences, University of Oslo, Oslo, Norway

Correspondence to

Dr Laura Zwaan, Institute of Medical Education Research Rotterdam (iMERR)/Erasmus MC, Wytemaweg 80, 3015CN Rotterdam, The Netherlands; l.zwaan@erasmusmc.nl

Received 7 January 2019

Accepted 7 January 2019



► <http://dx.doi.org/10.1136/bmjqs-2018-008390>



© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Zwaan L, Hautz WE. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2018-009078

Picture yourself in a casino sitting at the roulette table. Will you put your money on either black or red, or will you go for a riskier choice and place it on just a single number? But before you can make up your mind, the elderly, obese smoker to your right reaches for his chest and collapses. ‘What is wrong with him?’, cries his accompanying wife, looking at you in shock.

In this issue, Lawton and colleagues ask whether more experienced clinicians are better able to tolerate uncertainty and manage risks.¹ Doctors working in three emergency departments (ED) read four clinical vignettes each accompanied by four quite distinct options for management, all of which might be deemed clinically acceptable. Participants used 5-point Likert scales to indicate their agreement with each of the management plans. By design, the four options offered for each vignette included two management plans one would characterise as ‘risk averse’—requiring further tests and/or admission of the patient to hospital. The other two management plans may be seen as less risk averse, usually involving referring a patient back to the general practitioner or offering reassurance to the patient. In addition to rating their agreement with each of the possible management plans for the four vignettes, participants reported their experience (number of years worked in an ED setting) and answered a 15-item ‘Physicians Reactions to Uncertainty’ scale. Analysing the responses from 90 doctors, who had worked in EDs for between 5 weeks and 21 years, Lawton and colleagues report a medium to large correlation between experience and tolerance of risk. More

experienced clinicians made less risk averse decisions.

We would like to provide three conceptual considerations potentially relevant to interpret these results from Lawton and colleagues and identify possible follow-up research.

WHAT IS RISK AND HOW DOES IT DIFFER FROM UNCERTAINTY?

Roulette serves as a classic example for *risk*: unlike any clinical situation, all possible outcomes are known, as are their respective probabilities. You only have to decide how much risk you can tolerate. Whether you are an experienced emergency physician or not, the unfortunate gambling fellow from above poses a fundamentally different challenge—an *uncertain* situation. You can only observe the patient’s obvious cardiac risk factors obesity and smoking and apply some general knowledge of the several life-threatening conditions that may cause chest pain and collapse, including, for instance, not just the obvious possibility of a heart attack, but also a blood clot in the lungs or possibly a dissection of the aorta. In addition to these usual medical causes to consider, however, there is also the possibility he may just be acting to avoid paying his debt.

In contrast to roulette, in uncertain situations we do *not* know all possible outcomes and their respective probabilities. We also do not know how the likelihoods of different outcomes change when more information becomes available. In fact, likelihoods quickly become non-calculable in clinical practice. For instance, an experienced physician will be aware that chest pain in itself has a positive predictive

value of only around 20% for a heart attack and may also be aware that obesity doubles the risk for coronary artery disease. However, calculating the three most likely causes of collapse for a given patient such as the elderly obese smoker with chest pain is already next to impossible. Uncertainty is thus characterised by incomplete knowledge about the potential outcomes (there could be diagnoses we did not think of) and complex relationships between cues and criterion.² Your profession and professional experience may change the degree of uncertainty the collapsing gambler poses to you, for example, for a medical doctor versus somebody without medical knowledge, but even for medical doctors it is an uncertain situation because they encounter it without the accompanying information and diagnostic possibilities they would have in a hospital situation.

Emergency medicine (EM), where Lawton and colleagues place their study,¹ is a particularly uncertain—or ill-defined—environment. Decisions in EM often have to be based on incomplete and unreliable information,^{3,4} are made under time pressure and with potentially devastating consequences.⁴⁻⁶ However, the degree of uncertainty any environment poses is by no means independent of one's experience in it: humans are outstanding in identifying among the myriad of cues available in any given situation, those which best predict the criterion of interest and then combining the few important cues into simple decision rules ('heuristics').^{7,8}

This mechanism allows people to 'become an expert by making routine what to the novice requires creative problem-solving ability.'⁹ It does, however, require extensive exposure to the respective environment to develop. The more experience one has with a given situation, the better one will be able to foresee its future course.¹⁰ Thus, experience within a given field reduces uncertainty in that field. It is consequently of little surprise that physicians with limited EM experience tend to take seemingly safer decisions, whereas more experienced physicians seem to tolerate more uncertainty as found by Lawton and colleagues. Instead of explaining this phenomenon as a tolerance of uncertainty increasing with experience as Lawton and colleagues conclude, we would argue that the EM environment simply becomes less uncertain over time. Physicians who have been practising in EM for longer increasingly regard familiar problems as less uncertain—that is, it is not that they tolerate uncertainty better so much as they have less uncertainty to tolerate.

MAKING DECISIONS WHILE ACKNOWLEDGING UNCERTAINTY

A second question is how to define tolerance of uncertainty. The literature lacks a clear definition.¹¹ Lawton *et al* operationalise tolerance of uncertainty simply as allowing for more risky decisions in a well-justified effort to reduce the number of diagnostic tests.¹ And indeed, while junior physicians often aim for certainty as a prerequisite for action, more experienced physicians are

more confident in the appropriateness of their actions even in situations where they do not (yet) fully understand a patient's problem.^{12,13} Experts become more 'comfortable with not knowing.' Ilgen and colleagues recently reviewed the concepts of 'comfort with uncertainty' and suggested that skilful clinicians can be comfortable in their management plan and at the same time be not completely confident about their diagnosis.¹⁴ The reason for this is that skilful clinicians rely on their experience to be confident about having ruled out all conditions that require immediate action. They are able to be comfortable with their management plan while still being uncertain about the precise diagnosis. They then use an iterative process of pragmatic empiricism, forward planning, monitoring and ongoing reconceptualisation to find the exact diagnosis. This helps them to maintain the ability to act even in ill-defined environments that are characterised by an inherent uncertainty. Junior physicians by contrast first need to obtain experience and learn how to cope with uncertainty by identifying cues to monitor their actions.¹⁵ For example, if a patient reacts to a treatment as expected, this signals to a physician that (s)he is 'on track', but if the patient is not responding to treatment, physicians need to reconsider their diagnosis and ask for help because the conditions may be beyond their abilities to control or manage.¹⁴

When humans make decisions, we not only have access to the internal and external cues our decision is based on, we also monitor our cognitive operations. Such metacognition enables us to recognise how confident we are in a decision,¹⁶ thus, for instance, recognising when to slow down when we should¹⁷⁻¹⁹ and to over-rule first, intuitive impressions with more deliberate reasoning. The most prominent of these metacognitive impressions is fluency—the subjective impression of how easy or difficult a decision is.²⁰ Several researchers have suggested that fluency is not an absolute judgement of the ease or difficulty of a given cognitive operation, but rather results from the difference between expected and experienced difficulty.²⁰ Fluency does emerge from judgements and decisions, and from perception, retrieval or any other cognitive operation. Objects, for example, are fluent when we have seen them frequently or recently, memory is fluent when relied upon often. Thus, 'the interpretation of a fluency experience relies on past experiences and the current context,'²⁰ that is, the more exposure we have had or the more expert we are in a given context, the smaller the perceived difficulty of operating in it. We would suggest that this effect also results from learning to use the right cues to monitor whether we are on track.¹⁵ This hypothesis can and should be tested, because it could inform the 'interventions that better support junior doctors to deal with uncertainty' Lawton and colleagues call for.¹ Interventions that help learners to identify the right cues to monitor themselves are defensible and unlikely to cause harm¹⁵ while for reasons of patient safety, it would obviously not be advisable to ask junior physicians to simply

be more risk seeking when it comes to diagnostic testing. As pointed out earlier, junior physicians need more experience to be able to make good decisions.²¹ We argue here that junior physicians need to go through the phase where uncertainty is acknowledged, to facilitate the learning process needed to develop more confidence and to base that confidence judgement on the right cues. Ultimately, this could result in reducing the number of tests acquired on any patient. The question is whether such a learning process can be facilitated in a way that junior physicians reach the level of expert physicians faster.

DECISION ACCURACY

This brings us to the third—but most crucial—question: what about accuracy? Regardless of how experienced we are within an uncertain environment, it is of little value if our decision is wrong. And even if a decision is right, we better be sure it is right for the right reasons, not just because it has gone well in the past.²² Otherwise, we will be wrong in a similar situation eventually. In previous research, physicians were found to be overconfident in their diagnoses,²³ and experienced physicians in particular are less likely to seek advice, even when they were wrong.²⁴ It has also been shown that confidence increases with seniority—but its calibration with accuracy not necessarily.²⁵ Based on these results we would argue that looking only at confidence, and suggesting that higher is better, is in fact rather dangerous. A study by Friedman *et al* showed that medical students were more likely to get the diagnosis wrong than more experienced physicians, but were better calibrated than the more experienced physicians.²⁶ In this study, the experienced physicians were more often not confident in their diagnosis rather than being overconfident,²⁶ which is different from the study by Meyer *et al* who found a strong effect of overconfidence.²⁴ This is only to show that more research is needed to unravel the complex relations between adequacy, accuracy and calibration in conjunction with confidence, risk management and uncertain situations even aside from experience.

CALIBRATION IS KEY

How do junior doctors get to the point where they can safely decide that no further tests are needed? Generally, experienced physicians are often considered to be overconfident and junior physicians, as is shown by Lawton and colleagues, are less confident. However, it could very well be the case that even experienced physicians are not confident in cases with which they have little experience, and that junior physicians are overconfident in certain situations too. In the end, what really matters is that the level of confidence is well calibrated with diagnostic accuracy. We want physicians (junior physicians and experienced physicians alike) to be more risk averse when encountering patients with symptoms they are unfamiliar with, while this might be an inappropriate course of action when they encounter a more familiar

case. Good calibration is key for physicians of all levels of expertise and experience, because it will result in the right decisions. One way to improve calibration is to provide performance feedback (ie, informing physicians whether their diagnosis and/or treatment was correct or not).²⁷ Several recent articles have pointed out the poor levels of calibration and the lack of feedback in diagnostic reasoning.^{28–30} Feedback is—besides experience—an important prerequisite for becoming an expert. Physicians, especially in the ED, often do not get feedback about their diagnostic performance, that is, they do not hear back about changes in the diagnosis after a patient has been admitted to the ward or was sent home. Such feedback is essential to enable physicians to recognise cues that a situation differs from the familiar situation they initially thought it was, and to decrease the potentially dangerous situation mentioned above of experienced physicians being confident in their diagnosis but still being wrong. In the field of medical education, interventions have been developed that have proven to be efficient in improving calibration of uncertainty and diagnostic performance.²⁷ Mixed deliberate practice with feedback is what ultimately makes novices expert.^{31 32}

IN CONCLUSION

The study of Lawton and colleagues showed that experienced physicians make less risk averse decisions than junior physicians.¹ This reflects an important difference in clinical decision-making, but it only represents one piece of the puzzle. Physicians do not know when they are right or wrong, which sometimes results in overconfident decisions and other times in defensive medicine. Next steps should include studying tolerance of uncertainty in combination with overconfidence and diagnostic accuracy to determine where most miscalibration occurs. Furthermore, a focus on reducing miscalibration using strategies to enhance feedback and identifying the right cues to inform confidence in judgements are important ways to improve clinical reasoning.

Funding Dr. Zwaan is supported by the Netherlands Organization for Scientific Research (NWO Veni Grant nr: 451-16-032)

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

- 1 Lawton R, Robinson O, Harrison R, *et al*. Are more experienced clinicians better able to tolerate uncertainty and manage risks? A vignette study of doctors in three NHS emergency departments in England. *BMJ Qual Saf* 2018;Bmjqs-2018-008390.
- 2 Baron J. *Thinking and Deciding*. Cambridge University Press, 2000.
- 3 Ilgen JS, Humbert AJ, Kuhn G, *et al*. Assessing diagnostic Reasoning: a consensus statement summarizing

- theory, practice, and future needs. *Acad Emerg Med* 2012;19:1454–61.
- 4 Brown TW, McCarthy ML, Kelen GD, *et al.* An epidemiologic study of closed emergency department malpractice claims in a national database of physician malpractice insurers. *Acad Emerg Med* 2010;17:553–60.
 - 5 Schaffer AC, Jena AB, Seabury SA, *et al.* Rates and characteristics of paid malpractice claims among US physicians by specialty, 1992-2014. *JAMA Intern Med* 2017;177:710.
 - 6 Kachalia A, Gandhi TK, Puopolo AL, *et al.* Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med* 2007;49:196–205.
 - 7 Gigerenzer G, Gaissmaier W. Heuristic decision making. *Annu Rev Psychol* 2011;62:451–82.
 - 8 Gigerenzer G, Todd PM. *Simple Heuristics That Make Us Smart*. Oxford: Oxford Univ Press, 2001.
 - 9 Anderson JR. *Cognitive psychology and its implications*. WH Freeman/Times Books/Henry Holt & Co, 1985.
 - 10 Pusic M, Pecaric M, Boutis K. How much practice is enough? Using Learning curves to assess the deliberate practice of radiograph interpretation. *Acad Med* 2011;86:731–6.
 - 11 Bhise V, Rajan SS, Sittig DF, *et al.* Defining and measuring diagnostic uncertainty in medicine: a systematic review. *J Gen Intern Med* 2018;33:103–15.
 - 12 Lingard L, Garwood K, Schryer CF, *et al.* A certain art of uncertainty: case presentation and the development of professional identity. *Soc Sci Med* 2003;56:603–16.
 - 13 Cristancho S, Lingard L, Forbes T, *et al.* Putting the puzzle together: the role of 'problem definition' in complex clinical judgement. *Med Educ* 2017;51:207–14.
 - 14 Ilgen JS, Eva KW, de Bruin A, *et al.* Comfort with uncertainty: reframing our conceptions of how clinicians navigate complex clinical situations. *Adv in Health Sci Educ* 2018;19.
 - 15 de Bruin ABH, Dunlosky J, Cavalcanti RB. Monitoring and regulation of learning in medical education: the need for predictive cues. *Med Educ* 2017;51:575–84.
 - 16 Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof* 2008;28:14–19.
 - 17 Moulton CA, Regehr G, Lingard L, *et al.* 'Slowing down when you should': initiators and influences of the transition from the routine to the effortful. *J Gastrointest Surg* 2010;14:1019–26.
 - 18 Huang GC, Kriegel G, Wheaton C, *et al.* Implementation of diagnostic pauses in the ambulatory setting. *BMJ Qual Saf* 2018;27:492–7.
 - 19 Lambe KA, O'Reilly G, Kelly BD. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Qual Saf* 2016:bmjqs-2015-004417.
 - 20 Oppenheimer DM. The secret life of fluency. *Trends Cogn Sci* 2008;12:237–41.
 - 21 Norman GR, Eva KW. Diagnostic error and clinical Reasoning. *Med Educ* 2010;44:94–100.
 - 22 Hautz WE. When I say... diagnostic error. *Med Educ* 2018;52:896–7.
 - 23 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121(5 Suppl):S2–S23.
 - 24 Meyer AND, Payne VL, Meeks DW. Confidence, and resource requests: a vignette study. *JAMA Intern Med* 2013;173:1952–8.
 - 25 Hautz WE, Schubert S, Schaubert SK. Accuracy of self-monitoring: does experience, ability, or case difficulty matter? *Med Educ*.
 - 26 Friedman CP, Gatti GG, Franz TM, *et al.* Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *J Gen Intern Med* 2005;20:334–9.
 - 27 Nederhand ML, Tabbers HK, Splinter TAW. The effect of performance standards and medical experience on diagnostic calibration accuracy. *Health Prof Educ*. 2018.
 - 28 Omron R, Kotwal S, Garibaldi BT, *et al.* The diagnostic performance feedback "Calibration Gap": why clinical experience alone is not enough to prevent serious diagnostic errors. *AEM Educ Train* 2018;2:339–42.
 - 29 Meyer AND, Singh H. Calibrating how doctors think and seek information to minimise errors in diagnosis. *BMJ Qual Saf* 2017;26:436–8.
 - 30 Norman G. Clinical practice, deliberate practice, and "big data". *Adv in Health Sci Educ* 2018;23:863–6.
 - 31 Hatala RM, Brooks LR, Norman GR. Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Adv Health Sci Educ* 2003;8:17–26.
 - 32 Anders Ericsson K, Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med* 2008;15:988–94.