

BMJ Open Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany

Mirjam Annina Jenny,¹ Niklas Keller,² Gerd Gigerenzer¹

To cite: Jenny MA, Keller N, Gigerenzer G. Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany. *BMJ Open* 2018;**8**:e020847. doi:10.1136/bmjopen-2017-020847

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-020847>).

Received 29 November 2017
Revised 16 May 2018
Accepted 3 July 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Harding Center for Risk Literacy, Max Planck Institute for Human Development, Berlin, Germany

²Department of Anesthesiology and Operative Intensive Care Medicine, Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Berlin, Germany

Correspondence to

Dr Mirjam Annina Jenny;
jenny@mpib-berlin.mpg.de

ABSTRACT

Objectives To assess minimal medical statistical literacy in medical students and senior educators using the 10-item Quick Risk Test; to assess whether deficits in statistical literacy are stable or can be reduced by training.

Design Prospective observational study on the students, observational study on the university lecturers.

Setting Charité University Medicine medical curriculum for students and a continuing medical education (CME) course at a German University for senior educators.

Participants 169 students taking part in compulsory final-year curricular training in medical statistical literacy (63% female, median age 25 years). Sixteen professors of medicine and other senior educators attending a CME course on medical statistical literacy (44% female, age range=30–65 years).

Interventions Students completed a 90 min training session in medical statistical literacy. No intervention for the senior educators.

Outcome measures Primary outcome measure was the number of correct answers out of four multiple-choice alternatives per item on the Quick Risk Test.

Results Final-year students answered on average half (median=50%) of the questions correctly while senior educators answered three-quarters correctly (median=75%). For comparison, chance performance is 25%. A 90 min training session for students increased the median percentage correct from 50% to 90%. 82% of participants improved their performance.

Conclusions Medical students and educators do not master all basic concepts in medical statistics. This can be quickly assessed with the Quick Risk Test. The fact that a 90 min training session on medical statistical literacy improves students' understanding from 50% to 90% indicates that the problem is not a hard-wired inability to understand statistical concepts. This gap in physicians' education has long-lasting effects; even senior medical educators could answer only 75% of the questions correctly on average. Hence, medical students and professionals should receive enhanced training in how to interpret risk-related medical statistics.

INTRODUCTION

For healthcare to be effective, medical professionals require literacy in health, the healthcare system and medical statistics.

Strengths and limitations of this study

- The Quick Risk Test is the first test to measure minimal medical statistical literacy in physicians across disciplines.
- Only a single site was included in each study.
- A large student population was tested (N=169; ~60% of a cohort).
- Only a small population of senior educators was tested.
- No parallel instruments for convergent validity were tested at the same time.

Health literacy entails basic knowledge about diseases and the ability to identify trustworthy medical and health information. Similarly, health system literacy entails basic knowledge of the healthcare system, the incentives that different players face and the effect that these can have on care (eg, defensive medicine). Finally, medical statistical literacy entails the ability to critically assess the numbers that are communicated in health information as well as basic statistical knowledge (eg, understanding of false-negative rates and false-positive rates).¹

Recent efforts to improve healthcare delivery have focused on decisional aspects rather than on health and medical statistical literacy. For example, physicians are urged to ensure that their care is in line with patients' values and to transfer control over their patients' lives to the patients themselves.² This process, however, is impeded by many patients' low health and statistical literacy.^{3 4} Accordingly, other publications have stressed that physicians need to be aware of their patients' low levels of health literacy and numeracy and should take measures to ensure that patients understand what is communicated to them. At the same time, institutions are called to provide rigorously developed medical information formats that

are based on evidence-based communication principles for physicians and patients.⁵

These are all crucial points that need to be addressed, yet they overlook one critical issue. Discussions about patient values require that physicians understand medical statistics, including the nature and likelihood of benefits and harms of diagnostic, intervention or treatment options, as well as the rates at which tests produce false results and the subsequent interpretation of positive and negative test results. More broadly, a healthcare system in which decisions are based on scientific evidence needs medical students and physicians who are literate in medical statistics. Physicians may well have high levels of health literacy and health system literacy yet an insufficient level of statistical literacy.⁶ The few studies to have addressed physicians' statistical literacy indicate that many do not understand key concepts and can be manipulated by misleading statistical formats.^{1 6-8} For instance, only 21% of 160 gynaecologists in one study could correctly name the positive predictive value of a screening mammogram.⁶ A recent study of obstetricians and gynaecologists found low statistical literacy in these groups.⁹ In the absence of statistical literacy, physicians' recommendations can be influenced by framing (eg, mortality vs survival rates) or intransparent risk measures (eg, relative risks). Thus, physicians lacking minimal medical statistical literacy cannot provide the best care to their patients. There is a debate whether lack of statistical literacy in laypeople and experts is something that we must live with or whether it can be overcome by training, just as the inability to read and write can be overcome by education. For instance, Thaler and Sunstein¹⁰ argue that statistical errors are as stable as visual illusions and thereby justify governmental paternalism, popularly known as 'nudging'. Gigerenzer,¹¹ on the other hand, argues that statistical errors can be substantially reduced by training and thereby calls for enhancing statistical literacy by means of educational programmes in schools and medical curricula.

However, although frugal instruments exist to measure numeracy⁴ and minimal medical knowledge,¹² low-threshold, easily applicable and scalable tools for assessing medical statistical literacy are currently not available. One available instrument measures statistical literacy in obstetricians and gynaecologists and includes items that are limited to these professional groups, such as questions about the base rate of specific illnesses.¹³ To fill this gap, we provide a test that is applicable to all professional groups in healthcare: the Quick Risk Test. In this test, we define 10 elementary medical statistical concepts needed for evaluating medical tests, treatments and interventions as well as their results, which constitute what we call minimal medical statistical literacy. Medical statistics that concern patients are mostly related to medical testing. Thus, the 10 concepts were chosen to cover a basic understanding of medical testing (understanding sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), prevalence and Bayesian reasoning), and medical testing in

screening (risk reduction, mortality rate, lead-time bias and overdiagnosis bias). Some of these concepts such as absolute and relative risk reduction are relevant for medical interventions more broadly. Note that one solution for computing the positive predictive value requires Bayes theorem, which is challenging to apply. Another, much simpler solution requires the application of natural frequencies and natural frequency trees. This solution only requires a few simple mathematical computations and serves as a simple strategy for Bayesian reasoning that can be taught easily.¹⁴ This strategy is taught during our short training session on medical statistics. We then present the 10-item multiple-choice Quick Risk Test and apply it to both final-year medical students and professors, senior physicians and university lecturers of medicine in order to measure their levels of medical statistical literacy. Finally, we address the question of whether literacy in medical statistics can be efficiently taught in a 90 min intervention for medical students. In sum, we present the Quick Risk Test as a frugal tool to measure medical professionals' minimal statistical literacy and show that this type of literacy can be increased simply by a short training session. Consequently, we advocate that more effort and resources be channelled into improving these skills.

METHOD

The Quick Risk Test (table 1) measures understanding of 10 central concepts: sensitivity, specificity, PPV, NPV, prevalence, Bayes rule, relative risk, mortality rate, lead time-bias and overdiagnosis bias (table 1). Questions were constructed in multiple-choice format to reflect standard medical assessment and enable quick scoring. The test was administered to two groups: medical students and professionals engaged in teaching. We focused on these two groups to identify possible gaps both in the medical school curriculum and in physicians' continuous education. Proficient teachers are obviously the first step towards enhancing medical statistical literacy; thus, we wanted to avoid missing knowledge gaps in that group. Any such gaps indicate that medical school curricula and continuing education programmes need to be adapted. First, the test was administered over the course of a week in the summer semester of 2016 to 169 medical students (~60% of the semester cohort) in the final year of medical studies at the Charité University Medicine in Berlin. The course is a compulsory part of the medical curriculum, but participation in the test was voluntary and anonymous; students did not have to provide reasons for not participating or dropping out. This group received the Quick Risk Test before and after a 3-hour course on evidence-based medicine, the first 90 min of which deal with risk literacy and diagnostic risk assessment followed by 90 min training on extraction and communication of medical evidence from scientific articles. During the training session on risk literacy, students were taught two tools: *natural frequency trees*, to facilitate the calculation of PPVs and NPVs and *PPV/NPV curves*, to enhance understanding of the interplay between PPV/

Table 1 Quick Risk Test

| Question | Possible answers |
|--|--|
| 1. A test's sensitivity is a central criterion for its quality as a diagnostic tool. The sensitivity is | A) the proportion of people with a positive test result among those who are sick. *** B) the proportion of people with a negative test result among those who are sick. C) the proportion of people with a positive test result among those who are healthy. D) the proportion of people with a negative test result among those who are healthy. |
| 2. A test's specificity is a central criterion for its quality as a diagnostic tool. The specificity is | A) the proportion of people with a positive test result among those who are sick. B) the proportion of people with a negative test result among those who are sick. C) the proportion of people with a positive test result among those who are healthy. D) the proportion of people with a negative test result among those who are healthy. *** |
| 3. Which test characteristic quantifies the probability that a person with a positive test result actually has the disease? | A) Positive predictive value *** B) Negative predictive value C) Specificity D) Sensitivity |
| 4. Which test characteristic quantifies the probability that a person with a negative test result does not have the disease? | A) Sensitivity B) Positive predictive value C) Negative predictive value *** D) Sensitivity |
| 5. A medical test's manufacturer tells you the sensitivity and the specificity of its test. You would like to tell your patient the probability that they are sick if they have a positive test result. Which measurement do you need for your calculation? | A) Mortality B) Prevalence *** C) Coherence D) Latency |
| 6. Mammography is often used as a screening-test to detect breast cancer early. The probability that a woman has breast cancer is 1%. When a woman has breast cancer her probability of receiving a positive mammogram is 90%. When a woman does not have breast cancer her probability of nevertheless receiving a positive mammogram is 9%. What is the best estimate for the number of women with a positive screening mammogram who actually have breast cancer? | A) 9 in 10 B) 8 in 10 C) 1 in 10 *** D) 1 in 100 |
| 7. In a medical publication you read that screening with mammography lowers the probability of dying from breast cancer by 20%. This number is | A) a relative risk reduction. *** B) an absolute risk reduction. C) a specific risk reduction. D) an evident risk reduction. |
| 8. A patient asks you about the benefits of cancer screening. Which criterion should you consider here? | A) 5-year survival rate B) Incidence C) Mortality rate *** D) Prevalence |
| 9. Imagine two groups of people who all die of cancer at age 70. In group A, cancer is detected via screening at the age of 60. In this group, the 5-year survival rate is 100%. Group B is not screened. In this group, cancer is detected at age 68. Everyone dies at age 70. Thus, the 5-year survival rate is 0%. Which bias explains why both groups have different 5-year survival rates? | A) Selection bias B) Overdiagnosis bias C) Lead-time bias *** D) Performance bias |
| 10. A higher screening rate results in more positive diagnoses. In screening, if anomalies are discovered, which because of their extremely slow growth would never cause symptoms or an early death, this is called | A) selection bias. B) attrition bias. C) lead-time bias. D) overdiagnosis bias. *** |

Questions and multiple-choice answers of the 10-item Quick Risk Test (***) denotes the correct answer).

NPV, sensitivity, specificity and prevalence. This training session consisted of a 15min theoretical introduction, a 45min small-group exercise in which students calculated the PPV/NPV of four commonly used diagnostic procedures (sigmoidoscopy/HIVcombined test/neckfold-test/

amniocentesis) and a subsequent 30min discussion on the numerical and ethical implications of diagnostic risk assessment. The 90min intervening task consisted of training on how to extract evidence from medical articles using the *PICO method* and then translate this information into *fact*

boxes for transparent patient communication. All students completed the pretest, but 65 (38.5%) did not complete the post-test.

The test was also administered to 16 university professors, senior physicians and lecturers in medicine, all with a special interest in medical education (referred to as senior educators below) in a continuing medical education (CME) workshop at a German Faculty of Medicine held in October 2017. This group was tested only at the beginning of the workshop and participants were therefore not specifically trained on the topic by us. Participation was also voluntary in this group. In both groups, participation in the test was not required in order to receive the university credits or CME points that could be earned by participating in the courses. All students and educators were asked whether they would like to participate, meaning that both the student and the senior educator group were convenience samples. Both groups gave informed consent before participation.

Patient and public involvement

Neither the patients nor the public were involved in these studies since it concerns medical students and medical professionals.

Data analysis

The data were mainly descriptively analysed using percentages, medians, ranges and IQRs. The item discrimination index (point-biserial correlation) was calculated to test whether the items discriminated between students of different performance levels. Finally, inferential statistics were used in the form of χ^2 tests to test for group differences.

RESULTS

Among the students, 62.5% were female with a median age of 25 years (IQR=24–26) and 61.5% (n=104) completed both pretest and post-test. Among the senior educators, 44% were female with an age range of <30–65 years. Among the senior educators, we only asked participants to give age ranges in order to grant anonymity in the rather small sample. Neither group had any missing data. Final-year students answered on average half (median=50%) of the questions correctly. For comparison, chance performance is 25%. The data of students who dropped out were analysed only in the first round of the test. For the student population, the pretest median percentage (n=169) of correct responses across all 10 questions was 53.8% (IQR=44.4%–68.5%). Questions 6 and 8 (Bayes rule/mortality rate as measure of screening-success) obtained the fewest correct answers (22.5% and 17.2% correct), even below chance performance (25% with four multiple-choice answers). By contrast, questions 1 and 7 (sensitivity/relative risk reduction) obtained the highest number of correct answers (79.3% and 85.2% correct) (figure 1). In the student data before the training, the Quick Risk Test's median item discrimination index was 0.23 (IQR=0.14–0.28). Three questions had values below 0.2, which are considered low indices (question 5=0.10; question 8=0.11; question 10=0.10). All other questions had values between 0.20 and 0.40. The item discrimination index was calculated as the point-biserial correlation between a question's score and the total score and indicates the extent to which an item discriminated between students with higher and lower total scores. Note that a high discrimination index (high homogeneity) is not the goal when concepts are not dependent. The proportion

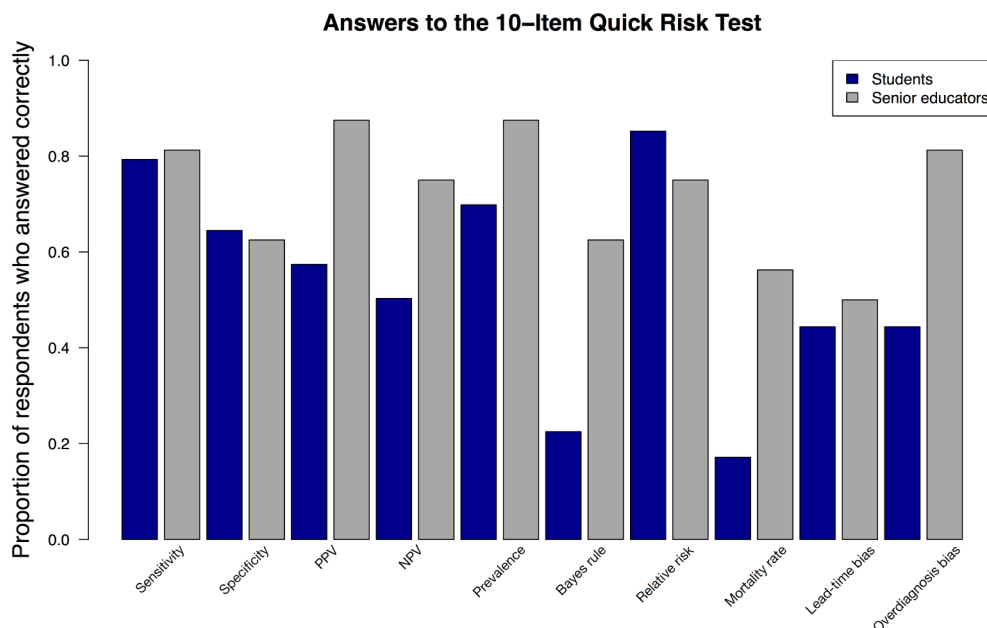


Figure 1 The proportion of correct answers to each of the 10 questions in the Quick Risk Test, for final-year medical students as well as professors, senior physicians and university lecturers. The test measures minimal medical statistical literacy, as defined by understanding 10 basic concepts. PPV, positive predictive value; NPV, negative predictive value.

of students who answered the questions correctly before receiving training did not differ between those students who took the test twice and those who only took it once (median difference in correct answers per question=6.2%, $\chi^2=0.8$, $df=1$, $p=0.4$).

Senior educators answered on average three-quarters of the questions correctly (median=75%). Among the senior educators, the median percentage correct across all 10 questions was 75% (IQR=62.5%–81.2%). **Figure 1** compares the group of senior educators with that of the students before training. On three of the questions—sensitivity, specificity and lead-time bias—students responded about as accurately as senior educators did. On the question of relative risk, students performed even somewhat better. The most difficult concepts for senior educators were mortality rate as opposed to 5-year survival rates (question 8), and lead-time bias (question 9) as a measure of the benefit of screening. Note that even the senior educators were not sure about the meaning of all 10 basic concepts; for instance, only 81% could identify the correct definition of sensitivity, and only 63% the correct definition of specificity. Question 9 (lead-time bias) was the most difficult (50% correct) and questions 3 (PPV) and 5 (prevalence necessary to compute the PPV) were the easiest (88% correct).

The students ($n=104$), but not the senior educators, then completed a 90 min training session on medical statistical literacy as part of the medical curriculum of the Charité University Medicine. The training session increased the median percentage correct from 50% to 90%. Eighty-two per cent of participants improved their performance. After the 90 min session (and an unrelated task of another 90 min), their performance improved to a median of 92.3% (IQR=83.2%–94.2%)

correct answers per question ($\chi^2=300$, $df=1$, $p<2e-16$). Additionally, each question obtained more correct answers after training, even the question with the smallest pretest difference in proportion correct answers, namely question 7 on relative risk ($\chi^2=7$, $df=1$, $p=0.004$); 81.7% of the students performed better after the training than beforehand. Whereas question 6 on estimating the PPV for mammography screening (using Bayes rule) showed substantial improvement, from 22.5% to 87.5% correct answers, question 8 (46.2% correct) on the appropriate measure of screening success (mortality rate, not 5-year survival rate) still proved to be the most difficult one. The lead-time bias and the overdiagnosis bias also were among the more difficult concepts to understand.

Both students and senior educators struggled with applying Bayes rule to identify the PPV of a diagnostic test and with concepts relevant to screening, including the lead-time bias, overdiagnosis and identifying mortality rates as the most informative criterion to quantify the benefits of screening programmes. The training session for students included teaching how to use natural frequencies instead of conditional probabilities (such as sensitivity), an effective method for understanding how to calculate the PPV.¹⁴ **Figure 2** shows the strong effect of this part of the training, with students reaching an average of close to 90% correct, compared with only about 60% among the senior educators who did not receive training by us (**figure 1**).

DISCUSSION

The Quick Risk Test presented here measures minimal medical statistical literacy as defined by the 10

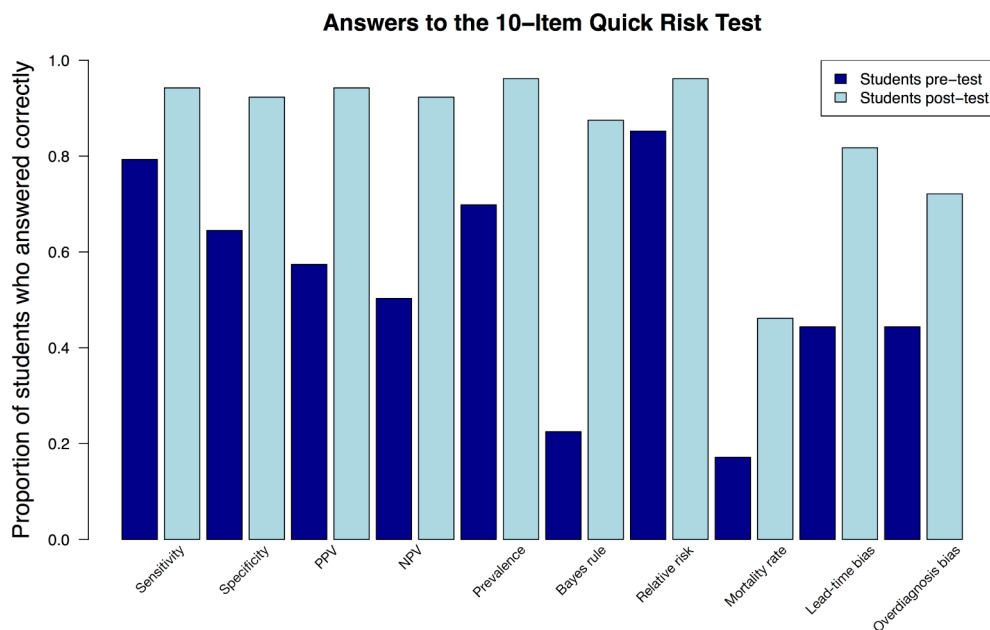


Figure 2 The proportion of correct answers to each of the 10 questions in the Quick Risk Test, for final-year medical students before and after a 90 min training session in risk literacy and diagnostic risk assessment. NPV, negative predictive value; PPV, positive predictive value.

elementary concepts. It can also be used to track performance improvement in risk literacy training. In contrast to claims that lack of statistical literacy is something we must live with, the present study shows the encouraging result that final-year medical students can greatly improve their understanding of medical statistics in as little as 90 min. Note that the training took place a week prior to the students' final year-exams without being relevant to these exams. Student engagement was increased by using real tests selected from areas of medicine taught in the final semester (eg, gynaecology), and dedicating the majority of the session to practice and discussion of the implications.

Although most questions and the test as a whole are able to discriminate between different levels of proficiency, this is not the main goal. Students and professionals should be able to answer all of the questions correctly and thereby demonstrate understanding of the 10 basic concepts that comprise *minimal* medical statistical literacy. The Quick Risk Test can identify knowledge gaps and track progress in medical statistical literacy. Instead of ranking students, the goal is thus to identify knowledge gaps that then have to be addressed immediately.

These results concern single-site studies with voluntary participation and thus risk of selection bias. The student study did, however, assess performance on over 50% of that year's student cohort in the final year of studies. Nevertheless, it is an empirical question whether our results generalise to other student cohorts, which will depend on students' statistical training in individual medical schools. In German-speaking Europe, statistical literacy is very rarely taught in medical school. We therefore expect similar results for other sites including students' promising learning rate. Further validation samples in different educational systems are planned for future studies.

One limitation of our study on the student population is that it looked solely at a retention interval of 90 min. However, the fact that students practised the use of the tools (natural frequency trees and PPV/NPV curves) on actual tests using their real statistical properties supports long-term retention of these tools. With regard to natural frequency trees, studies showed that high application accuracy is maintained in a non-medical population after up to 3-month follow-up.¹⁵ No evidence for long-term retention of PPV/NPV curves currently exists. Our studies did not measure how minimal medical statistical literacy affects outcomes. However, a national survey in the USA suggests that physician's understanding of medical statistics affects their recommendations.⁷ Finally, in contrast to other studies that have looked at statistical literacy of specific subdisciplines,⁹ the Quick Risk Test is the first test to measure minimal medical statistical literacy in physicians across disciplines.

Medical statistical literacy is insufficient among medical students¹⁶ and professionals, even those active in teaching medicine. The generally low understanding of the screening-related concepts may also be due to the widespread

use of misleading information in health pamphlets and publications, such as 5-year survival rates to communicate the supposed benefits of screening.¹⁶⁷ The fact that almost 20% of medical professors and lecturers could not identify the correct definition of sensitivity and 40% could not correctly identify the definition of specificity highlights the need for more rigorous training in medical schools and in physicians' continuous medical education programmes. As we have shown, just 90 min of training on medical statistical literacy can make a big difference. We urge medical schools and organisers of CME to include medical statistical literacy in their curricula so that physicians can become fully competent in assessing medical risks. Because the test is geared towards assessing basic medico-statistical knowledge in medical practitioners, additional tools would have to be developed to educate and test patients. Future research should concern the validation of the Quick Risk Test with other tools such as numeracy tests and with other groups such as student groups in other medical schools.

Acknowledgements The authors would like to thank Clara Schirren for testing the senior educators and Jana Hinneburg, Felix Rebitschek and Anna Held for their input concerning the wording of the items in the Quick Risk Test.

Contributors MAJ and GG developed the Quick Risk Test. MAJ analysed the data and wrote the manuscript. NK developed and ran the intervention study with the students and revised the manuscript. GG ran the study with the senior educators and revised the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent Not required.

Ethics approval The study protocol for the students was approved by the Charité University Medicine's ethics committee (ID: EA4/067/15) and the study protocol for the senior educators was approved by the ethics committee at the Max Planck Institute for Human Development (ID 19102017).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional unpublished data are available from the study.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Gigerenzer G, Gray JAM. Launching the century of the patient. In: Gigerenzer G, Gray J, eds. *Better doctors, better patients, better decisions*. California: Better doctors, 2011.
- Berwick DM. Era 3 for medicine and health care. *JAMA* 2016;315:1329–30.
- Mata J, Frank R, Gigerenzer G. Symptom recognition of heart attack and stroke in nine European countries: a representative survey. *Health Expect* 2014;17:376–87.
- Cokely ET, Galesic M, Schulz E, et al. Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making* 2012;7:25–47.
- Koh HK, Rudd RE. The Arc of Health Literacy. *JAMA* 2015;314:1225–6.
- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, et al. Helping Doctors and Patients Make Sense of Health Statistics. *Psychol Sci Public Interest* 2007;8:53–96.

7. Wegwarth O, Schwartz LM, Woloshin S, *et al.* Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med* 2012;156:340–9.
8. Herzog SM, Lim HW, Williams MS, *et al.* Sun Protection Factor Communication of Sunscreen Effectiveness: A Web-Based Study of Perception of Effectiveness by Dermatologists. *JAMA Dermatol* 2017;153:348–50.
9. Anderson BL, Williams S, Schulkin J. Statistical literacy of obstetrics-gynecology residents. *J Grad Med Educ* 2013;5:272–5.
10. Thaler RH, Sunstein CR. *Nudge*. Yale: Yale University Press, 2008.
11. Gigerenzer G. *Risk Savvy: How to make good decisions*. New York: Viking Penguin, 2014.
12. Bachmann LM, Gutzwiller FS, Puhan MA, *et al.* Do citizens have minimum medical knowledge? A survey. *BMC Med* 2007;5:14.
13. Anderson BL, Gigerenzer G, Parker S, *et al.* Statistical literacy in obstetricians and gynecologists. *J Healthc Qual* 2014;36:5–17.
14. McDowell M, Jacobs P. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychol Bull* 2017;143:1273–312.
15. Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. *J Exp Psychol Gen* 2001;130:380–400.
16. Hoffrage U, Lindsey S, Hertwig R, *et al.* Medicine. Communicating statistical information. *Science* 2000;290:2261–2.