BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Measuring Minimal Medical Statistical Literacy Using the Quick Risk Test

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020847
Article Type:	Research
Date Submitted by the Author:	29-Nov-2017
Complete List of Authors:	Jenny, Mirjam; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy Keller, Niklas; Charité Universitätsmedizin Berlin, Department of Anesthesiology and Operative Intensive Care Medicine Gigerenzer, Gerd; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Communication
Keywords:	MEDICAL EDUCATION & TRAINING, STATISTICS & RESEARCH METHODS, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Measuring Minimal Medical Statistical Literacy Using the Quick Risk Test

Mirjam Jenny, Niklas Keller, Gerd Gigerenzer

Word Count (main body of text): 1606

Mirjam Jenny (Corresponding author)
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany
Phone: +49 030 82406-720
E-mail: jenny@mpib-berlin.mpg.de

Niklas Keller, Dr. rer. nat.
Department 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

Gerd Gigerenzer, Dr. phil., Professor
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany

Abstract

Objectives: To assess minimal medical statistical literacy in medical students and senior educators using a 10-item Quick Risk Test; to assess whether deficits in statistical literacy are stable or can be reduced by a 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 course at Leipzig University for educators.

Participants: 169 students taking part in compulsory final-year curricular training on medical statistical literacy (63% female, median age 25 years, interquartile range (IQR) = 24 – 26). Sixteen senior educators attending a CME training on medical statistical literacy (44% female, age range = 30 – 65 years).

Interventions: Students received a 1.5-hour training 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 could answer on average half (median = 50%, IQR = 40% – 70%) of the questions correctly while senior educators answered three-quarters correctly (median = 75%, IQR = 60% – 90%). For comparison, chance performance is 25%. A 90-minutes training for students increased the median percentage correct from 50% to 90% (80% – 100%). 82% of participants improved their performance.

Conclusions: On average, students in their final year could only answer half of the questions, while senior educators could answer on average 75% of the questions correctly. The fact that a 90-min training improves students' understanding from 50% to 90% shows that the problem is not a hard-wired inability to understand statistical concepts, but a lack of training at medical universities. It is time that medical students and professionals are taught how to think about risks.

Strength and Limitations of this Study:

- Administration and completion of the test require only 10-minutes
- The test tracks improvement after a 1.5-hour training intervention
- No parallel instruments for convergent validity were tested at the same time
- While a large student population was tested, only a small population of senior educators was tested.

Keywords:
Medical education & training, Statistics & Research Methods, Risk Management

Introduction

Effective healthcare requires health literacy, health system literacy, and medical statistical literacy. Health literacy entails basic knowledge about diseases and the ability to identify trustworthy medical and health information. Health system literacy necessitates basic knowledge of the healthcare system, the incentives that different players face, and the effect that those can have on care (e.g., 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 (e.g., understanding of false negative rates and false positive rates).[1]

Recent efforts to improve healthcare delivery have focused on the decisional aspects rather than health and medical statistical literacy. For example, physicians should ensure that their care is in line with patients' values and that they should transfer the control over their patients' lives to the patients themselves.[2] Patients' notoriously low health and statistical literacy impedes this, however. Accordingly, other publications have stressed that physicians need to be aware of their patients' low health literacy and numeracy and that they should take measures to ensure that patients understand what is communicated to them. At the same time, institutions should provide rigorously developed medical information formats that are based on evidence-based communication principles for physicians and patients.[3]

While these are all crucial points that need to be addressed they overlook one critical issue. Discussions about value 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. Physicians may have a high health literacy and health system literacy. However, their level of statistical literacy may not be as required. There is a debate whether lack of statistical literacy is something that we must live with, or whether it can be overcome by training, just like lack of literacy can be overcome by education. For instance, Thaler & Sunstein (2008)[4] argue that statistical errors are stable like visual illusions, calling for governmental paternalism, known as "nudging", as the solution. Gigerenzer (2014)[5], on the other hand, argues that statistical errors can be substantially reduced by training, calling to enhance statistical literacy by means of educational programs in schools and medical curricula as the solution. However, while there are frugal instruments available to measure numeracy[6] and minimal medical knowledge[7], low-threshold, easily applicable, and scalable tools for assessing medical statistical literacy are currently not available. To fill this gap, we define ten elementary medical statistical concepts necessary to evaluate medical tests, treatments, and interventions as well as their results, which constitute what we call minimal medical statistical literacy. We also present a 10-item multiple-choice test, the Quick Risk Test, and apply it to final-year medical students as well as to professors, senior physicians, and university lecturers of medicine in order to measure the present level of understanding. Finally, we address the question of whether literacy in medical statistics can be efficiently taught in a 90-minutes intervention for medical students.

Method

The Quick Risk Test (see Appendix) measures understanding of ten central concepts: sensitivity, specificity, positive predictive value, negative predictive value, prevalence, Bayes rule, relative risk, mortality rate, lead time-bias, and overdiagnosis bias (see Appendix 1). Questions were constructed in multiple-choice format to reflect standard medical assessment and allow quick scoring. The test was administered to two groups: First, to 169 medical students at the Charité University Medicine in Berlin, just prior to a training on medical statistical literacy in the final year of medical studies. This group received the same test before and after a 90-minutes training in risk literacy and diagnostic risk assessment, including a 1.5-hour unrelated intervening task. Second, the test was 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) training at the Faculty of Medicine of Leipzig University. Participation in the test was voluntary in both groups and was not required to receive the university credits or CME points that could be earned by participating in the trainings. Among the students, 62.5% were female with a median age of 25 years (interquartile range (IQR) = 24 – 26) and 61.5% (N = 104) completed both pre-and post-tests.

Among the senior educators, 44% were female with an age range of <30 – 65. 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).

Results

For the student population, the pre-test median percentage 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) got the fewest correct answers (22.5% and 17.2% correct), even below chance performance (25% with four multiple-choice answers). Questions 1 and 7 (sensitivity / relative risk reduction) got most correct answers (79.3% and 85.2% correct). 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 to that of the students before training. On three of the questions – sensitivity, specificity, and lead-time bias, students did about as well as senior educators. On the question of relative risk, students even did somewhat better. The most difficult concepts for senior educators were mortality rate as a measure of the benefit of screening, as opposed to 5-year survival rates (Question 8), and lead-time bias (Questions 9). Note that even the senior educators were not sure what all of the 10 basic concepts meant; for instance, only 81% understood what sensitivity means, and only 63% what specificity means. Question 9 (lead-time bias) was the hardest (50% correct) and questions 3 (positive predictive value) and 5 (prevalence necessary to compute the positive predictive value) were the easiest (88% correct).

The students, but not the senior educators, then undertook a 90-minute training on medical statistical literacy as part of the medical curriculum of the Charité University Medicine. After the 90-minute training (and an unrelated task of another 90 minutes), the performance improved to a median of 92.3% (IQR = 83.2% – 94.2%) correct answers per question. 81.7% of the students performed better after the training than before. While 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-suvival 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.

Discussion

A health system in which decisions are based on scientific evidence requires that medical students are trained and physicians are literate in medical statistics. However, the few studies that have addressed physicians' statistical literacy indicate that many do not understand key concepts and can be manipulated with misleading statistical formats.[1,8-10] For instance, only 21% of 160 gynaecologists in one study could correctly name the positive predictive value of a screening mammogram.[9] In the absence of statistical literacy, physicians' recommendations can be influenced by framing (e.g., mortality vs. survival rates) or intransparent risk measures (e.g., relative risks). Thus, even with high health and health system literacy, physicians lacking minimal medical statistical literacy cannot provide the best care to their patients. Consequently, more effort and resources need to be channelled to improve these skills.

Both students and senior educators struggled greatly with applying Bayes' rule to identify the positive predictive value 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–programs. 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.[11] Figure 2 shows the strong effect of this part of the training, reaching an average of close to 90% correct, compared to only about 60% among the senior educators (Figure 1). The generally low understanding of the screening-related concepts may 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.[1,5-7]

The Quick Risk Test presented here measures minimal medical statistical literacy, as defined by the ten concepts. It can also be used to track performance improvement of a risk literacy training. The finding 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 contrast to claims that lack of statistical literacy is something we must live with, the present study shows the encouraging result that last-year medical students can greatly improve their understanding of medical statistics in as little as 90 minutes. The Quick Risk Test can identify knowledge gaps and track progress in medical statistical literacy.

Contributorship Statement
MJ and GG developed the Quick Risk Test. MJ 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.

Acknowledgments: We thank Clara Schirren for testing the senior educators and we thank her as well as Jana Hinneburg and Felix Rebitschek for their input concerning the wording of the items in the Quick Risk Test.

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

Competing Interests Statement
The authors declare that they have no competing interests as a result of associations with commercial entities or similar financial associations involving spouse or children.

Data Sharing Statement
No additional unpublished data are available from the study.

Figure Legends

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

Figure 2. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students before and after a 90-minute training in risk literacy and diagnostic risk assessment.

Reference list

1 Gigerenzer G, Gray J. Launching the century of the patient. In: Gigerenzer G, Gray J, eds. *Better doctors, better patients, better decisions*. The MIT Press 2011.

2 Berwick DM. Era 3 for medicine and health care. *JAMA : The journal of the American Medical Association* 2016;**315**:1329–30. doi:10.1001/jama.2016.1509

3 Koh HK, Rudd RE. The Arc of Health Literacy. *JAMA : The journal of the American Medical Association* 2015;**314**:1225–6. doi:10.1001/jama.2015.9978

4 Thaler RH, Sunstein CR. *Nudge*. Yale University Press 2008.

5 Gigerenzer G. *Risk Savvy: How to make good decisions*. New York: Viking Penguin 2014.

6 Cokely ET, Galesic M, Schulz E, *et al.* Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making* 2012;**7**:25–47.

7 Bachmann LM, Gutzwiller FS, Puhan MA, *et al.* Do citizens have minimum medical knowledge? A survey. *BMC Med* 2007;**5**:14. doi:10.1186/1741-7015-5-14

8 Wegwarth O, Schwartz LM, Woloshin S. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Annals of Internal Medicine* 2012.

9 Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al.* Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest* 2007;**8**:53–96. doi:10.1111/j.1539-6053.2008.00033.x

10 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 Dermatology* 2017;**153**:348–50. doi:10.1001/jamadermatol.2016.4924

11 McDowell M, Jacobs P. Meta-Analysis of the Effect of Natural Frequencies on Bayesian Reasoning. *Psychological Bulletin* Published Online First: 19 October 2017. doi:10.1037/bul0000126

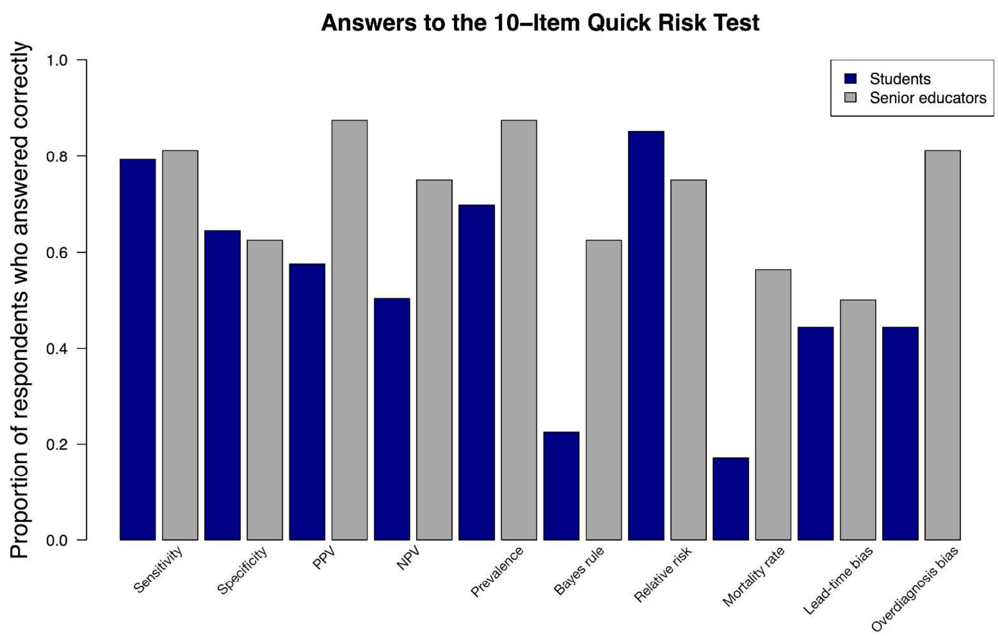

Figure 1. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students as well as professors, senior physicians, and university lecturers. The test measures minimal medical statistical literacy, as defined by understanding ten basic concepts. PPV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

361x270mm (300 x 300 DPI)

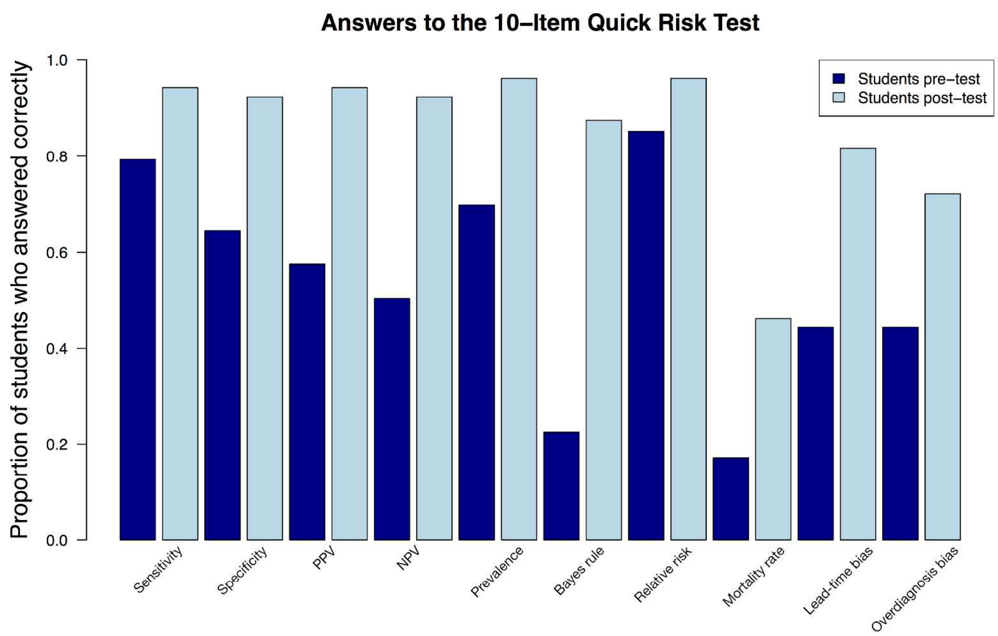

Figure 2. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students before and after a 90-minute training in risk literacy and diagnostic risk assessment.

361x270mm (300 x 300 DPI)

Appendix: 10-item Quick Test (* denotes correct answer)**

1. A test's sensitivity is a central criterion for its quality as a diagnostic tool.
The sensitivity describes:
A) The proportion of sick people who receive a positive rest result. ***
B) The proportion of sick people who receive a negative rest result.
C) The proportion of healthy people who receive a positive test result.
D) The proportion of healthy people who receive a negative rest result.

2. A test's specificity is a central criterion for its quality as a diagnostic tool.
The specificity describes:
A) The proportion of sick people who receive a positive rest result.
B) The proportion of sick people who receive a negative rest result.
C) The proportion of healthy people who receive a positive test result. ***
D) The proportion of healthy people who receive a negative rest result.

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) Prevalence ***
B) Mortality
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. Also in this group everyone dies at age 70. Thus, the 5-year survival rate is 0%. Which bias is used here to describe the benefits of screening?
A) Lead-time bias ***
B) Overdiagnosis bias
C) Selection 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, we call this...
A) Selection bias
B) Attrition bias
C) Lead-time bias
D) Overdiagnosis bias ***

BMJ Open

Can We Measure Minimal Medical Statistical Literacy Using the Quick Risk Test? A Prospective Observational Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020847.R1
Article Type:	Research
Date Submitted by the Author:	15-Mar-2018
Complete List of Authors:	Jenny, Mirjam; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy Keller, Niklas; Charité Universitätsmedizin Berlin, Department of Anesthesiology and Operative Intensive Care Medicine Gigerenzer, Gerd; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Communication, Medical education and training
Keywords:	MEDICAL EDUCATION & TRAINING, STATISTICS & RESEARCH METHODS, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Can We Measure Minimal Medical Statistical Literacy Using the Quick Risk Test? A Prospective Observational Study

Mirjam A. Jenny, Niklas Keller, Gerd Gigerenzer

Word Count (main body of text): 1606

Mirjam Annina Jenny, Dr. phil. (Corresponding author)
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany
Phone: +49 030 82406-720
E-mail: jenny@mpib-berlin.mpg.de

Niklas Keller, Dr. rer. nat.
Department 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

Gerd Gigerenzer, Dr. phil., Professor
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany

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 Leipzig University for 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 1.5-hour 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-minute training session for students increased the median percentage correct from 50% to 90% . 82% of participants improved their performance.

Conclusions: Medical students' and educators' medical statistical literacy is insufficient. This can be quickly assessed with the Quick Risk Test. The fact that a 90-minute training session improves students' understanding from 50% to 90% indicates that the problem is not a hard-wired inability to understand statistical concepts but instead inadequate training at medical universities. This shortcoming 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 need enhanced training in how to interpret risk-related 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.
- A large student population was tested (N=169; ~60% of a cohort).
- Only a single site was included in each study.
- Only a small population of senior educators was tested.
- No parallel instruments for convergent validity were tested at the same time.

Keywords:
Medical Education & Training, Statistics & Research Methods, Risk Management

Introduction

For healthcare to be effective, medical professionals require literacy in health, the healthcare system, and medical statistics. 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 (e.g., 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 (e.g., 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 patients' notoriously low health and statistical literacy. 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.[3]

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 inadequate level of statistical literacy.[4] 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.[1,4-6] For instance, only 21% of 160 gynaecologists in one study could correctly name the positive predictive value of a screening mammogram.[4] A recent study of obstetricians and gynaecologists found low statistical literacy in these groups.[7] In the absence of statistical literacy, physicians' recommendations can be influenced by framing (e.g., mortality vs. survival rates) or intransparent risk measures (e.g., 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 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 & Sunstein (2008)[8] argue that statistical errors are as stable as visual illusions and thereby justify governmental paternalism, popularly known as "nudging." Gigerenzer (2014)[9], 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 programs in schools and medical curricula. However, although frugal instruments exist to measure numeracy[10] and minimal medical knowledge[11], 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.[12] To fill this gap and provide a test that is applicable to all professional groups in health care, 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. We also present a 10-item multiple-choice test, the 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-minute 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 (see Appendix) measures understanding of 10 central concepts: sensitivity, specificity, positive predictive value, negative predictive value, prevalence, Bayes rule, relative risk, mortality rate, lead time-bias, and overdiagnosis bias (see Appendix 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 gaps both in the medical school curriculum and in physicians' continuous education. Proficient teachers are obviously the first step toward enhancing medical statistical literacy; thus we wanted to avoid missing knowledge gaps in that group. Any such gaps indicated that not only medical school curricula but also continuing education programs 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 test was administered just prior to a training course on medical statistical literacy. 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 same test before and after a 90-minute training session in risk literacy and diagnostic risk assessment, including an unrelated 1.5-hour intervening task. During the training session, students were taught two tools: *natural frequency trees*, to facilitate the calculation of positive and negative predictive values and *PPV / NPV-curves*, to enhance understanding of the interplay between PPV / NPV, sensitivity, specificity, and prevalence. The training session consisted of a 15-minute theoretical introduction of these tools, a 45-minute small-group exercise in which students calculated the PPV / NPV of four commonly used diagnostic procedures (sigmoidoscopy / HIV combined test / neck-fold-test / amniocentesis) and a subsequent 30-minute discussion on the numerical and ethical implications of diagnostic risk assessment. The 90-minute 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 pre-test, but 65 (38,5%) did not complete the post-test. Second, the test was 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 the Faculty of Medicine of Leipzig University 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. Among the students, 62.5% were female with a median age of 25 years (interquartile range (IQR) = 24 – 26) and 61.5% (N = 104) completed both pre-and post-tests. Among the senior educators, 44% were female with an age range of <30 – 65. 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). Both groups gave informed consent before participation.

Results

Neither group had any missing data. The data of students who dropped out were analysed only in the first round of the test. For the student population, the pre-test 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). With respect to 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.

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 to 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% understood what sensitivity means, and only 63% what specificity means. Question 9 (lead-time bias) was the most difficult (50% correct) and Questions 3 (positive predictive value) and 5 (prevalence necessary to compute the positive predictive value) were the easiest (88% correct).

The students (N = 104), but not the senior educators, then completed a 90-minute training session on medical statistical literacy as part of the medical curriculum of the Charité University Medicine. After the 90-minute session (and an unrelated task of another 90 minutes), their performance improved to a median of 92.3% (IQR = 83.2% – 94.2%) correct answers per question (X-squared = 300, df = 1, p-value <2e-16). Additionally, each question obtained more correct answers after training, even the question with the smallest pre–post difference in proportion correct answers, namely Question 7 on relative risk (X-squared = 7, df = 1, p-value = 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 positive predictive value 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 programs. 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.[13] Figure 2 shows the strong effect of this part of the training, with students reaching an average of close to 90% correct, compared to only about 60% among the senior educators (Figure 1). 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.[1,5-7]

Discussion

The Quick Risk Test presented here measures minimal medical statistical literacy as defined by the 10 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 minutes. The Quick Risk Test can identify knowledge gaps and track progress in medical statistical literacy.

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. Instead of ranking students, the goal is thus to identify knowledge gaps that then have to be addressed immediately.

One limitation of our study on the student population is that it looked solely at a retention interval of 90 minutes. However, the fact that students practiced 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 regards to natural frequency trees, studies showed that high application accuracy is maintained in a nonmedical population after up to 3-month follow-up [14]. No evidence for long-term retention of PPV / NPV curves currently exists. Furthermore, these are 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. Our studies did not measure how minimal medical statistical literacy affects outcomes. However, a national survey in the US suggests that physician's understanding of medical statistics affects their recommendations[15]. Finally, in contrast to other studies that have looked at statistical literacy of specific subdisciplines[7], the Quick Risk Test is the first test to measure minimal medical statistical literacy in physicians across disciplines.

Medical statistical literacy is clearly inadequate amongst medical students and professionals, even those active in teaching medicine. 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 not only in medical schools but also in physicians' continuous medical education programs. As we have shown, just 90 minutes of training can make a big difference. We urge every medical school and every organizer of CME to include medical statistical literacy in their curricula so that physicians are fully competent in assessing medical risks. Because the test is geared toward 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.

Contributorship Statement

MJ and GG developed the Quick Risk Test. MJ 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.

Acknowledgments: We thank Clara Schirren for testing the senior educators and thank her along with Jana Hinneburg and Felix Rebitschek for their input concerning the wording of the items in the Quick Risk Test.

Funding Statement

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

Competing Interests Statement

The authors declare that they have no competing interests as a result of associations with commercial entities or similar financial associations involving spouse or children.

Data Sharing Statement

No additional unpublished data are available from the study.

Patient and public involvement

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

Figure Legends

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.

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-minute training session in risk literacy and diagnostic risk assessment.

Reference list

1 Gigerenzer G, Gray JAM. Launching the century of the patient. In: Gigerenzer G, Gray JAM, eds. *Better doctors, better patients, better decisions*. Better doctors 2011.

2 Berwick DM. Era 3 for medicine and health care. *JAMA* 2016;**315**:1329–30. doi:10.1001/jama.2016.1509

3 Koh HK, Rudd RE. The Arc of Health Literacy. *JAMA* 2015;**314**:1225–6. doi:10.1001/jama.2015.9978

4 Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al.* Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest* 2007;**8**:53–96. doi:10.1111/j.1539-6053.2008.00033.x

5 Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier, W, Gigerenzer, G. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Annals of Internal Medicine* 2012; **156**:340-349. doi: 10.7326/0003-4819-156-5-201203060-00005

6 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 Dermatology* 2017;**153**:348–50. doi:10.1001/jamadermatol.2016.4924

7 Anderson BL, Williams S, Schulkin J. Statistical Literacy of Obstetrics-Gynecology Residents. *Journal of Graduate Medical Education* 2013;**5**:272–5. doi:10.4300/JGME-D-12-00161.1

8 Thaler RH, Sunstein CR. *Nudge*. Yale University Press 2008.

9 Gigerenzer G. *Risk Savvy: How to make good decisions*. New York: Viking Penguin 2014.

10 Cokely ET, Galesic M, Schulz E, *et al.* Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making* 2012;**7**:25–47.

11 Bachmann LM, Gutzwiller FS, Puhan MA, *et al.* Do citizens have minimum medical knowledge? A survey. *BMC Med* 2007;**5**:14. doi:10.1186/1741-7015-5-14

12 Anderson BL, Gigerenzer G, Parker S, *et al.* Statistical Literacy in Obstetricians and Gynecologists. *Journal for Healthcare Quality* 2012;**36**:5–17.

13 McDowell M, Jacobs P. Meta-Analysis of the Effect of Natural Frequencies on Bayesian Reasoning. *Psychological Bulletin* 2017;:1–41. doi:10.1037/bul0000126

14 Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology General* 2001;**130**:380–400. doi:10.1037//0096-3445.130.3.380

15 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. *Annals of Internal Medicine* 2012;**156**:340–9. doi:10.7326/0003-4819-156-5-201203060-00005

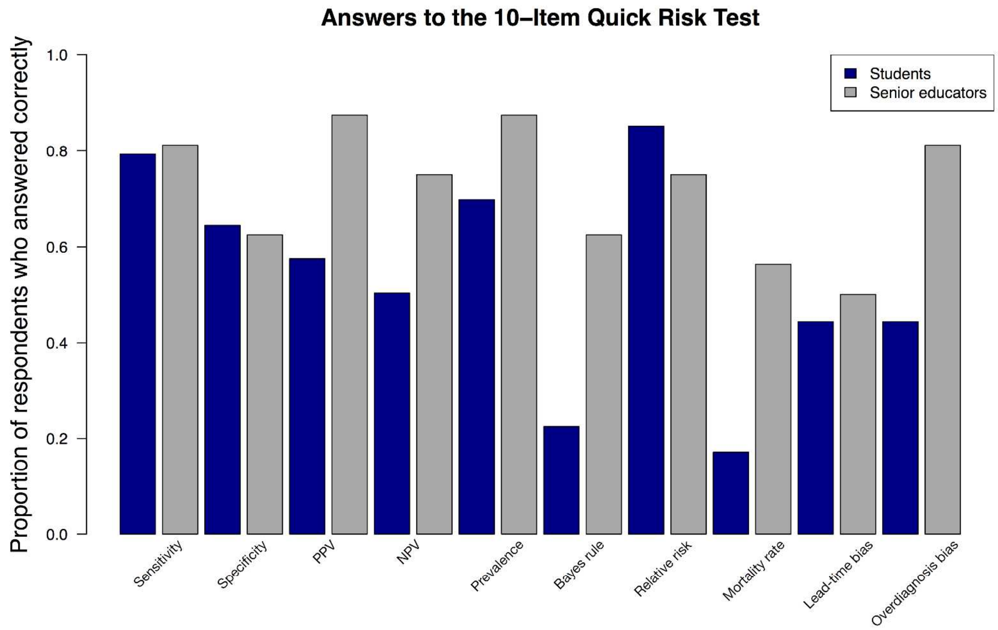

Figure 1. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students as well as professors, senior physicians, and university lecturers. The test measures minimal medical statistical literacy, as defined by understanding ten basic concepts. PPV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

361x270mm (300 x 300 DPI)

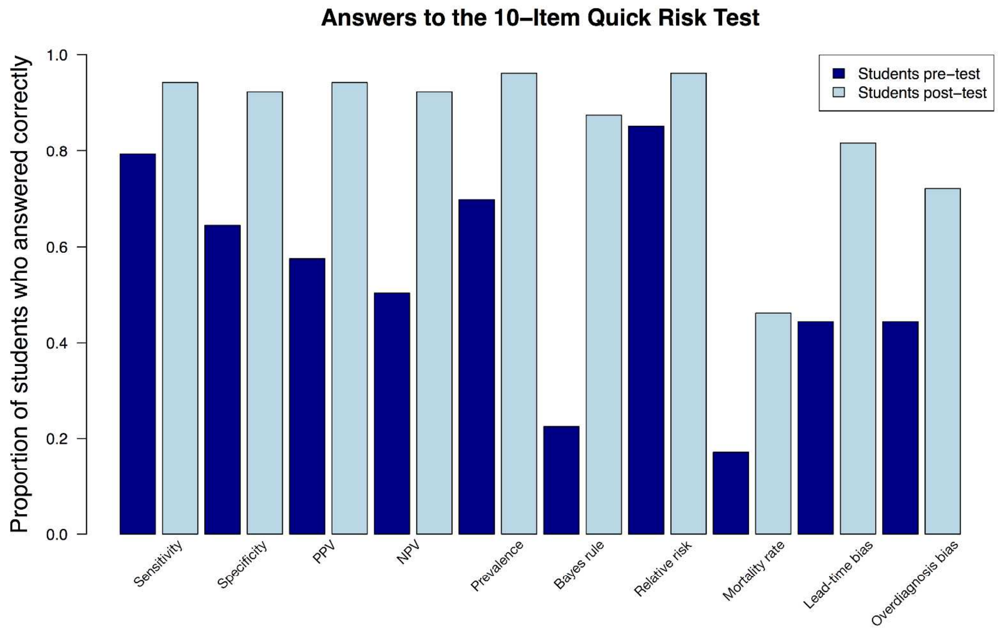

Figure 2. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students before and after a 90-minute training in risk literacy and diagnostic risk assessment.

361x270mm (300 x 300 DPI)

Appendix: 10-item Quick Test (* denotes correct answer)**

1. A test's sensitivity is a central criterion for its quality as a diagnostic tool.
The sensitivity describes:
A) The proportion of sick people who receive a positive rest result. ***
B) The proportion of sick people who receive a negative rest result.
C) The proportion of healthy people who receive a positive test result.
D) The proportion of healthy people who receive a negative rest result.

2. A test's specificity is a central criterion for its quality as a diagnostic tool.
The specificity describes:
A) The proportion of sick people who receive a positive rest result.
B) The proportion of sick people who receive a negative rest result.
C) The proportion of healthy people who receive a positive test result. ***
D) The proportion of healthy people who receive a negative rest result.

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) Prevalence ***
B) Mortality
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. Also in this group everyone dies at age 70. Thus, the 5-year survival rate is 0%. Which bias is used here to describe the benefits of screening?
A) Lead-time bias ***
B) Overdiagnosis bias
C) Selection 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, we call this...
A) Selection bias
B) Attrition bias
C) Lead-time bias
D) Overdiagnosis bias ***

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **Title page and abstract**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **Is provided**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Page 3**
Objectives	3	State specific objectives, including any prespecified hypotheses **Page 3**
Methods		
Study design	4	Present key elements of study design early in the paper **Pages 4**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Pages 4**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **Pages 4**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **Not applicable**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Pages 3 and 4**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Pages 4**
Bias	9	Describe any efforts to address potential sources of bias **Convenience sampling, but large portion of student cohort (~60%), p.4**
Study size	10	Explain how the study size was arrived at **Convenience sample, page 4**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Pages 4 and 5**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **Page 4 and 5**
		(*b*) Describe any methods used to examine subgroups and interactions **Not applicable**
		(*c*) Explain how missing data were addressed **Page 4 and 5**
		(*d*) If applicable, explain how loss to follow-up was addressed **Not applicable**
		(*e*) Describe any sensitivity analyses **Not applicable**

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **Pages 4**
		(b) Give reasons for non-participation at each stage **Page 4**
		(c) Consider use of a flow diagram **Not applicable**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **Page 4**
		(b) Indicate number of participants with missing data for each variable of interest **Page 4**
		(c) Summarise follow-up time (eg, average and total amount) **Page 4**
Outcome data	15*	Report numbers of outcome events or summary measures over time **Page 4 and 5**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **Not applicable**
		(*b*) Report category boundaries when continuous variables were categorized **Not applicable**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **Not applicable**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Not applicable**
Discussion		
Key results	18	Summarise key results with reference to study objectives **Pages 4 and 5**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Page 5 and 6**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Page 5 and 6**
Generalisability	21	Discuss the generalisability (external validity) of the study results **Page 5 and 6**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Page 6**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

Assessing Minimal Medical Statistical Literacy Using the Quick Risk Test
A Prospective Observational Study in Germany

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020847.R2
Article Type:	Research
Date Submitted by the Author:	25-Apr-2018
Complete List of Authors:	Jenny, Mirjam; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy Keller, Niklas; Charité Universitätsmedizin Berlin, Department of Anesthesiology and Operative Intensive Care Medicine Gigerenzer, Gerd; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Communication, Medical education and training
Keywords:	MEDICAL EDUCATION & TRAINING, STATISTICS & RESEARCH METHODS, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Assessing Minimal Medical Statistical Literacy Using the Quick Risk Test
A Prospective Observational Study in Germany

Mirjam A. Jenny, Niklas Keller, Gerd Gigerenzer

Word Count (main body of text): 2969

Mirjam Annina Jenny, Dr. phil. (Corresponding author)
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany
Phone: +49 030 82406-720
E-mail: jenny@mpib-berlin.mpg.de

Niklas Keller, Dr. rer. nat.
Department 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

Gerd Gigerenzer, Dr. phil., Professor
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany

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 Leipzig University for 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-minute 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-minute training session for students increased the median percentage correct from 50% to 90%. 82% of participants improved their performance.

Conclusions: Medical students and educators struggle with basic concepts in medical statistics. This can be quickly assessed with the Quick Risk Test. The fact that a 90-minute 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 shortcoming 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 need enhanced training in how to interpret risk-related 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.
- A large student population was tested (N=169; ~60% of a cohort).
- Only a single site was included in each study.
- Only a small population of senior educators was tested.
- No parallel instruments for convergent validity were tested at the same time.

Keywords:
Medical Education & Training, Statistics & Research Methods, Risk Management

Introduction

For healthcare to be effective, medical professionals require literacy in health, the healthcare system, and medical statistics. 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 (e.g., 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 (e.g., 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 inadequate 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 (e.g., mortality vs. survival rates) or intransparent risk measures (e.g., 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 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 & Sunstein (2008)[10] argue that statistical errors are as stable as visual illusions and thereby justify governmental paternalism, popularly known as "nudging." Gigerenzer (2014)[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 programs in schools and medical curricula.

However, although frugal instruments exist to measure numeracy[4] 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 and provide a test that is applicable to all professional groups in health care: 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 ten concepts where chosen to cover a basic understanding of medical testing (understanding sensitivity, specificity, positive predictive value, negative predictive value, 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 solutions only requires a few simple mathematical computation 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-minute 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 (see Appendix) measures understanding of 10 central concepts: sensitivity, specificity, positive predictive value, negative predictive value, prevalence, Bayes rule, relative risk, mortality rate, lead time-bias, and overdiagnosis bias (see Appendix 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 gaps both in the medical school curriculum and in physicians' continuous education. Proficient teachers are obviously the first step toward enhancing medical statistical literacy; thus we wanted to avoid missing knowledge gaps in that group. Any such gaps indicate that not only medical school curricula but also continuing education programs 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 minutes of which deal with risk literacy and diagnostic risk assessment followed by 90-minutes 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 positive and negative predictive values and *PPV / NPV-curves*, to enhance understanding of the interplay between PPV / NPV, sensitivity, specificity, and prevalence. This training session consisted of a 15-minute theoretical introduction, a 45-minute small-group exercise in which students calculated the PPV / NPV of four commonly used diagnostic procedures (sigmoidoscopy / HIV combined test / neck-fold-test / amniocentesis) and a subsequent 30-minute discussion on the numerical and ethical implications of diagnostic risk assessment. The 90-minute 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 pre-test, 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 the Faculty of Medicine of Leipzig University 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. 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). Both groups gave informed consent before participation.

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

Results

Among the students, 62.5% were female with a median age of 25 years (interquartile range (IQR) = 24 – 26) and 61.5% (N = 104) completed both pre-and post-tests. Among the senior educators, 44% were female with an age range of <30 – 65. Amongst 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 pre-test 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). With respect to 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 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%, χ^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 to 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 (positive predictive value) and 5 (prevalence necessary to compute the positive predictive value) were the easiest (88% correct).

The students (N = 104), but not the senior educators, then completed a 90-minute 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%. 82% of participants improved their performance. After the 90-minute session (and an unrelated task of another 90 minutes), their performance improved to a median of 92.3% (IQR = 83.2% – 94.2%) correct answers per question (χ^2 = 300, df = 1, p <2e-16). Additionally, each question obtained more correct answers after training, even the question with the smallest pre–post difference in proportion correct answers, namely Question 7 on relative risk (χ^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 positive predictive value 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 programs. 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.[15] Figure 2 shows the strong effect of this part of the training, with students reaching an average of close to 90% correct, compared to only about 60% among the senior educators (Figure 1).

Discussion

The Quick Risk Test presented here measures minimal medical statistical literacy as defined by the 10 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 minutes. 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 (e.g., 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 generalize 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 minutes. However, the fact that students practiced 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 regards to natural frequency trees, studies showed that high application accuracy is maintained in a nonmedical population after up to 3-month follow-up[16]. 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 US 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 clearly inadequate amongst medical students[17] 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 (e.g., [1,5,6]). 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 not only in medical schools but also in physicians' continuous medical education programs. As we have shown, just 90 minutes of training on medical statistical literacy can make a big difference. We urge every medical school and every organizer of CME to include medical statistical literacy in their curricula so that physicians are fully competent in assessing medical risks. Because the test is geared toward 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.

Contributorship Statement
MJ and GG developed the Quick Risk Test. MJ 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.

Acknowledgments: We thank Clara Schirren for testing the senior educators and thank her along with Jana Hinneburg and Felix Rebitschek for their input concerning the wording of the items in the Quick Risk Test.

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

Competing Interests Statement
The authors declare that they have no competing interests as a result of associations with commercial entities or similar financial associations involving spouse or children.

Data Sharing Statement
No additional unpublished data are available from the study.

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

Figure Legends

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.

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-minute training session in risk literacy and diagnostic risk assessment.

Reference list

1 Gigerenzer G, Gray JAM. Launching the century of the patient. In: Gigerenzer G, Gray J, eds. *Better doctors, better patients, better decisions*. Better doctors 2011.

2 Berwick DM. Era 3 for medicine and health care. *JAMA* 2016;**315**:1329–30. doi:10.1001/jama.2016.1509

3 Mata J, Frank R, Gigerenzer G. Symptom recognition of heart attack and stroke in nine European countries: a representative survey. *Health Expectations* 2012;**17**:376–87. doi:10.1111/j.1369-7625.2011.00764.x

4 Cokely ET, Galesic M, Schulz E, *et al.* Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making* 2012;**7**:25–47.

5 Koh HK, Rudd RE. The Arc of Health Literacy. *JAMA* 2015;**314**:1225–6. doi:10.1001/jama.2015.9978

6 Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al.* Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest* 2007;**8**:53–96. doi:10.1111/j.1539-6053.2008.00033.x

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. *Annals of Internal Medicine* 2012;**156**:340–9. doi:10.7326/0003-4819-156-5-201203060-00005

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 Dermatology* 2017;**153**:348–50. doi:10.1001/jamadermatol.2016.4924

9 Anderson BL, Williams S, Schulkin J. Statistical Literacy of Obstetrics-Gynecology Residents. *Journal of Graduate Medical Education* 2013;**5**:272–5. doi:10.4300/JGME-D-12-00161.1

10 Thaler RH, Sunstein CR. *Nudge*. 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. doi:10.1186/1741-7015-5-14

13 Anderson BL, Gigerenzer G, Parker S, *et al.* Statistical Literacy in Obstetricians and Gynecologists. *Journal for Healthcare Quality* 2012;**36**:5–17.

14 McDowell M, Jacobs P. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin* 2017;**143**:1273–312. doi:10.1037/bul0000126

15 McDowell M, Jacobs P. Meta-Analysis of the Effect of Natural Frequencies on Bayesian Reasoning. *Psychological Bulletin* 2017;:1–41. doi:10.1037/bul0000126

16 Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General* 2001;**130**:380–400. doi:10.1037//0096-3445.130.3.380

17 Hoffrage U, Lindsey S, Hertwig R, *et al.* Communicating Statistical Information. *Science* 2000;**290**:2261–2. doi:10.1126/science.290.5500.2261

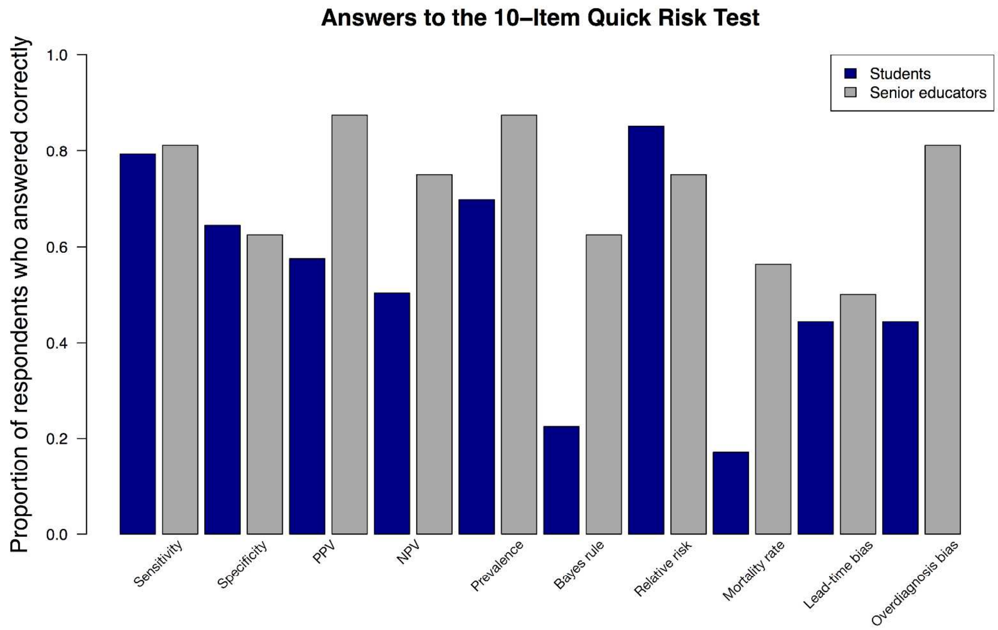

Figure 1. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students as well as professors, senior physicians, and university lecturers. The test measures minimal medical statistical literacy, as defined by understanding ten basic concepts. PPV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

361x270mm (300 x 300 DPI)

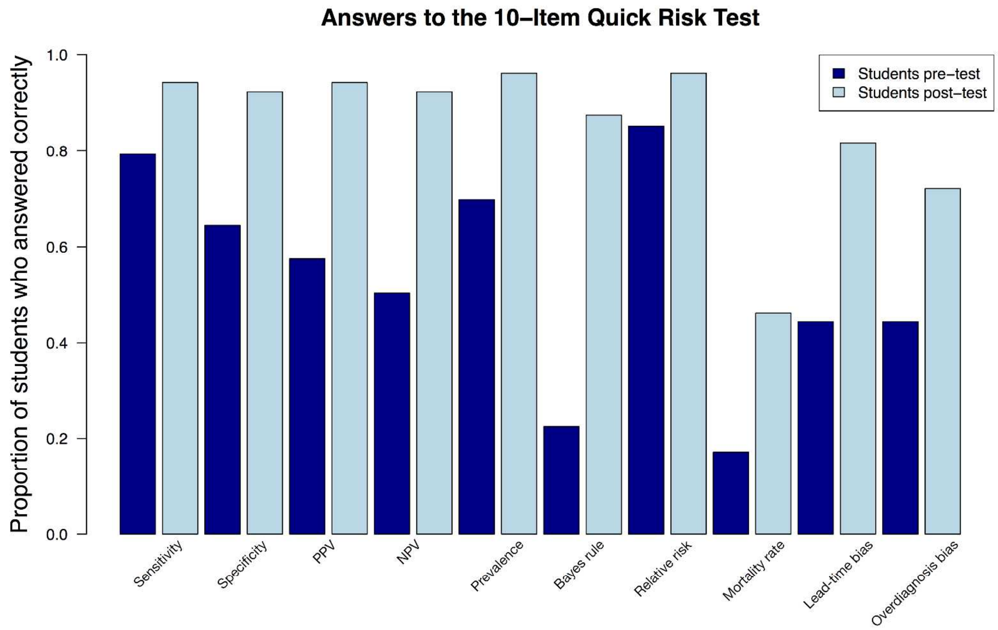

Figure 2. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students before and after a 90-minute training in risk literacy and diagnostic risk assessment.

361x270mm (300 x 300 DPI)

Appendix: 10-item Quick Test (* denotes correct answer)**

1. A test's sensitivity is a central criterion for its quality as a diagnostic tool.
The sensitivity describes
A) the proportion of sick people who receive a positive rest result. ***
B) the proportion of sick people who receive a negative rest result.
C) the proportion of healthy people who receive a positive test result.
D) the proportion of healthy people who receive a negative rest result.

2. A test's specificity is a central criterion for its quality as a diagnostic tool.
The specificity describes
A) the proportion of sick people who receive a positive rest result.
B) the proportion of sick people who receive a negative rest result.
C) the proportion of healthy people who receive a positive test result. ***
D) the proportion of healthy people who receive a negative rest result.

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) Prevalence ***
B) Mortality
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. Also in this group everyone dies at age 70. Thus, the 5-year survival rate is 0%. Which bias is used here to describe the benefits of screening?

A) Lead-time bias ***
B) Overdiagnosis bias
C) Selection 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, we call this

A) selection bias.
B) attrition bias.
C) lead-time bias.
D) overdiagnosis bias. ***

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **Title page and abstract**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **Is provided**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Page 3**
Objectives	3	State specific objectives, including any prespecified hypotheses **Page 3**
Methods		
Study design	4	Present key elements of study design early in the paper **Pages 4**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Pages 4**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **Pages 4**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **Not applicable**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Pages 3 and 4**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Pages 4**
Bias	9	Describe any efforts to address potential sources of bias **Convenience sampling, but large portion of student cohort (~60%), p.4**
Study size	10	Explain how the study size was arrived at **Convenience sample, page 4**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Pages 4 and 5**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **Page 4 and 5**
		(*b*) Describe any methods used to examine subgroups and interactions **Not applicable**
		(*c*) Explain how missing data were addressed **Page 4 and 5**
		(*d*) If applicable, explain how loss to follow-up was addressed **Not applicable**
		(*e*) Describe any sensitivity analyses **Not applicable**

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **Pages 4**
		(b) Give reasons for non-participation at each stage **Page 4**
		(c) Consider use of a flow diagram **Not applicable**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **Page 4**
		(b) Indicate number of participants with missing data for each variable of interest **Page 4**
		(c) Summarise follow-up time (eg, average and total amount) **Page 4**
Outcome data	15*	Report numbers of outcome events or summary measures over time **Page 4 and 5**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **Not applicable**
		(*b*) Report category boundaries when continuous variables were categorized **Not applicable**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **Not applicable**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Not applicable**
Discussion		
Key results	18	Summarise key results with reference to study objectives **Pages 4 and 5**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Page 5 and 6**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Page 5 and 6**
Generalisability	21	Discuss the generalisability (external validity) of the study results **Page 5 and 6**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Page 6**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

Assessing Minimal Medical Statistical Literacy Using the Quick Risk Test
A Prospective Observational Study in Germany

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020847.R3
Article Type:	Research
Date Submitted by the Author:	16-May-2018
Complete List of Authors:	Jenny, Mirjam; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy Keller, Niklas; Charité Universitätsmedizin Berlin, Department of Anesthesiology and Operative Intensive Care Medicine Gigerenzer, Gerd; Max-Planck-Institut fur Bildungsforschung, Harding Center for Risk Literacy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Communication, Medical education and training
Keywords:	MEDICAL EDUCATION & TRAINING, STATISTICS & RESEARCH METHODS, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Assessing Minimal Medical Statistical Literacy Using the Quick Risk Test
A Prospective Observational Study in Germany

Mirjam A. Jenny, Niklas Keller, Gerd Gigerenzer

Word Count (main body of text): 2979

Mirjam Annina Jenny, Dr. phil. (Corresponding author)
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany
Phone: +49 030 82406-720
E-mail: jenny@mpib-berlin.mpg.de

Niklas Keller, Dr. rer. nat.
Department 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

Gerd Gigerenzer, Dr. phil., Professor
Harding Center for Risk Literacy,
Max Planck Institute for Human Development,
Berlin, Germany

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 Leipzig University for 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-minute 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-minute 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-minute 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.

Strengths and Limitations of this Study:

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

Keywords:
Medical Education & Training, Statistics & Research Methods, Risk Management

Introduction

For healthcare to be effective, medical professionals require literacy in health, the healthcare system, and medical statistics. 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 (e.g., 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 (e.g., 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 (e.g., mortality vs. survival rates) or intransparent risk measures (e.g., 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 & Sunstein (2008)[10] argue that statistical errors are as stable as visual illusions and thereby justify governmental paternalism, popularly known as "nudging." Gigerenzer (2014)[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 programs in schools and medical curricula.

However, although frugal instruments exist to measure numeracy[4] 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 health care: 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 ten concepts where chosen to cover a basic understanding of medical testing (understanding sensitivity, specificity, positive predictive value, negative predictive value, 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 computation 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-minute 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 (see Appendix) measures understanding of 10 central concepts: sensitivity, specificity, positive predictive value, negative predictive value, prevalence, Bayes rule, relative risk, mortality rate, lead time-bias, and overdiagnosis bias (see Appendix 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 toward enhancing medical statistical literacy; thus we wanted to avoid missing knowledge gaps in that group. Any such gaps indicate that not only medical school curricula but also continuing education programs 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 minutes of which deal with risk literacy and diagnostic risk assessment followed by 90-minutes 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 positive and negative predictive values and *PPV / NPV-curves*, to enhance understanding of the interplay between PPV / NPV, sensitivity, specificity, and prevalence. This training session consisted of a 15-minute theoretical introduction, a 45-minute small-group exercise in which students calculated the PPV / NPV of four commonly used diagnostic procedures (sigmoidoscopy / HIV combined test / neck-fold-test / amniocentesis) and a subsequent 30-minute discussion on the numerical and ethical implications of diagnostic risk assessment. The 90-minute 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 pre-test, 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 the Faculty of Medicine of Leipzig University 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. 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). 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 interquartile ranges. The item discrimination index (point-biseral correlation) was calculated to test whether the items discriminated between students of different performance levels. Finally, inferential statistics were used in the form of chi-squared tests to test for group differences.

Results
Among the students, 62.5% were female with a median age of 25 years (interquartile range (IQR) = 24 – 26) and 61.5% (N = 104) completed both pre-and post-tests. Among the senior educators, 44% were female with an age range of <30 – 65. Amongst 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 pre-test 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). With respect to 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 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 to 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 (positive predictive value) and 5 (prevalence necessary to compute the positive predictive value) were the easiest (88% correct).

The students (N = 104), but not the senior educators, then completed a 90-minute 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%. 82% of participants improved their performance. After the 90-minute session (and an unrelated task of another 90 minutes), 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 pre–post 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 positive predictive value 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 programs. 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.[15] Figure 2 shows the strong effect of this part of the training, with students reaching an average of close to 90% correct, compared to 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 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 minutes. 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 (e.g., 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 generalize 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 minutes. However, the fact that students practiced 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 regards to natural frequency trees, studies showed that high application accuracy is maintained in a nonmedical population after up to 3-month follow-up[16]. 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 US 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 amongst medical students[17] 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 (e.g., [1,5,6]). 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 not only in medical schools but also in physicians' continuous medical education programs. As we have shown, just 90 minutes of training on medical statistical literacy can make a big difference. We urge medical schools and every organizers 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 toward 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.

Contributorship Statement
MJ and GG developed the Quick Risk Test. MJ 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.

Acknowledgments: We thank Clara Schirren for testing the senior educators and thank her along with Jana Hinneburg, Felix Rebitschek, and Anna Held for their input concerning the wording of the items in the Quick Risk Test.

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

Competing Interests Statement
The authors declare that they have no competing interests as a result of associations with commercial entities or similar financial associations involving spouse or children.

Data Sharing Statement
No additional unpublished data are available from the study.

Figure Legends

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.

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-minute training session in risk literacy and diagnostic risk assessment.

Reference list

1 Gigerenzer G, Gray JAM. Launching the century of the patient. In: Gigerenzer G, Gray J, eds. *Better doctors, better patients, better decisions*. Better doctors 2011.

2 Berwick DM. Era 3 for medicine and health care. *JAMA* 2016;**315**:1329–30. doi:10.1001/jama.2016.1509

3 Mata J, Frank R, Gigerenzer G. Symptom recognition of heart attack and stroke in nine European countries: a representative survey. *Health Expectations* 2012;**17**:376–87. doi:10.1111/j.1369-7625.2011.00764.x

4 Cokely ET, Galesic M, Schulz E, *et al.* Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making* 2012;**7**:25–47.

5 Koh HK, Rudd RE. The Arc of Health Literacy. *JAMA* 2015;**314**:1225–6. doi:10.1001/jama.2015.9978

6 Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al.* Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest* 2007;**8**:53–96. doi:10.1111/j.1539-6053.2008.00033.x

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. *Annals of Internal Medicine* 2012;**156**:340–9. doi:10.7326/0003-4819-156-5-201203060-00005

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 Dermatology* 2017;**153**:348–50. doi:10.1001/jamadermatol.2016.4924

9 Anderson BL, Williams S, Schulkin J. Statistical Literacy of Obstetrics-Gynecology Residents. *Journal of Graduate Medical Education* 2013;**5**:272–5. doi:10.4300/JGME-D-12-00161.1

10 Thaler RH, Sunstein CR. *Nudge*. 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. doi:10.1186/1741-7015-5-14

13 Anderson BL, Gigerenzer G, Parker S, *et al.* Statistical Literacy in Obstetricians and Gynecologists. *Journal for Healthcare Quality* 2012;**36**:5–17.

14 McDowell M, Jacobs P. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin* 2017;**143**:1273–312. doi:10.1037/bul0000126

15 McDowell M, Jacobs P. Meta-Analysis of the Effect of Natural Frequencies on Bayesian Reasoning. *Psychological Bulletin* 2017;:1–41. doi:10.1037/bul0000126

16 Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General* 2001;**130**:380–400. doi:10.1037//0096-3445.130.3.380

17 Hoffrage U, Lindsey S, Hertwig R, *et al.* Communicating Statistical Information. *Science* 2000;**290**:2261–2. doi:10.1126/science.290.5500.2261

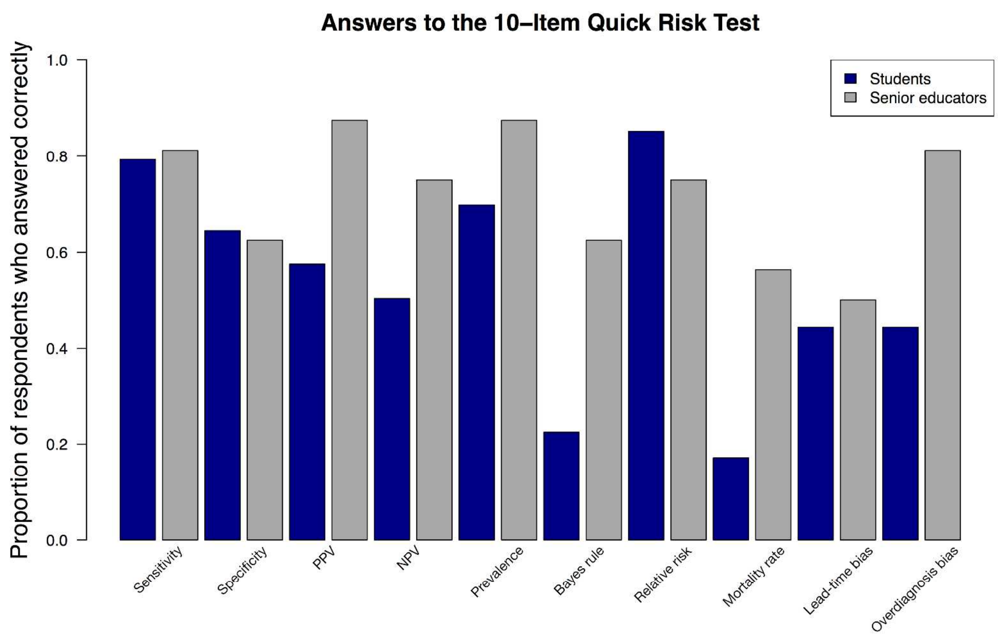

Figure 1. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students as well as professors, senior physicians, and university lecturers. The test measures minimal medical statistical literacy, as defined by understanding ten basic concepts. PPV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

361x270mm (300 x 300 DPI)

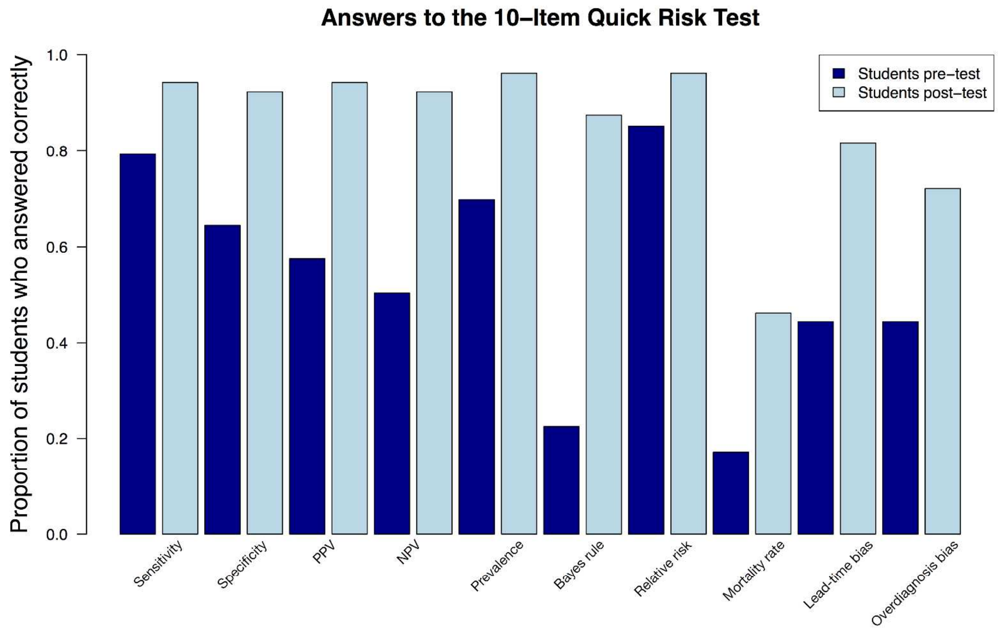

Figure 2. The proportion of correct answers for each of the 10 questions of the Quick Risk Test, for last-year medical students before and after a 90-minute training in risk literacy and diagnostic risk assessment.

361x270mm (300 x 300 DPI)

Appendix: 10-item Quick Test (* denotes correct answer)**

1. A test's sensitivity is a central criterion for its quality as a diagnostic tool.
The sensitivity describes
A) the proportion of sick people who receive a positive rest result. ***
B) the proportion of sick people who receive a negative rest result.
C) the proportion of healthy people who receive a positive test result.
D) the proportion of healthy people who receive a negative rest result.

2. A test's specificity is a central criterion for its quality as a diagnostic tool.
The specificity describes
A) the proportion of sick people who receive a positive rest result.
B) the proportion of sick people who receive a negative rest result.
C) the proportion of healthy people who receive a positive test result. ***
D) the proportion of healthy people who receive a negative rest result.

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) Prevalence ***
B) Mortality
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. Also in this group everyone dies at age 70. Thus, the 5-year survival rate is 0%. Which bias is used here to describe the benefits of screening?

A) Lead-time bias ***
B) Overdiagnosis bias
C) Selection 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, we call this

A) selection bias.
B) attrition bias.
C) lead-time bias.
D) overdiagnosis bias. ***

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **Title page and abstract**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **Is provided**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Page 3**
Objectives	3	State specific objectives, including any prespecified hypotheses **Page 3**
Methods		
Study design	4	Present key elements of study design early in the paper **Pages 4**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Pages 4**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **Pages 4**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **Not applicable**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Pages 3 and 4**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Pages 4**
Bias	9	Describe any efforts to address potential sources of bias **Convenience sampling, but large portion of student cohort (~60%), p.4**
Study size	10	Explain how the study size was arrived at **Convenience sample, page 4**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Pages 4 and 5**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **Page 4 and 5**
		(*b*) Describe any methods used to examine subgroups and interactions **Not applicable**
		(*c*) Explain how missing data were addressed **Page 4 and 5**
		(*d*) If applicable, explain how loss to follow-up was addressed **Not applicable**
		(*e*) Describe any sensitivity analyses **Not applicable**

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **Pages 4**
		(b) Give reasons for non-participation at each stage **Page 4**
		(c) Consider use of a flow diagram **Not applicable**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **Page 4**
		(b) Indicate number of participants with missing data for each variable of interest **Page 4**
		(c) Summarise follow-up time (eg, average and total amount) **Page 4**
Outcome data	15*	Report numbers of outcome events or summary measures over time **Page 4 and 5**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **Not applicable**
		(*b*) Report category boundaries when continuous variables were categorized **Not applicable**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **Not applicable**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Not applicable**
Discussion		
Key results	18	Summarise key results with reference to study objectives **Pages 4 and 5**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Page 5 and 6**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Page 5 and 6**
Generalisability	21	Discuss the generalisability (external validity) of the study results **Page 5 and 6**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Page 6**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.