

Avaliação de letramento de risco em estudantes de medicina

Risk literacy assessment of medical students in Brazil

Evaluación de la alfabetización del riesgo en estudiantes de medicina en Brasil

Lucas Magalhães Moreira¹ , Fabio Yuji Furukawa¹ , Itamar Santos¹ , Gustavo Diniz Ferreira Gusso¹ 

¹Faculdade de Medicina da Universidade de São Paulo – São Paulo (SP), Brasil.

Resumo

Introdução: Para o processo de decisão compartilhada é essencial que profissionais da saúde interpretem dados estatísticos das melhores evidências disponíveis para que essas informações possam ser comunicadas para seus pacientes. Nesse contexto, o letramento de risco é a capacidade de avaliar riscos e benefícios de determinada ação. Apesar da importância dessa habilidade, estudos têm mostrado que muitos profissionais e estudantes possuem dificuldade na compreensão de conceitos estatísticos e de probabilidade e, dessa forma, baixo letramento de risco.

Objetivo: Este estudo teve como objetivo avaliar o letramento de risco em estudantes de medicina e como isso impacta a capacidade de resolver um problema de cálculo de valor preditivo positivo de um exame de rastreamento. **Métodos:** Foram convidados estudantes do 4º, 5º e 6º anos da Faculdade de Medicina da Universidade de São Paulo para responder a um questionário composto pelo *Berlin Numeracy Test* (BNT), instrumento validado para a mensuração de numeracia, e um problema clínico sobre cálculo de valor preditivo positivo (VPP) em rastreamento de câncer de mama com mamografia. Avaliar qual o grau de letramento de risco em estudantes de medicina e verificar se existe associação entre o número de acertos no BNT e a capacidade de resolução do cenário clínico sobre VPP. **Resultados:** Obtivemos 97 respostas, em que 19 (19,52%) participantes acertaram 3 das 4 questões do BNT, e 61 (62,89%) acertaram todas as questões. Já na pergunta sobre VPP do rastreamento de câncer de mama houve 43 respostas corretas (44,33%). A média de pontuação no BNT da amostra de participantes foi de 3,41. Entre os estudantes que acertaram o cálculo do VPP, a média foi 3,67 e, entre os que erraram, foi de 3,21. **Conclusões:** Apesar da numeracia alta medida pelo BNT, os estudantes apresentam baixa taxa de acerto no caso clínico. Este estudo reforça os achados prévios de que o letramento de risco é uma habilidade difícil de ser aprendida, mesmo em pessoas com alta numeracia. Entretanto, o baixo número de respostas dificulta a interpretação mais precisa dos resultados.

Palavras-chave: Educação médica; Valor preditivo dos testes; Programas de rastreamento; Risco; Medidas de associação, exposição, risco ou desfecho.

Autor correspondente:

Lucas Magalhães Moreira
E-mail: lucas.magalhaes3151@gmail.com

Fonte de financiamento:

não se aplica.

Parecer CEP:

CAAE: 59183722.0.0000.0068, Parecer nº:
5.753.756 CEP HC-FMUSP.

Procedência:

não encomendado.

Avaliação por pares:

externa.

Recebido em: 09/06/2023.

Aprovado em: 07/09/2023.

Editores convidados:

Maria Inez Padula Anderson e
Marcello Dala Bernardina Dalla.

Como citar: Moreira LM, Furukawa FY, Santos I, Gusso GDF. Avaliação de letramento de risco em estudantes de medicina. Rev Bras Med Fam Comunidade. 2023;18(45):3802. [https://doi.org/10.5712/rbmfc18\(45\)3802](https://doi.org/10.5712/rbmfc18(45)3802)



Abstract

Background: In order to accomplish the shared decision-making process, it is essential that health professionals are able to understand statistical data of the best available evidence, so that this information can be communicated to their patients. In this context, risk literacy is the ability to evaluate risks and benefits of a given action. Despite the relevance of this skill, research has demonstrated that many professionals and students have difficulty interpreting concepts of statistics and probability, therefore having low risk literacy. **Objectives:** This study aimed to evaluate risk literacy in medical students and how it impacts their ability to solve a problem concerning the positive predictive value (PPV) of a mass screening. **Methods:** Medical students from the 4th, 5th, and 6th years of the School of Medicine of Universidade de São Paulo were invited to answer a questionnaire comprised of the Berlin Numeracy Test (BNT), a validated instrument to measure numeracy, and a clinical problem regarding the PPV of a mammogram in the context of mass screening. To measure the level of risk literacy of medical students and to investigate whether there is an association between score in the BNT and the ability to correctly answer the clinical problem regarding PPV. **Results:** A total of 97 responses were collected, of which 19 (19.52%) participants answered 3 out of 4 questions of the BNT, and 61 (62.89%) correctly answered all the questions. In the clinical problem about PPV of cancer screening, there were 43 correct answers (44.33%). The mean BNT score of the participants was 3.41. Among the students who correctly answered the problem, the mean score was 3.76, and among the ones who answered incorrectly, it was 3.21. **Conclusions:** Despite the high numeracy measured by the BNT, students had a poor outcome in the clinical problem. This study reinforces previous findings that risk literacy is a difficult skill to be learned, even in individuals with high numeracy. However, the low response rate hinders a more precise interpretation of the results.

Keywords: Education, medical; Predictive value of tests; Mass screening; Risk; Measures of association, exposure, risk or outcome.

Resumen

Introducción: Para el proceso de toma de decisión compartida, es esencial que los profesionales de la salud sean capaces de interpretar datos estadísticos a partir de las mejores evidencias científicas disponibles para que las informaciones puedan ser comunicadas a sus pacientes. En ese contexto, la alfabetización de riesgo es la capacidad de evaluar riesgos y beneficios de una determinada acción. No obstante la importancia de esa habilidad, estudios han demostrado que muchos profesionales y estudiantes tienen dificultades para comprender conceptos de estadística y de probabilidad y, por consecuencia, baja alfabetización de riesgo. **Objetivos:** Este estudio tiene como objetivo evaluar la alfabetización de riesgo en estudiantes de medicina y cómo eso impacta la capacidad de resolver un problema de cálculo del valor predictivo positivo (VPP) de un examen de tamizaje. **Métodos:** Estudiantes de 4^º, 5^º y 6^º año de la Facultad de Medicina de la Universidad de São Paulo fueron invitados a responder un cuestionario compuesto por el *Berlin Numeracy Test* (BNT), un instrumento validado para medir la numeracia, y un problema clínico sobre el cálculo del VPP en el tamizaje de cáncer de mama con mamografía. Evaluar el grado de alfabetización de riesgo en estudiantes de medicina y verificar si existe asociación entre el número de aciertos en BNT y la capacidad de resolución del escenario clínico de VPP. **Resultados:** Se obtuvieron 97 respuestas, de las que 19 (19,52%) participantes respondieron correctamente 3 de las 4 preguntas del BNT y 61 (62,89%) respondieron correctamente todas las preguntas. En la pregunta sobre el VPP del tamizaje de cáncer de mama hubo 43 respuestas correctas (44,33%). La puntuación media de BNT de los participantes fue de 3,41. Entre los estudiantes que calcularon correctamente el VPP, el promedio fue de 3,67 y entre los que se equivocaron, fue de 3,21. **Conclusiones:** A pesar de la alta numeracia medida por el BNT, los estudiantes tienen una baja tasa de éxito en el caso clínico. Este estudio reforzó los hallazgos previos de que la alfabetización de riesgo es una habilidad difícil de aprender, incluso en personas con un alto nivel de numeracia. Sin embargo, el bajo número de respuestas dificulta una interpretación más precisa de los resultados.

Palabras clave: Educación médica; Valor predictivo de las pruebas; Tamizaje masivo; Riesgo; Medidas de asociación, exposición, riesgo o desenlace.

INTRODUÇÃO

O contexto da prática médica atual é cada vez mais o de uma decisão compartilhada com o paciente.^{1,2} Nesse modelo, não cabe ao profissional de saúde decidir pelo paciente, mas sim traduzir a melhor evidência científica disponível sobre o plano terapêutico (diagnóstico, tratamento e prognóstico) em termos compreensíveis aos indivíduos.

No entanto, pesquisas vêm mostrando que os próprios profissionais da saúde (especialmente médicos), muitas vezes, são incapazes de interpretar corretamente os dados estatísticos, prejudicando sua habilidade para a comunicação de risco.³⁻⁵ A capacidade de compreender os dados de estatística e probabilidade é uma área particular de *numeracy* (por vezes traduzida como numeracia) que alguns autores têm chamado de *statistical literacy*.^{6,7} Optamos por traduzir como letramento estatístico.⁸ Letramento de risco (*risk literacy*), por sua vez, é um nível mais elevado de numeracia que é fortemente influenciado pelo

letramento estatístico e diz respeito à habilidade prática de entender e avaliar os riscos e benefícios de determinada ação no contexto de uma decisão informada.^{5,8}

Entre estudantes de medicina, algumas pesquisas mostraram que, apesar do maior foco que a medicina baseada em evidências vem recebendo nos currículos, os estudantes não têm desempenho satisfatório em tarefas relacionadas à interpretação de dados estatísticos.⁸⁻¹²

O letramento de risco é fundamental para a comunicação clínica de probabilidades.^{13,14} Pesquisas demonstraram que, apesar de uma numeracia considerada acima da média da população geral⁵, médicos têm desempenho ruim na aplicação desse conhecimento a seus cenários de prática. Gigerenzer et al. sugeriram que a probabilidade de ginecologistas calcularem corretamente o valor preditivo positivo (VPP) de uma mamografia foi pior do que uma resposta ao acaso.¹⁵ Bramwell, West e Salmon reportaram taxa de acertos entre obstetras de 34% para o VPP de trissomia do 21 dado um teste de rastreamento positivo.¹⁶ Médicos de atenção primária dos Estados Unidos optaram por evidência irrelevante (aumento da detecção precoce e da sobrevivência em cinco anos) em relação a relevante (aumento da incidência com queda na mortalidade por câncer).¹⁷ Os poucos estudos neste campo feitos com estudantes e residentes vêm mostrando que não há grande diferença entre graduandos, residentes e formados.⁸⁻¹²

Objetivos

As perguntas principais a serem respondidas pela pesquisa eram:

- Qual o grau de numeracia em estudantes de medicina?
- Existe relação entre numeracia e a capacidade de interpretação de um cenário envolvendo o VPP de um teste de rastreamento?

MÉTODOS

Materiais

O instrumento utilizado foi uma versão adaptada pelos autores da tradução para o português europeu do *Berlin Numeracy Test* (BNT), cedida gentilmente pelos criadores (disponível como material suplementar em <https://zenodo.org/record/8331044>). Optou-se pelo uso do BNT por ser um questionário já validado para a avaliação de numeracia estatística e letramento de risco em diversos países (por exemplo, Alemanha, Japão, EUA, Paquistão) e testado em diferentes amostras (por exemplo, médicos, população geral).¹⁸ Adicionalmente, foi acrescida uma pergunta sobre o VPP do rastreamento de câncer de mama, a fim de comparar o desempenho neste cenário com o resultado do BNT (também no material suplementar). A pergunta sobre o rastreamento do câncer de mama foi escolhida por ser paradigmática nos estudos sobre letramento de risco e raciocínio bayesiano em profissionais de saúde, o que também possibilita a comparação com a literatura já existente.^{5,9,19,20}

Participantes

Trata-se de um estudo transversal com base na aplicação de questionário em alunos do 4º, 5º e 6º ano do curso de medicina da Faculdade de Medicina da Universidade de São Paulo (FMUSP).

Os estudantes foram convidados a responder ao questionário voluntariamente após leitura e concordância com o Termo de Consentimento Livre e Esclarecido. Optou-se por essa população por se tratar de período de importante inserção na prática clínica durante a graduação, sendo utilizados rotineiramente conceitos de epidemiologia e risco. O estudo foi aprovado no Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de São Paulo, sob o Parecer nº 5.753.756.

O questionário foi inserido na plataforma Google Forms, sendo gerado um *link* para resposta virtual pelos voluntários. O convite para a participação na pesquisa foi realizado em algumas ocasiões presencialmente, como em aulas e provas práticas, porém o principal método de recrutamento de participantes foi por meio de mensagens via *e-mail*, aplicativo de mensagens *WhatsApp* e pela rede social Instagram.

Os convites para participação, nesta primeira fase, foram realizados durante o período de 29 de julho de 2022 a 5 de dezembro de 2022.

Análise estatística

O resultado principal analisado foi a correlação entre alto desempenho no BNT e resposta correta da pergunta clínica. Considerando-se o nível de significância alfa de 5% e poder do estudo de 80%, calculou-se, com o uso do *software* R, que com uma amostra de 270 participantes (80% dos alunos de duas turmas) seria possível identificar uma diferença de 15% entre os grupos. Para esta análise, utilizamos como teste estatístico o qui-quadrado.

RESULTADOS

Durante o período em que o questionário esteve aberto, obtivemos 97 respostas válidas. Destas, 31 (32%) foram de alunos do 4º ano, 42 (43,3%) do 5º ano, e 24 (24,7%) do 6º ano. Com relação ao gênero dos participantes, 42 (43,2%) eram do gênero feminino, 54 (55,7%) eram do gênero masculino, e 1 (1,1%) era não binário.

A distribuição de acertos no BNT está disposta na Tabela 1.

Já na pergunta sobre VPP do rastreamento de câncer de mama houve 43 respostas corretas (44,33%).

A média de pontuação no BNT da amostra de participantes foi de 3,41 (Gráfico 1). Entre os estudantes que acertaram o cálculo do VPP, a média foi de 3,67 e, entre os que erraram, foi de 3,21. Não houve significância estatística entre ter BNT \geq 3 e acertar o cálculo do VPP ($p=0,085$) nem entre ter BNT=4 e acertar o cálculo do VPP ($p=0,106$).

DISCUSSÃO

A média de pontuação no BNT obtida neste estudo foi substancialmente maior do que a de outros. Na publicação seminal do BNT, Cokely et al. indicaram, para o formato de papel e lápis, média populacional de 1,6; mediana de 2; e média do 4º quartil de 3,3.¹⁸ Friederichs et al. encontraram média de 2,34 em alunos do 3º ano na Westphalian Wilhelms University e de 2,33 em médicos de família e comunidade da Alemanha.⁵ Em outro trabalho do mesmo grupo, encontraram média de 2,03 entre 600 estudantes do 1º ao 6º ano da Muenster University, com menor média de 1,76 entre alunos do 3º ano e maior de 2,38 entre alunos do 6º ano.¹² Em um grupo composto de 119 residentes de 14 especialidades diferentes e 54 estudantes do 6º e 7º ano da Cayetano Heredia University, em Lima, Petrova et al. verificaram média de 1,01.⁶

Com relação ao cálculo do VPP de mamografias, Jenny, Keller e Gigerenzer encontraram taxa de acerto de 22,5% (abaixo do 25% de chance de acerto ao acaso para o formato de múltipla escolha utilizado no estudo) em estudantes do último ano da Charité University, em Berlim, e de aproximadamente 60% em profissionais envolvidos com ensino.⁹ No estudo de Friederichs et al., apenas 23,5% dos médicos de família e comunidade realizaram corretamente o cálculo.⁵ Entre ginecologistas-obstetras americanos, Anderson et al. observaram 49% de acerto.²⁰

Ao nosso conhecimento, este é o primeiro estudo a avaliar simultaneamente o BNT e o cálculo do VPP de uma mamografia em estudantes. Na pesquisa com médicos de família e comunidade,⁵ a média de BNT dos participantes que acertaram o cálculo foi de 3,00, enquanto a dos que erraram foi de 2,11. Schwartz et al. usaram outro instrumento para medir a numeracia de mulheres não médicas e sua compreensão do benefício calculável do rastreamento de câncer de mama.²¹ Mesmo as mulheres que acertaram todos os itens do instrumento só obtiveram a resposta correta sobre o benefício da mamografia 40% das vezes. No estudo de Bramwell com gestantes, acompanhantes, parteiras e obstetras, 34% das obstetras foram capazes de computar corretamente o VPP de um teste de rastreamento, enquanto apenas 15% dos acompanhantes, 9% das gestantes e 0% das parteiras foram capazes de fazê-lo.¹⁶ Usando-se escolaridade como uma *proxy* para numeracia, observa-se correlação positiva entre numeracia e capacidade de cálculo do VPP, mas ainda assim com a maioria das respostas incorreta.

Apesar de a amostra de participantes deste estudo apresentar alta numeracia medida pelo BNT (média de 3,41), chama atenção a baixa taxa de acerto no cálculo do VPP (44,33%). Chapman e Liu reconhecem que problemas bayesianos:²²

Tabela 1. Número de acertos no *Berlin Numeracy Test*.

Pontuação no BNT	Número de participantes com a pontuação
0	0 (0%)
1	4 (4,12%)
2	13 (13,40%)
3	19 (19,59%)
4	61 (62,89%)

Fonte: elaboração própria.

BNT: *Berlin Numeracy Test*.

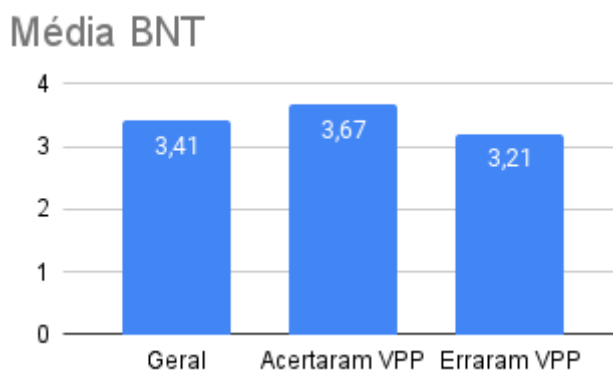


Gráfico 1. Médias de acerto no *Berlin Numeracy Test*.

são obviamente bem difíceis para participantes, e níveis modestos de acurácia são encontrados mesmo com a condição que o problema apareça em um formato de frequência e que o participante tenha um alto grau de numeracia (tradução dos autores).

Existem na literatura da psicologia cognitiva várias hipóteses acerca dos processos mentais e fatores contextuais que podem estar implicados na capacidade de realizar atualizações bayesianas. Com relação ao formato em que a pergunta é apresentada (probabilidades nesta pesquisa), há evidências de que o formato de frequências naturais (“9 em cada 100 mulheres” para 9%) facilita a resolução, porém com algumas limitações.²²⁻²⁴ Por exemplo, em alguns estudos o benefício só foi observado ou é bastante maior em grupos com alta numeracia, enquanto em outros o inverso foi verdadeiro. No caso do estudo de Bramwell com obstetras e gestantes, a apresentação no formato de frequência natural aumentou a taxa de acerto exclusivamente no grupo de obstetras. Em pesquisa com universitários usando um problema bayesiano, Brase encontrou taxa de acerto de 35,4%, mesmo apresentando o problema em formato de frequência natural.²⁵ Ottley et al. mostraram que a adição da pergunta “Quantas pessoas testaram positivo?” antes da pergunta “Das que testaram positivo, quantas pessoas têm a doença?” aumentou o acerto de 5,4 para 26,7%, e inserir o problema no contexto de uma narrativa breve aumentou para 42,4%.²⁶

Outros autores apontam que a familiaridade com o problema em questão (isto é, o quanto a resposta correta se aproxima de uma atualização bayesiana *naive* ou pela experiência) e as crenças tidas pelo indivíduo afetam seu desempenho. A familiaridade diz respeito não apenas ao contexto das probabilidades dadas no problema, mas também à taxa base (*base rate*) em questão. Binder encontrou taxa de acerto significativamente maior a um problema cuja taxa base era de 32,5% em comparação a um em que era de 1%.²⁷ As crenças dizem respeito a vieses implícitos que podem estar envolvidos no raciocínio. Se um estudante de medicina tem uma crença forte de que uma mamografia é capaz de *confirmar ou descartar* um diagnóstico de câncer de mama, ele terá maior tendência a superestimar o VPP.²⁸

É possível também que, durante a resolução do problema, o participante se atenha a estratégias aprendidas previamente, mesmo que não sejam as mais simples ou adequadas para a resolução do problema em questão. Este fenômeno é chamado de efeito Einstellung e mesmo *experts* estão sujeitos a ele.^{24,29} É bastante provável que a maioria (senão todos) os participantes desta pesquisa já tenha sido apresentada a uma forma de cálculo do VPP de um teste diagnóstico, e ainda assim menos da metade foi capaz de chegar à resposta correta. Entretanto, 62,89% deles acertaram todas as questões do BNT e poderiam, por um raciocínio análogo, calcular o VPP. É plausível que parte da falha em chegar à resposta se deva a tentar reproduzir um padrão a que se foi exposto previamente (por exemplo, tentar memorizar quadrantes fixos para a construção de uma tabela 2x2 do tipo “teste positivo/teste negativo” vs. “doente/saudável”) e fracassar na sua aplicação.

LIMITAÇÕES

Nosso estudo possui algumas limitações. Primeiramente, o convite para a participação na pesquisa foi realizado majoritariamente de maneira virtual, o que pode ter contribuído para a baixa taxa de respostas que obtivemos em relação ao número total de estudantes matriculados no 4º, 5º e 6º anos da faculdade. É possível que o pequeno número de respostas possa ter selecionado participantes com maior afinidade e interesse pelo assunto, o que poderia superestimar a numeracia e o grau de letramento de risco na amostra.

O número de respostas inferior ao tamanho de amostra calculado também pode ter sido a causa de não ter sido encontrada significância estatística na associação entre o BNT e o cálculo do VPP.

Outro ponto a ser citado é que, até o momento, não existe uma versão traduzida validada do BNT para o português do Brasil. Por conta disso, fizemos uma adaptação própria das perguntas do BNT. Entretanto, por se tratar de texto objetivo, com perguntas com enfoque no raciocínio numérico, acreditamos que esse processo não tenha dificultado a compreensão do questionário.

CONCLUSÃO

Não foi possível obter uma estimativa bastante confiável do grau de numeracia e letramento de risco dos estudantes de medicina da FMUSP em razão de um provável viés de seleção.

Apesar da numeracia alta medida pelo BNT, os estudantes apresentam baixo letramento de risco (o qual, no entanto, está condizente com a literatura científica). Este estudo reforça os achados prévios de que o letramento de risco (avaliado pela capacidade de resolver problemas bayesianos) é uma habilidade difícil de ser aprendida, mesmo em pessoas com alta numeracia.

AGRADECIMENTOS

Agradecemos aos alunos do curso de medicina da FMUSP, que tornaram esta pesquisa possível.

CONFLITO DE INTERESSE

Nada a declarar.

CONTRIBUIÇÕES DOS AUTORES

LMM e FYF: conceituação, investigação, metodologia, administração do projeto, escrita – primeira redação, escrita - revisão e edição. IS: conceituação, metodologia, administração do projeto, software, supervisão, escrita – primeira redação. GDFG: conceituação, metodologia, administração do projeto, supervisão, escrita – primeira redação.

REFERÊNCIAS

1. Stewart M, Brown JB, Weston WW, McWhinney IR, McWilliam CL, Freeman TR. O terceiro componente: elaborando um plano conjunto de manejo dos problemas. In: Medicina centrada na pessoa: transformando o método clínico. 3. ed. ARTMED; 2017. p. 146-86.
2. Montori VM, Elwyn G, Devereaux P, Strauss SE, Haynes RB, Guyatt G. Decision Making and the Patient. In: Users' guides to the medical literature. 3. ed. New York: McGraw-Hill Medical; 2015. p. 389-415.
3. Wegwarth O, Gigerenzer G. The barrier to informed choice in cancer screening: statistical illiteracy in physicians and patients. *Recent Results Cancer Res* 2018;210:207-221. https://doi.org/10.1007/978-3-319-64310-6_13
4. Soto-Mota A, Carrillo Maravilla E, Fragoso JM, Castejón A, González Herrero A, Ponce S. Evaluation of statistical illiteracy in Latin American clinicians and the piloting evaluation of a short course across multiple timepoints. *BMC Med Educ* 2022;22(1):54. <https://doi.org/10.1186/s12909-022-03128-w>
5. Friederichs H, Birkenstein R, Becker JC, Marschall B, Weissenstein A. Risk literacy assessment of general practitioners and medical students using the Berlin Numeracy Test. *BMC Fam Pract* 2020;21(1):143. <https://doi.org/10.1186/s12875-020-01214-w>
6. Rothman RL, Montori VM, Cherrington A, Pignone MP. Perspective: The Role of Numeracy in Health Care. *J Health Commun* 2008;13(6):583-95. <https://doi.org/10.1080/10810730802281791>

7. Schmidt FM, Zottmann JM, Sailer M, Fischer MR, Berndt M. Statistical literacy and scientific reasoning & argumentation in physicians. *MS J Med Educ* 2021;38(4):Doc77. <https://doi.org/10.3205/zma001473>
8. Petrova D, Mas G, Navarrete G, Rodriguez TT, Ortiz PJ, Garcia-Retamero R. Cancer screening risk literacy of physicians in training: An experimental study. *PLoS One* 2018;14(7):e0218821. <https://doi.org/10.1371/journal.pone.0218821>
9. Jenny MA, Keller N, Gigerenzer G. Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany. *BMJ Open* 2018;8(8):e020847. <http://doi.org/10.1136/bmjopen-2017-020847>
10. Johnson TV, Abbasi A, Schoenberg ED, Kellum R, Speake LD, Spiker C, et al. Numeracy among trainees: are we preparing physicians for evidence-based medicine? *J Surg Educ* 2014;71(2):211-5. <http://doi.org/10.1016/j.jsurg.2013.07.013>
11. Sheridan SL, Pignone M. Numeracy and the medical student's ability to interpret data. *Eff Clin Pract* 2002;5(1):35-40. PMID: 11874195
12. Friederichs H, Schölling M, Marschall B, Weissenstein A. Assessment of risk literacy among german medical students: a cross-sectional study evaluating numeracy skills. *HERA* 2014;20(4):1139-47. <https://doi.org/10.1080/10807039.2013.821909>
13. Ghosh AK, Ghosh K. Translating evidence-based information into effective risk communication: Current challenges and opportunities. *J Lab Clin Med* 2005;145(4):171-80. <https://doi.org/10.1016/j.lab.2005.02.006>
14. Petrova D, Kostopoulou O, Delaney BC, Cokely ET, Garcia-Retamero R. Strengths and gaps in physicians' risk communication: a scenario study of the influence of numeracy on cancer screening communication. *Med Decis Making* 2018;38(3):355-365. <https://doi.org/10.1177/0272989X17729359>
15. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8(2):53-96. <https://doi.org/10.1111/j.1539-6053.2008.00033.x>
16. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *BMJ* 2006;333(7562):284. <https://doi.org/10.1136/bmj.38884.663102.AE>
17. Wegwarth O. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the united states. *Ann Intern Med* 2012;156(5):340-9. <https://doi.org/10.7326/0003-4819-156-5-201203060-00005>
18. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: the berlin numeracy test. *Judgm Decis Mak* 2012;7(1):25-47. <https://doi.org/10.1017/S1930297500001819>
19. Eddy D. Probabilistic reasoning in clinical medicine: Problems and opportunities. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press; 1982. p. 249-67. <https://doi.org/10.1017/CBO9780511809477.019>
20. Anderson BL, Gigerenzer G, Parker S, Schulkin J. Statistical literacy in obstetricians and gynecologists. *J Healthc Qual* 2014;36(1):5-17. <https://doi.org/10.1111/j.1945-1474.2011.00194.x>
21. Schwartz LM. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med* 1997;127(11):966-72. <https://doi.org/10.7326/0003-4819-127-11-199712010-00003>
22. Chapman GB, Liu J. Numeracy, frequency, and Bayesian reasoning. *Judgm Decis Mak* 2009;4(1):34-40. <https://doi.org/10.1017/S1930297500000681>
23. Garcia-Retamero R, Hoffrage U. Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Soc Sci Med* 2013;83:27-33. <https://doi.org/10.1016/j.socscimed.2013.01.034>
24. Weber P, Binder K, Krauss S. Why can only 24% solve bayesian reasoning problems in natural frequencies: frequency phobia in spite of probability blindness. *Front Psychol* 2018;9:1833. <https://doi.org/10.3389/fpsyg.2018.01833>
25. Brase GL. Pictorial representations in statistical reasoning. *Appl Cogn Psychol* 2009;23(3):369-81. <https://doi.org/10.1002/acp.1460>
26. Ottley A, Peck EM, Harrison LT, Afergan D, Ziemkiewicz C, Taylor HA, et al. Improving Bayesian reasoning: the effects of phrasing, visualization, and spatial ability. *EEE Trans Vis Comput Graph* 2016;22(1):529-38. <https://doi.org/10.1109/TVCG.2015.2467758>
27. Binder K, Krauss S, Bruckmaier G. Effects of visualizing statistical information – an empirical study on tree diagrams and 2 × 2 tables. *Front Psychol* 2015;6:1186. <https://doi.org/10.3389/fpsyg.2015.01186>
28. Reani M, Davies A, Peek N, Jay C. Evidencing how experience and problem format affect probabilistic reasoning through interaction analysis. *Front Psychol* 2019;10:1548. <https://doi.org/10.3389/fpsyg.2019.01548>
29. Bilalić M, McLeod P, Gobet F. Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition* 2008;108(3):652-61. <https://doi.org/10.1016/j.cognition.2008.05.005>