Poor Individual Risk Classification From Adverse Childhood Experiences Screening

Alan J. Meehan, PhD,[1,2] Jessie R. Baldwin, PhD,[1,3] Stephanie J. Lewis, MRCPsych,[4]
Jelena G. MacLeod, MD,[2,5] Andrea Danese, MD, PhD[1,4,6]

Introduction: Adverse childhood experiences confer an increased risk for physical and mental health problems across the population, prompting calls for routine clinical screening based on reported adverse childhood experience exposure. However, recent longitudinal research has questioned whether adverse childhood experiences can accurately identify ill health at an individual level.

Methods: Revisiting data collected for the Adverse Childhood Experience Study between 1995 and 1997, this study derived approximate area under the curve estimates to test the ability of the retrospectively reported adverse childhood experience score to discriminate between adults with and without a range of common health risk factors and disease conditions. Furthermore, the classification accuracy of a recommended clinical definition for high-risk exposure ($\geq$4 versus 0–3 adverse childhood experiences) was evaluated on the basis of sensitivity, specificity, positive and negative predictive values, and positive likelihood ratios.

Results: Across all health outcomes, the levels of discrimination for the continuous adverse childhood experience score ranged from very poor to fair (area under the curve=0.50–0.76). The binary classification of $\geq$4 versus 0–3 adverse childhood experiences yielded high specificity (true-negative detection) and negative predictive values (absence of ill health among low-risk adverse childhood experience groups). However, sensitivity (true-positive detection) and positive predictive values (presence of ill health among high-risk adverse childhood experience groups) were low, whereas positive likelihood ratios suggested only minimal-to-moderate increases in health risks among individuals reporting $\geq$4 adverse childhood experiences versus that among those reporting 0–3.

Conclusions: These findings suggest that screening based on the adverse childhood experience score does not accurately identify those individuals at high risk of health problems. This can lead to both allocation of unnecessary interventions and lack of provision of necessary support.
Am J Prev Med 2022;62(3):427–432. © 2021 American Journal of Preventive Medicine. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

INTRODUCTION

Adverse childhood experiences (ACEs), including abuse, neglect, and household dysfunction (e.g., parental incarceration, mental illness, substance use), are consistently associated with elevated risk for physical and mental health conditions.[1,2] Given the global ubiquity of violence against children and broader family adversity, various regional, national, and international health bodies have advocated for routine ACE screening.[3,4] These efforts, they argue, would help to mitigate detrimental health effects by identifying

From the [1]Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom; [2]Child Study Center, Yale School of Medicine, New Haven, Connecticut; [3]Clinical, Educational and Health Psychology, UCL Psychology and Language Sciences, University College London, London, United Kingdom; [4]Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom; [5]Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut; and [6]National and Specialist CAMHS Clinic for Trauma, Anxiety, and Depression, South London and Maudsley NHS Foundation Trust, London, United Kingdom

Address correspondence to: Andrea Danese, MD, PhD, Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, 16 De Crespigny Park, London SE5 8AF, United Kingdom. E-mail: andrea.danese@kcl.ac.uk.

individuals who may benefit from preventive interventions or other support services.

However, the clinical utility of widespread ACE screening has been questioned.[5−8] In particular, evidence linking high-risk ACE exposure (i.e., experiencing ≥4 different types of ACEs)[2,9] with ill health is largely based on population-level risk estimates (ORs/risk ratios). These metrics can quantify the mean differences in health risks between groups exposed to ≥4 ACEs and those exposed to 0−3 ACEs. However, to identify at-risk individuals, ACE scores must demonstrate sufficient discrimination (i.e., accurately differentiate between those with and without ill health).[7]

Recent analyses of 2 population-representative birth cohorts found that the ability of the established high-risk cut off to discriminate individuals with later health problems from those without was universally poor, casting doubt on the ability of ACE screening to inform individual risk classification.[10] However, because these general population findings relied on ad hoc ACE measures that only partly reflect current screening practices, there is a need to estimate generalizability to higher-risk, treatment-seeking samples that utilize the conventional ACE questionnaire.

METHODS

Analyses revisited the seminal ACE Study—the first to systematically link ACEs and ill health.[9] In brief, among 8,506 adults (mean age=56.1 years, 52.1% female) who retrospectively reported ACEs after medical evaluation between 1995 and 1997, those reporting ≥4 ACEs (versus those reporting none) were significantly more likely to exhibit the most common health risk factors (e.g., smoking, obesity) and disease conditions (e.g., cancer, stroke) in the U.S.

In 2021, data from the original publication[9] were used to quantify individual discrimination and classification in 2 ways. First, for the continuous ACE score, the approximate area under the curve (AUC)[11] statistics were derived by trapezoidal integration of observed discriminative ability across several dichotomous cut offs. Second, for high-risk ACE exposure (≥4 versus 0−3 ACEs), individual classification was quantified using the following measures: sensitivity, specificity, positive/negative predictive values, and positive likelihood ratios.[12]

RESULTS

Using established interpretive thresholds,[11] overall ACE score discrimination was very poor (AUC=0.5−0.6) for 57.9% (11 of 19) of health outcomes (e.g., smoking, obesity, cancer), poor (AUC=0.6−0.7) for 31.6% (6 of 19) of outcomes (e.g., depressed mood, alcoholism, drug use), and fair (AUC=0.7−0.8) for 10.5% (2 of 19) of outcomes (attempted suicide, drug injection) (Table 1, Panel A). No good-to-excellent discrimination (AUC=0.8−1.0)

was observed. As an illustrative example, an AUC of 0.52 for ischemic heart disease implied that the likelihood of a random adult with ischemic heart disease reporting more ACEs than a random adult without ischemic heart disease was 52% or just above chance (50%).

Meanwhile, the high-risk ACE cut off (≥4 versus 0−3 ACEs) generated low sensitivity (true-positive rate) across outcomes (range=0.07 [cancer] to 0.34 [suicide attempt]; mean=0.14), suggesting that only a small proportion of adults with ill health reported ≥4 ACEs (Table 1, Panel B).[12] By contrast, specificity (true-negative rate) was high across outcomes (range=0.93 [e.g., drug injection, cancer, stroke] to 0.96 [depressed mood]; mean=0.94), implying that the vast majority of those without health problems did not report ≥4 ACEs.

Predictive values (which, unlike sensitivity and specificity, are influenced by the outcome's prevalence) can be used to estimate the probability of detecting an outcome on the basis of a particular test result (i.e., where disease status is unknown). Positive predictive values were low across outcomes (range=0.03 [drug injection] to 0.51 [depressed mood]; mean=0.15), indicating that adults reporting ≥4 ACEs had low rates of health problems.[12] By contrast, high negative predictive values (range=0.79 [physical inactivity] to 0.99 [drug injection]; mean=0.93) implied that most adults reporting <4 ACEs did not have ill health.

Finally, positive likelihood ratios (range=0.97 [cancer] to 5.90 [suicide attempt]) were well below the conventional cut offs for screening tests (likelihood ratio <0.1 or >10).[13] Values indicated that individuals reporting ≥4 ACEs showed only minimal-to-moderate increases in their likelihood of exhibiting poor versus good health, suggesting limited value in using this definition of high-risk ACE exposure to classify ill health.

DISCUSSION

In replicating recent general population findings,[10] results obtained from the seminal ACE Study sample raise further concerns around the ACE score's ability to support individual classification of health risk. Proposed screening based on the advocated ≥4 ACEs cut off may help to rule out ill health among low-risk individuals (0−3 ACEs) on the basis of high specificity and negative predictive values. However, low sensitivity and positive predictive values suggest that such screening would not accurately detect ill health among high-risk individuals (≥4 ACEs), undermining this cut off's utility in allocating preventative interventions. Poor classification estimates may have been influenced by relatively low prevalence rates among some health outcomes (Figure 1,

Table 1. Diagnostic Accuracy for Specific Health Problems Based on ACEs

Health outcomes	Prevalence, %	Panel A: Continuous ACE score AUC[a]	Panel B: Categorical ACE score (≥4 versus 0—3 ACEs)					
			OR (95% CI)	Sens.	Spec.	PPV	NPV	LR+ (95% CI)
Health risk factors								
Current smoker	8.6	0.59	2.23 (1.76, 2.84)	0.13	0.94	0.17	0.92	2.08 (1.68, 2.57)
Severe obesity (BMI ≥35)	7.1	0.58	1.87 (1.42, 2.46)	0.11	0.94	0.12	0.93	1.77 (1.39, 2.26)
No leisure-time physical activity	21.0	0.54	1.38 (1.13, 1.69)	0.09	0.94	0.27	0.79	1.35 (1.12, 1.63)
Depressed mood (≥2 weeks)	22.0	0.65	4.05 (3.39, 4.84)	0.15	0.96	0.51	0.80	3.58 (3.05, 4.20)
Suicide attempt	3.5	0.76	8.42 (6.50, 10.91)	0.34	0.94	0.18	0.97	5.90 (4.92, 7.09)
Alcoholism	5.9	0.68	3.41 (2.65, 4.38)	0.18	0.94	0.16	0.95	2.98 (2.41, 3.67)
Illicit drug use	11.6	0.67	3.31 (2.71, 4.05)	0.16	0.95	0.28	0.89	2.94 (2.47, 3.50)
Drug injection	0.8	0.74	5.02 (2.91, 8.66)	0.26	0.93	0.03	0.99	3.97 (2.65, 5.96)
≥50 lifetime sexual partners	4.4	0.59	1.63 (1.12, 2.38)	0.10	0.94	0.07	0.96	1.57 (1.12, 2.20)
Sexually transmitted disease	8.2	0.61	2.43 (1.91, 3.09)	0.14	0.94	0.17	0.92	2.24 (1.81, 2.76)
Any health risk factor[b]	54.7	0.63	3.10 (2.55, 3.87)	0.10	0.97	0.77	0.47	2.94 (2.41, 3.58)
Disease conditions								
Ischemic heart disease	3.8	0.52	1.58 (1.08, 2.32)	0.10	0.93	0.06	0.96	1.52 (1.08, 2.15)
Cancer	1.9	0.50	0.97 (0.51, 1.85)	0.07	0.93	0.02	0.98	0.97 (0.53, 1.78)
Stroke	2.6	0.51	1.65 (1.06, 2.60)	0.11	0.93	0.04	0.98	1.58 (1.06, 2.38)
Chronic bronchitis/ emphysema	4.0	0.60	2.54 (1.83, 3.53)	0.14	0.94	0.09	0.96	2.32 (1.74, 3.08)
Diabetes	4.3	0.51	1.37 (0.94, 2.00)	0.09	0.93	0.06	0.96	1.33 (0.94, 1.89)
Skeletal fracture	3.9	0.52	1.25 (0.83, 1.89)	0.08	0.93	0.05	0.96	1.23 (0.84, 1.80)
Hepatitis/jaundice	6.5	0.57	1.85 (1.39, 2.47)	0.11	0.94	0.11	0.94	1.76 (1.36, 2.27)
Poor/fair self-rated health	18.2	0.55	1.89 (1.55, 2.30)	0.11	0.94	0.29	0.82	1.80 (1.50, 2.15)

Note: Data were derived from Felitti et al.[9] with ≥4 ACEs used to define a *high-risk ACE exposure* group.
[a]AUC was calculated by comparing sensitivity (true positive rate) and 1-specificity (false positive rate) at each individual ACE cut off (≥1, ≥2, ≥3, ≥4) before consolidating these in an omnibus AUC using trapezoidal integration by the trapz function within the caTools R package. Values range from 0.50 (random chance) to 1.00 (perfect discrimination), with performance interpreted using the following thresholds: very poor (0.5—0.6), poor (0.6—0.7), fair (0.7—0.8), good (0.8—0.9), and excellent (0.9—1.0).[11]
[b]Denotes the presence of at least 1 of the 10 individual health risk factors (smoking, severe obesity, physical inactivity, depressed mood, suicide attempt, alcoholism, illicit drug use, injected drug use, ≥50 lifetime sexual partners, or history of any sexually transmitted disease).
ACE, adverse childhood experience; AUC, area under the (receiver operating characteristic) curve; LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; Sens., sensitivity; Spec., specificity.

light gray and black tiles), which can stifle detection. However, predictive performance was similarly poor for more prevalent outcomes (e.g., depressed mood, physical inactivity). Classification accuracy is also influenced by the distribution of ACEs in the sample: despite the greater increases in risk associated with inclusion in the high-risk group (≥4 ACEs), because most individuals are classed as low risk (0—3 ACEs) (Figure 1, white and light gray tiles), most of those with ill health are similarly found within these low-risk groups (Figure 1, light gray tiles).

From these findings, implementation of ACE screening may have detrimental consequences for resource allocation. Specifically, because most high-risk adults (≥4 ACEs) do not exhibit ill health, screening based on ACE scores alone would generate substantial numbers of false-positive results, potentially leading to inappropriate or unnecessary interventions (e.g., invasive testing/imaging, referral for psychiatric assessment). At the same time, because most adults with ill health were in the low-risk group (0—3 ACEs), such screening efforts would generate a substantial number of false-negative results, potentially excluding many individuals from beneficial health interventions.

The continuous ACE score's poor discrimination suggests that similar consequences would arise from any cut
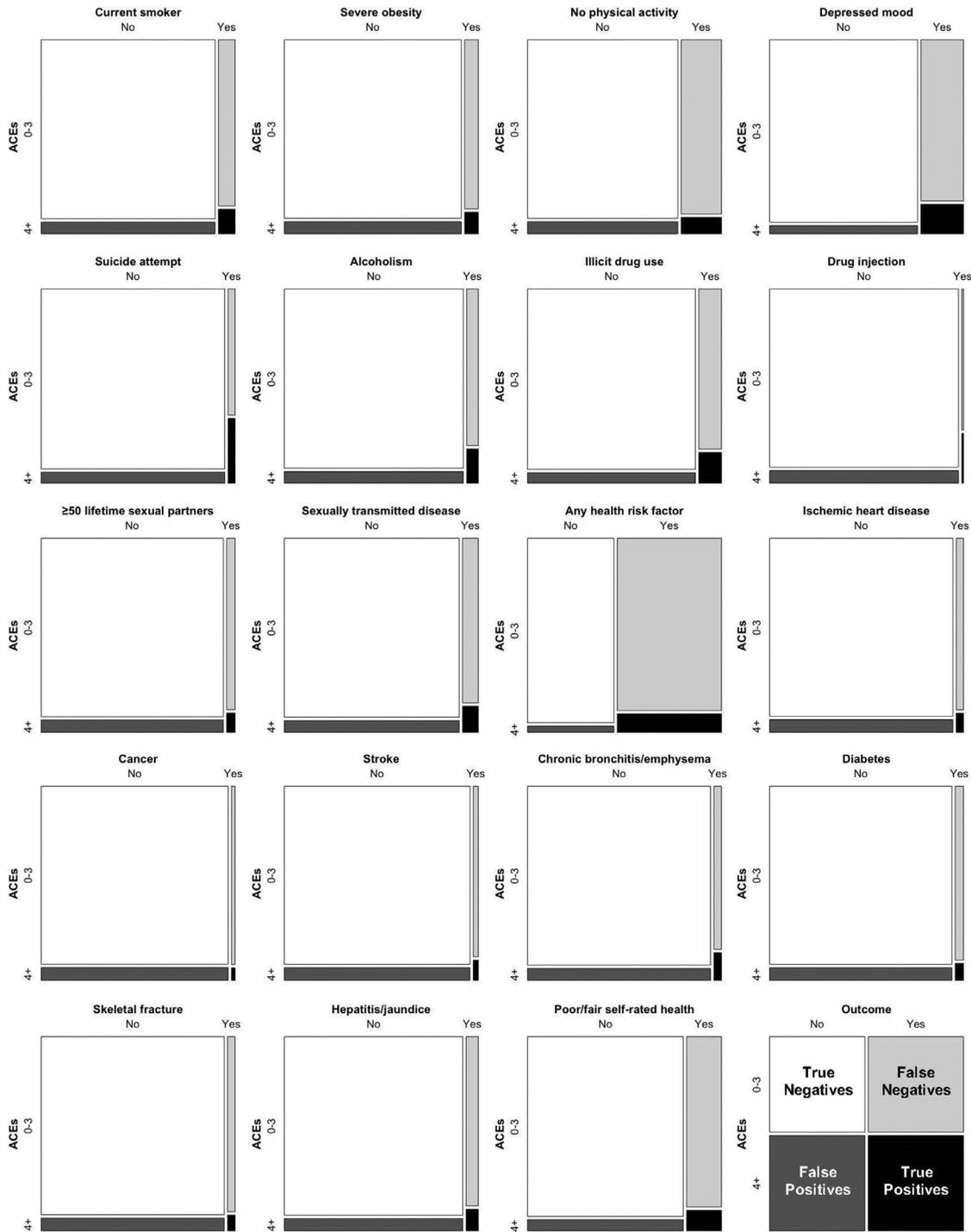
Figure 1. Prevalence of health problems in ACE Study adults reporting 0−3 versus those reporting ≥4 ACEs.
For each mosaic plot, the relative size of individual tiles corresponds to cell frequencies from 2-way contingency tables for each health outcome in the ACE Study. Specifically, tile height is proportional to the number of individuals with 0−3 ACEs versus those with ≥4 ACEs, whereas tile width is proportional to the absence versus presence of the health outcome. As illustrated by the example plot in the bottom right-hand corner, this provides a visualization of the overall proportion of true negative (white tiles), false negative (light gray tiles), false positive (dark gray tiles), and true positive (black tiles) cases for each outcome.
ACEs, adverse childhood experiences.

off applied to this questionnaire. Unlike successful screening tools within cardiology and oncology,[14,15] no AUC values met the suggestive threshold for clinical utility (AUC>0.80). Although reliance on the ACE Study's retrospective adult data limits the ability to draw firm conclusions on the utility of ACE screening for secondary prevention, prospective findings suggest that childhood ACE screening offers a similarly poor prediction of health problems.[10]

To enhance the clinical utility of ACE screening, prediction modeling techniques could be employed to ascribe differential weights to ACE questionnaire items on the basis of their relative contribution to discrimination of people with ill health from those without. Instead of relying on a simple count of exposures, this individualized approach could facilitate more nuanced risk indices from underlying components of the ACE score, without necessitating additional data collection.[6] In addition, multivariable prediction models that combine ACEs with other established health risk and protective factors may enhance individual classification by better capturing the multifactorial causes of disease. Several examples of such models have recently been derived for trauma-exposed children.[16−18] Of note, these ACE-based predictive algorithms should address common methodologic challenges (e.g., overfitting) and demonstrate reliable generalizability to new individuals.[7]

CONCLUSIONS

Although ACE scores are important risk indicators across general and clinical populations, their utility in identifying individuals at heightened risk of ill health appears to be overstated. By shifting toward individualized classification metrics (i.e., discrimination, calibration), considering more sophisticated operationalizations of current ACE measures, and capitalizing on prediction modeling approaches, future research has the potential to improve the identification of individual health risks and, in turn, ensure that ACE screening can provide clinically meaningful health benefits.

CREDIT AUTHOR STATEMENT

Alan J. Meehan: Conceptualization; Formal analysis; Investigation; Validation; Visualization; Writing - original draft. Jessie R. Baldwin: Writing - review & editing.

Stephanie J. Lewis: Writing - review & editing. Jelena G. MacLeod: Writing - review & editing. Andrea Danese: Conceptualization; Formal analysis; Project administration; Investigation; Supervision; Visualization; Writing - original draft.

ACKNOWLEDGMENTS

REFERENCES

1. Bellis MA, Hughes K, Ford K, Ramos Rodriguez G, Sethi D, Passmore J. Life course health consequences and associated annual costs of adverse childhood experiences across Europe and North America: a systematic review and meta-analysis. *Lancet Public Health.* 2019;4 (10):e517−e528. https://doi.org/10.1016/S2468-2667(19)30145-8.
2. Hughes K, Bellis MA, Hardcastle KA, et al. The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. *Lancet Public Health.* 2017;2(8):e356−e366. https://doi.org/10.1016/S2468-2667(17)30118-4.
3. Bhushan D, Kotz K, McCall J, et al. *Roadmap for resilience: the California Surgeon General's report on adverse childhood experiences, toxic stress, and health.* Sacramento, CA: Office of the California Surgeon General, 2020 https://osg.ca.gov/wp-content/uploads/sites/266/2020/12/Roadmap-For-Resilience_CA-Surgeon-Generals-Report-on-ACEs-Toxic-Stress-and-Health_12092020.pdf, Accessed May 25, 2021.
4. Hardcastle K, Bellis MA. *Asking about adverse childhood experiences (ACEs) in general practice: evaluation findings from a pilot study in Anglesey, North Wales.* Wrexham, United Kingdom: Public Health Wales, 2019 https://www.wales.nhs.uk/sitesplus/documents/888/Asking%20about%20ACEs%20in%20General%20Practice.pdf, Accessed May 25, 2021.
5. Finkelhor D. Screening for adverse childhood experiences (ACEs): cautions and suggestions. *Child Abuse Negl.* 2018;85:174−179. https://doi.org/10.1016/j.chiabu.2017.07.016.
6. Lacey RE, Minnis H. Practitioner review: twenty years of research with adverse childhood experience scores - advantages, disadvantages and applications to practice. *J Child Psychol Psychiatry.* 2020;61(2):116−130. https://doi.org/10.1111/jcpp.13135.
7. Danese A. Annual research review: rethinking childhood trauma-new research directions for measurement, study design and analytical strategies. *J Child Psychol Psychiatry.* 2020;61(3):236−250. https://doi.org/10.1111/jcpp.13160.

8. Anda RF, Porter LE, Brown DW. Inside the adverse childhood experience score: strengths, limitations, and misapplications. *Am J Prev Med.* 2020;59(2):293–295. https://doi.org/10.1016/j.amepre.2020.01.009.

9. Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The adverse childhood experiences (ACE) study. *Am J Prev Med.* 1998;14(4):245–258. https://doi.org/10.1016/s0749-3797(98)00017-8.

10. Baldwin JR, Caspi A, Meehan AJ, et al. Population vs individual prediction of poor health from results of adverse childhood experiences screening. *JAMA Pediatr.* 2021;175(4):385–393. https://doi.org/10.1001/jamapediatrics.2020.5602.

11. Pines JM, Carpenter CR, Raja AS, Schuur JD. *Evidence-Based Emergency Care: Diagnostic Testing and Clinical Decision Rules.* 2nd ed. Oxford, United Kingdom: Wiley, 2012.

12. Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. *BMJ.* 2012;345:e3999. https://doi.org/10.1136/bmj.e3999.

13. Parikh R, Parikh S, Arun E, Thomas R. Likelihood ratios: clinical application in day-to-day practice. *Indian J Ophthalmol.* 2009;57(3):217–221. https://doi.org/10.4103/0301-4738.49397.

14. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. *BMJ.* 2012;344:e4181. https://doi.org/10.1136/bmj.e4181.

15. Zappala SM, Scardino PT, Okrongly D, Linder V, Dong Y. Clinical performance of the 4Kscore test to predict high-grade prostate cancer at biopsy: a meta-analysis of U.S. and European clinical validation study results. *Rev Urol.* 2017;19(3):149–155. https://doi.org/10.3909/riu0776.

16. Lewis SJ, Arseneault L, Caspi A, et al. The epidemiology of trauma and post-traumatic stress disorder in a representative cohort of young people in England and Wales. *Lancet Psychiatry.* 2019;6(3):247–256. https://doi.org/10.1016/S2215-0366(19)30031-8.

17. Meehan AJ, Latham RM, Arseneault L, Stahl D, Fisher HL, Danese A. Developing an individualized risk calculator for psychopathology among young people victimized during childhood: a population-representative cohort study. *J Affect Disord.* 2020;262:90–98. https://doi.org/10.1016/j.jad.2019.10.034.

18. Latham RM, Meehan AJ, Arseneault L, Stahl D, Danese A, Fisher HL. Development of an individualized risk calculator for poor functioning in young people victimized during childhood: a longitudinal cohort study. *Child Abuse Negl.* 2019;98:104188. https://doi.org/10.1016/j.chiabu.2019.104188.