Article:
Kerr, W.T., McFarlane, K.N., Allas, C.H. et al. (2026) Quantifying the impact of a computer-aided diagnostic score on the clinical diagnosis of functional seizures. Epilepsia. ISSN: 0013-9580

https://doi.org/10.1002/epi.70069

Received: 28 September 2025 | Revised: 3 December 2025 | Accepted: 4 December 2025

DOI: 10.1002/epi.70069

Epilepsia®

RESEARCH ARTICLE

Quantifying the impact of a computer-aided diagnostic score on the clinical diagnosis of functional seizures

Wesley T. Kerr[1,2,3,4,5] | **Katherine N. McFarlane**[1] | **Corinne H. Allas**[2,3] | **Samuel W. Terman**[4] | **Markus Reuber**[6] | **Sung Hyun Seo**[4] | **Payge Barnard**[4] | **Adriana Y. Koek**[4,7] | **Amir H. Karimi**[2,3,8] | **Siddhika S. Sreenivasan**[2,3] | **Jena S. Grauer**[2,3,9] | **Di Sun**[7] | **Meagan Watson**[10] | **Elissa H. Patterson**[4,11] | **James Castellano**[1] | **Anto I. Bagić**[1] | **L. Brian Hickman**[2] | **Kenneth H. Kutschman**[4] | **Laurence Knowles**[12] | **Alistair Wardrope**[6,13] | **Piyush Ostwal**[14] | **Bridget K. MacDonald**[15,16] | **Melissa Berry**[4] | **Lomalan Reddy**[17] | **Najda Robinson-Mayer**[4,18] | **Nicholas J. Beimer**[4,11] | **Rachna Reddy**[1,19] | **Sarah K. Yaghoubi**[1] | **Laura A. Kirkpatrick**[1,20] | **Danielle R. Carns**[1,21] | **Alex J. Israel**[22] | **Jamie D. Feusner**[23,24] | **Zongqi Xia**[1] | **Yanshan Wang**[5,25,26] | **Laura A. Strom**[10] | **Page B. Pennell**[1] | **John M. Stern**[2] | **Lubomir Hadjiiski**[27] | **William C. Stacey**[4,28]

Correspondence
Wesley T. Kerr, Department of Neurology, University of Pittsburgh, 3471 Fifth Ave., Kaufmann Suite 810, Pittsburgh, PA 15213, USA.
Email: kerrw@pitt.edu

Funding information
National Institutes of Health, Grant/Award Number: K23NS135134, P20NS080131, R01NS033310, R25NS065723, R25NS089450, R90DA022768, R90DA023422, T32GM08042, T90DA022768 and U24NS107158; Christina Louise George Trust; W. M. Keck Foundation; American Academy of Neurology; National Institute of Neurological Disorders and Stroke, Grant/Award Number: K12NS098482

Abstract

Objective: The diagnosis of functional/dissociative seizures (FDS) without ictal video-electroencephalography is challenging. The Functional/Dissociative Seizures Likelihood Score (FSLS) is a machine learning-based diagnostic score that aims to help clinicians identify FDS. We evaluated whether a human-in-the-loop implementation of the FSLS improved the performance of clinicians identifying FDS as compared to epileptic seizures (ES).

Methods: We constructed 117 anonymized cases about patients with ictal video-electroencephalography-documented FDS, epilepsy, co-occurring ES and FDS, or physiological seizurelike events. Text-based clinical history was presented followed by the FSLS. Readers were asked the most likely diagnosis after each piece of information. We used mixture modeling combined with mixed effects logistic regression to perform data-driven grouping of participants based on observed patterns of diagnostic performance.

Results: Overall, 163 readers saw 1142 cases (median = 4 cases/reader), and 146 (90%) had a performance higher than chance. More formal training in seizures was associated with better performance (epileptologist accuracy = 67%, mental

For affiliations refer to page 9.

health clinician accuracy = 52%). Data-driven groups including 66% of readers benefitted from the FSLS (accuracy improvement = 12%–15%, $p < .05$), including those in the reference and near highest baseline performance group. Other groups had no net change in performance ($p > .75$).

Significance: Clinicians with more formal seizure training identified possible FDS more accurately than others, but formal training did not guarantee high diagnostic performance. Two performance-based groups, which included 66% readers, benefitted from the FSLS because they identified when to change their mind on the basis of the FSLS's suggestion. The implementation of machine learning in the diagnosis of FDS should focus on identifying clinical settings where it can effectively enhance clinicians' decision-making.

KEYWORDS

artificial intelligence, epilepsy, machine learning, observer study, psychogenic nonepileptic seizures (PNES)

1 | INTRODUCTION

Previously known as psychogenic nonepileptic seizures, functional/dissociative seizures (FDS) have profound negative impacts on patients' quality of life, capacity for employment, health care utilization, and mortality rate.[1–6] Although the pathophysiology of FDS is not well established, FDS are episodic neurological symptoms that commonly, but not exclusively, were associated with acute and chronic biopsychosocial stressors.[7] The average delay from first seizure to an accurate diagnosis of FDS was 8.6 years (median = 3 years), during which 80% of patients were empirically treated for epileptic seizures (ES) with antiseizure medications.[5,8,9] Even after determination that seizures were functional instead of epileptic with ictal video-electroencephalographic monitoring (VEM), longer delay to diagnosis was associated with worse outcomes.[10,11] Therefore, it is critical to decrease the delay to diagnosis, because shorter delays may contribute to superior treatment outcomes.

To address the delay to diagnosis, we used machine learning to develop the Functional/Dissociative Seizures Likelihood Score (FSLS), which had prospective accuracy of 77% and external validation accuracy of 81%.[12,13] The FSLS was trained on a retrospective VEM-based cohort of 1126 patients and validated with standardized interviews with 490 patients. The FSLS was built on a base of logistic regression to create a weighted combination of features obtained as part of the clinical history to identify patients with "possible" functional seizures (FS).[14] The International League Against Epilepsy (ILAE) certainty level of "possible" was defined as a history raising concern for FDS, but the seizure had not been observed by a clinician or on electroencephalography (EEG).[15] Although the FSLS is not the only clinical score to identify FDS,[16–26] only two scores have been validated prospectively and at an external center.[12,13] There also is evidence that large language models like ChatGPT (version 3.5 and 4) may assist in identifying patients with FDS.[27]

Each approach in identifying FDS may have unique patterns of benefits and limitations.[14,15,27] For example, when clinical epilepsy fellows viewed the same cases as the FSLS, their diagnostic performance was similar, but the errors did not overlap (Cohen kappa = 21%).[14]

Because the pattern of errors from humans and the FSLS did not overlap, we asked if a human-in-the-loop implementation that combined clinicians' impression

Key points

- Higher diagnostic accuracy of FDS and epileptic seizures was partially associated with more formal training in seizures.
- The FSLS improved the performance of two groups (66%) of readers: the average participants and near highest participants.
- The FSLS did not improve the accuracy of the highest performance group (Group 1) or intermediate groups (Groups 3 and 4).
- FSLS improved the average group's accuracy from 41% to 60%, compared to intermediate groups (55%–64%) and high-performing groups (72%–82%).
- Clinical history supplemented by the FSLS was insufficient to diagnose FS; further neurodiagnostic testing is required.

with the FSLS could improve diagnostic performance. Human-in-the-loop implementation emphasizes that machine learning and artificial intelligence likely are not accurate or reliable enough to be used in health care without human supervision; therefore, we evaluate how clinicians' insights improve upon the machine learning approach and vice versa.[28,29] Evaluation of the utilization of artificial intelligence medicine has identified distinct groups of clinicians: those who use artificial intelligence effectively, those who did not benefit from artificial intelligence support, and those who overrelied on artificial intelligence. Based on those groups, we hypothesized that epileptologists and those with specific expertise in FDS may have high enough pre-FSLS performance that they may not benefit from the FSLS, or they may trust their own expertise more than an unfamiliar FSLS.

Alternatively, clinicians with less expertise in FDS may benefit from the FSLS if they identify how to leverage its strengths while recognizing its limitations. However, that balanced approach to clinician–artificial intelligence collaboration may require high levels of experience and literacy with artificial intelligence. If clinicians with less expertise in FDS benefit from the FSLS, then subsequent naturalistic implementation studies could evaluate whether the FSLS addresses a primary challenge of delayed diagnosis: referral to epilepsy specialists or performing more definitive neurodiagnostic testing.[8] To evaluate these questions initially in a structured experimental setting, we developed online case modules based on real patient examples and asked readers to differentiate between ES and FDS based on either clinical history alone or clinical history supplemented by the FSLS.

2 | MATERIALS AND METHODS

2.1 | Patients on whom the case modules were based

We created online case modules regarding adult (>18 years old) patients with seizurelike movements with definitive diagnosis based on ictal VEM at a Level 4 Comprehensive Epilepsy Center during 2017–2019. To ensure the prevalence of each seizure type was maintained, patients were selected in reverse sequential order (most recent first) from the patients with "documented" certainty of FDS, EEG-observed ES, co-occurring FDS and ES, or nonepileptic nonfunctional physiological seizurelike events (PSLE, e.g., convulsive cardiogenic syncope), and with prospective standardized interviews that were used in the validation of the FSLS.[14,15] The patient portion of this study was approved by the University of California, Los Angeles (UCLA) Institutional Review Board (IRB#11-000916). The case module portion of this study was approved by IRBs at UCLA, University of Michigan, and University of Pittsburgh. For more information about patients, clinical history, and patient selection, see Supplemental Methods.

Online case modules were created based on a combination of manually deidentified and anonymized retrospective chart review and the prospective standardized interview. The section we call the history of present illness (HPI) includes all clinical information from the earliest neurologists' clinical notes, excluding EEG and neuroimaging results. The FSLS was calculated based on information obtained from the standardized interview conducted by research staff to obtain the clinical history (see Supplemental Methods for details).[14]

2.2 | Case module information

These cases were organized into case modules within Qualtrics with the three sections of HPI, FSLS, and the results of VEM (Figure 1). Readers were asked two questions after each sequential section: (1) "Which diagnosis is most likely?" (a) epileptic seizures, (b) functional seizures, (c) co-occurring epileptic and functional seizures, and (d) physiological seizurelike events; and (2) "What is your certainty or confidence in this diagnosis?" using a 5-point Likert scale. When the FSLS information was provided, the readers were told the predicted probability of the most likely prediction.[14] After each case, readers were shown the correct diagnosis based on subsequent VEM. For additional details about the case modules and the use of GPT-4 for comparison, see Supplemental Methods.[27]

2.3 | Readers

Readers were intended to be clinicians or researchers in seizures and were recruited through multiple sources to participate in the online case modules including word of mouth, emails to training program directors, academic presentations, and social media (see Supplemental Methods). Nonclinicians (e.g., patients and advocates) were included to estimate baseline performance that may be improved upon with clinical training. We categorized readers based on self-reported formal clinical training.

2.4 | Statistical methods

We evaluated the impact of clinical expertise, the FSLS, and the interaction between clinical expertise and the FSLS on diagnostic performance using mixed-effected

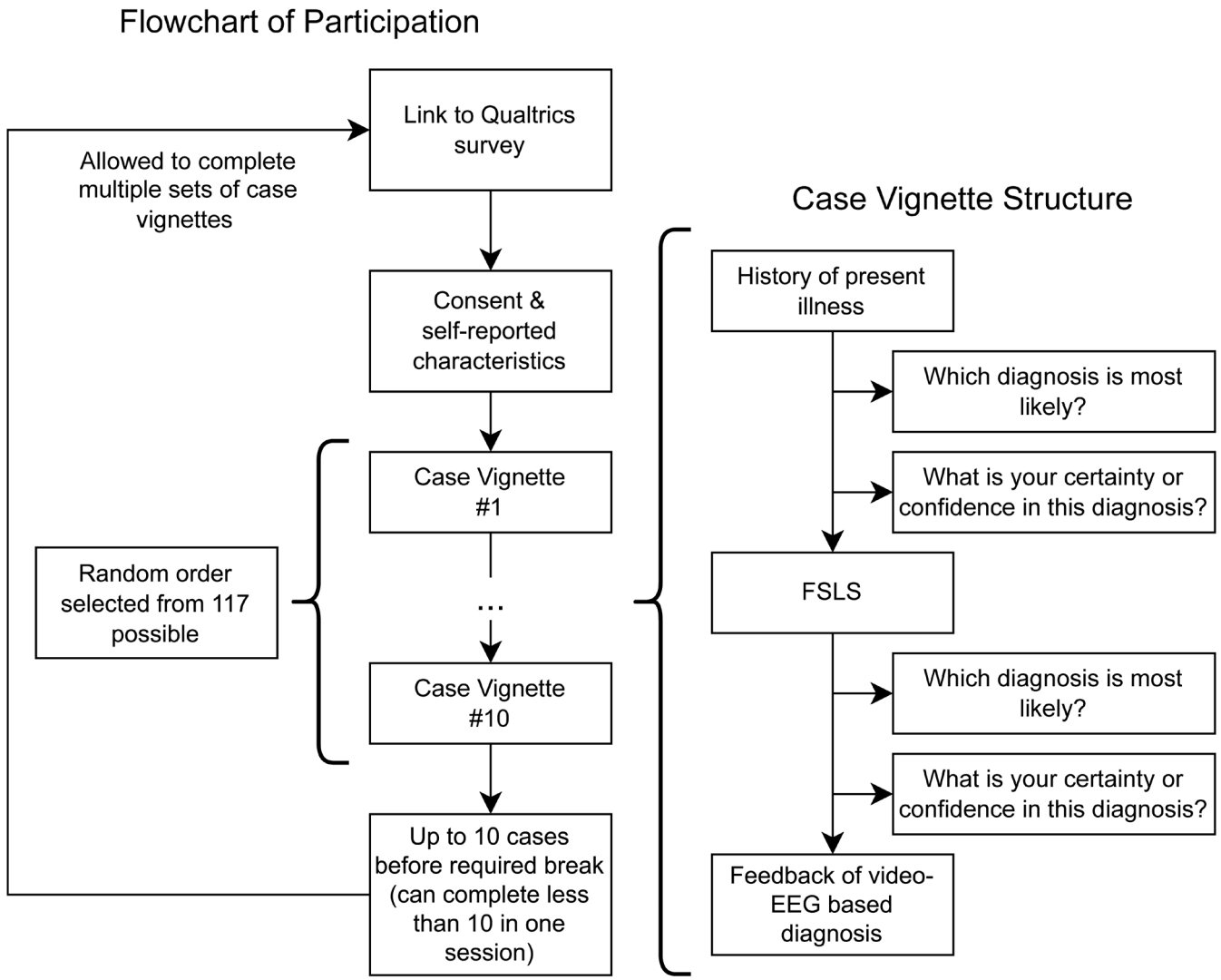

FIGURE 1 Flowchart of participation in the case modules. EEG, electroencephalography; FSLS, Functional Seizures Likelihood Score.

logistic regression predicting whether the reader made the correct diagnosis with crossed random effects accounting for within reader, within case, and repeated measures variability (R package lme4).[30,31] To perform data-driven grouping of participants based on observed level of performance, we superimposed mixture modeling upon this logistic regression model and forward selection to determine the number of groups (R package mixture; see Supplemental Methods).[32] Mixture modeling hypothesizes that the observed performance may be a mixture of performance patterns that, originally, come from more than one group. Based on the similarities of readers' performances across individual cases using clinical history alone or clinical history assisted by the FSLS, mixture modeling evaluated whether there were groups of readers who had different patterns of diagnostic performance and interaction with the FSLS. The area under the receiver operating curve was built based on certainty ratings.

We measured intraparticipant variability and learning from the educational feedback at the end of each case with a fixed effect term to indicate repeat views of the same case. Repeats occurred randomly when participants completed more than one session of case modules (see Supplemental Methods) and were a "positive control" where readers had additional insights from remembering the patient, as compared to the additional insights from the FSLS.

We evaluated the changes in reader-predicted diagnosis after viewing the FSLS. In this change analysis, we evaluated (1) how often readers changed their diagnosis after viewing the FSLS and (2) rates of change based on whether the FSLS was correct, the reader was correct, and the true diagnosis was either FDS or ES. The Supplemental Methods describes other details. We defined the following terms to describe these changes:

- Persuaded: An initially incorrect reader changed their diagnosis to agree with a correct FSLS.

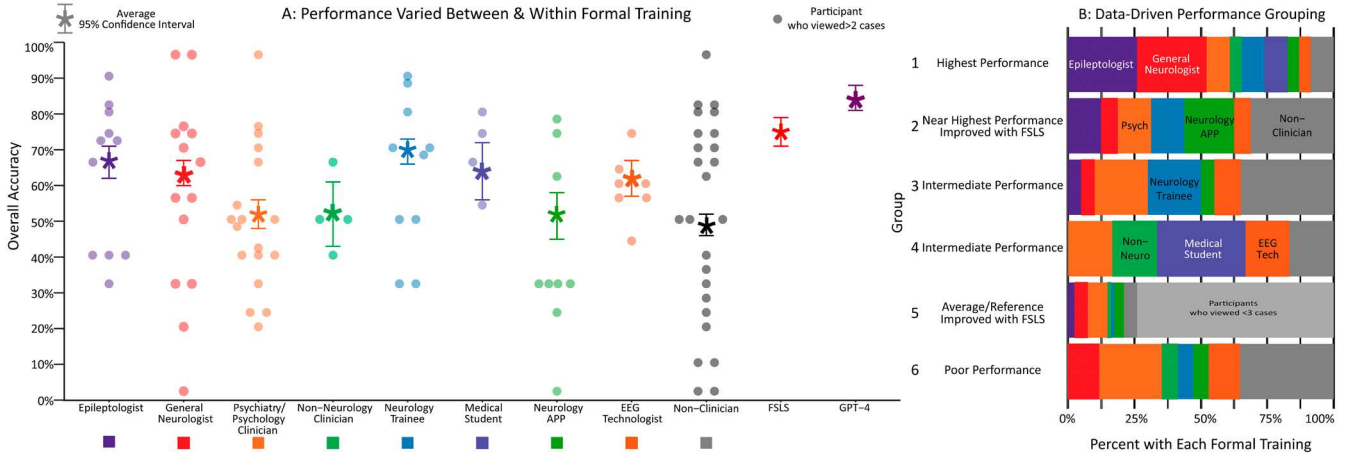

FIGURE 2 (A) Performance on cases varied between readers both between and within formal training categories. Only participants who viewed more than two cases are displayed. (B) Mixture modeling used performance data to form six groups irrespective of formal training. Additional performance metrics are described in Figure S2, Tables S1 and S2. APP, advanced practice provider; EEG, electroencephalography; FSLS, Functional Seizures Likelihood Score; GPT, generative pretrained transformer; Non-Neuro, nonneurology clinician; Tech, technologist.

- Misled: An initially correct reader changed their diagnosis to agree with an incorrect FSLS.
- Overruled: An initially correct reader did not change their diagnosis to agree with an incorrect FSLS.
- Not convinced: An initially incorrect reader did not change their diagnosis to agree with a correct FSLS.
- Error: An initially correct reader viewed a correct FSLS but erroneously changed their diagnosis.

3 | RESULTS

3.1 | Patient and reader information

The reverse sequential selection of patients with diagnostic VEM yielded 117 unique patients, comprised of 85 (73%) patients with ES only, 26 (22%) patients with FDS only, four (3%) patients with co-occurring ES and FS, and two (2%) patients with PSLE (Table S1).

Through all methods of recruitment, 163 unique readers saw at least one case (see Figure S1 for STROBE flowchart). The participants self-identified as follows: 17 epileptologists, 19 general neurologists, 22 psychiatry or psychology clinicians, 11 nonneurology clinicians, 12 neurology trainees (four epilepsy fellows, eight residents), four medical students, 18 neurology advanced practice providers (APPs), nine EEG technologists, and 51 nonclinicians (Table S2).

There were 1142 total case–reader pairs of diagnoses (median = 4 cases per reader, interquartile range = 1–8; 28% of readers viewed only one case; see Figure S1). The 117 unique cases were viewed a median of 10 times (interquartile range = 8–11). Readers learned from case-specific feedback (odds ratio = 2.23, 95% confidence interval = 1.07–4.67, $p = .033$; Figure S5). Before viewing the FSLS, readers' diagnoses had fair correspondence with the FSLS (Cohen kappa = 21.8%).

3.2 | Associations of performance with self-reported clinical expertise

Readers with more formal training in seizures had higher diagnostic performance than those with less formal training, but there was substantial variability (Figure 2; leaderboard in Table S8). The formal training category was the only reader characteristic associated with different performance (Figure S2, Tables S2 and S3). Within clinicians with formal training in neurology (accuracy = 66%, 327/492 case–reader pairs), there was no significant difference in diagnostic performance based on level of training (e.g., residency, fellowship, general neurology, epileptology). The pooled diagnostic accuracy of these neurology/seizure specialists was higher, although not significantly, than all other readers (66% vs. 53%, $p = .078$). The performance of the FSLS and GPT-4 are provided for comparison (Table S2).[27]

3.3 | Data-driven grouping of impact of FSLS on readers' performance

Readers' self-rater certainty reflected that when the reader felt they were more confident, then the likelihood of their chosen diagnosis was higher in most, but not all, cases (Figure 3A). Viewing the FSLS improved that relationship

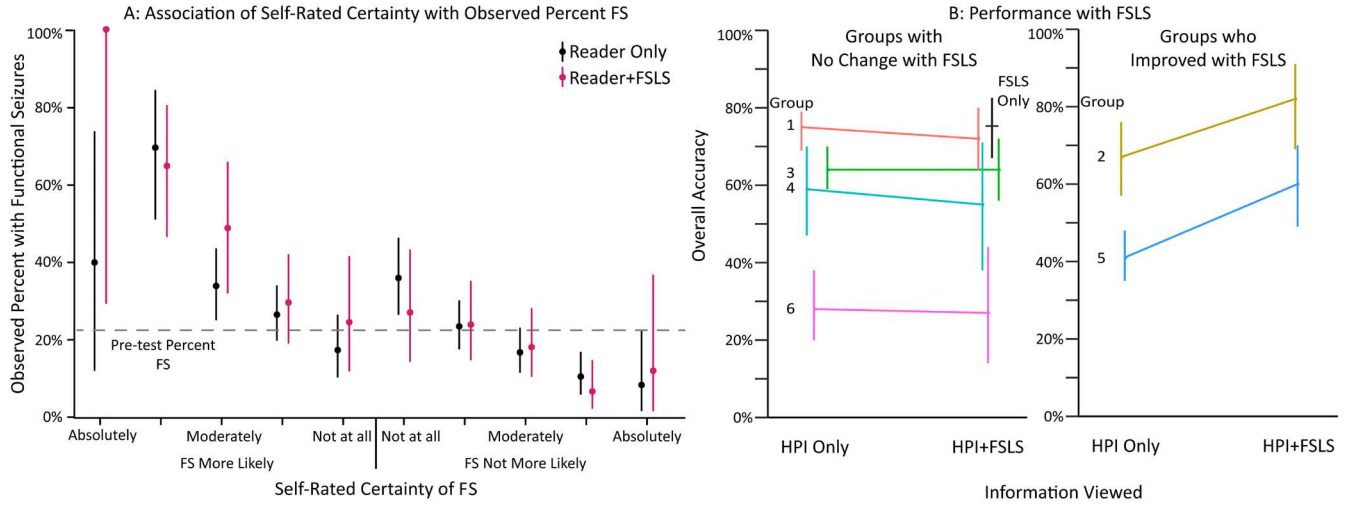

FIGURE 3 Performance changed after the Functional Seizures Likelihood Score (FSLS) was displayed, with higher predictive value for cases with high certainty of functional seizures (FS). (A) Predictive value of self-rated certainty with the observed percent FS. (B) Change in performance was different across data-driven groups. Raw FSLS-only performance is illustrated for comparison. Other performance metrics are provided in Figures S3 and S4. Error bars indicate 95% confidence intervals. HPI, history of present illness.

TABLE 1 Descriptive names of the data-driven groups and the change (Δ) in accuracy with the FSLS.

		Accuracy		
Data-driven group	**Descriptive name**	**HPI, %**	**HPI + FSLS, %**	**ΔFSLS, %**
1	Highest performance	75	72	−3
2	Near highest performance	67	82	15
3	Intermediate performance	64	64	0
4	Intermediate performance	59	55	−4
5	Reference/average	41	60	19
6	Poor performance	28	27	−1

Note: For additional performance metrics, see Table S7.

Abbreviations: FSLS, Functional Seizures Likelihood Score; HPI, history of present illness.

between certainty and likelihood of identifying the correct diagnosis (Figures 3A and S3).

Mixture modeling identified six groups of participants based on significantly different observed performance (Figure 3B, Table S4). Our descriptive names for each group reflect a gradient of performance (Figure 2B, Tables 1 and S5). Most (61%) of the highest performance group (Group 1) were seizure subspecialists (26% [6/23] epileptologists, 26% [6/23] neurologists, 9% [2/23] neurology residents or fellows). Of the remaining nine readers in Group 1, at least six had special experience in FS. The poor performance readers (Group 6) were comparable or worse than chance and were excluded from other analyses.

After viewing the FSLS, participants in the reference/average group (Group 5) had a 15% increase in accuracy (odds ratio = 2.44, 95% confidence interval = 1.15–5.15, $p = .02$), whereas the near highest performance group (Group 2) had a 12% increase in accuracy (odds ratio = 3.02, 95% confidence interval = 1.05–8.70, $p = .041$; Figure 3B, Table S6). The other groups (Groups 1, 3, 4, and 6) had a net accuracy change of <5% (p-value of change > .75).

The accuracy of the near highest performance group (Group 2) improved because they were persuaded by the FSLS, but they were never misled (Figures 4 and S4–S6). The accuracy of the reference group (Group 5) improved by 21% due to persuasion by the FSLS and reduced by 4% due to being misled. Readers and FSLS disagreed less often for higher performing groups than lower performing groups (Groups 1–3: 37%, Groups 4–5: 53%). When readers and FSLS disagreed, higher performing groups changed their diagnostic impression less often (Groups 1–3: 22%, Groups 4–5: 42%).

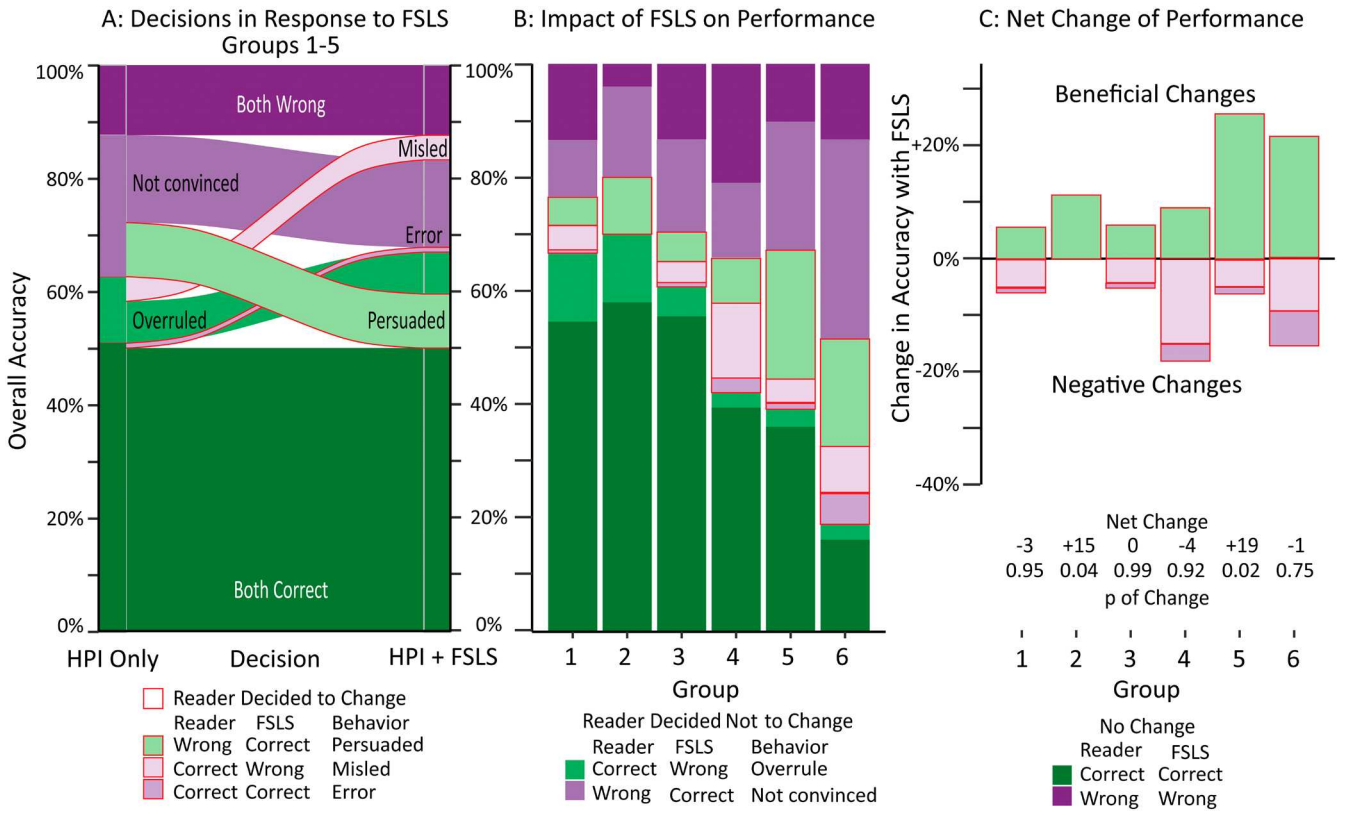

FIGURE 4 The impact of readers' decisions in response to the Functional Seizures Likelihood Score (FSLS) on performance varied based on data-driven group. Groups 2 and 5 were persuaded more often than misled by the FSLS, but other groups had no significant net change. (A) Alluvial to illustrate decisions in response to FSLS. (B) Potential impact of FSLS on overall accuracy are illustrated in lighter shades. Decisions to change initial impression are outlined in red. (C) Net impact of decisions on overall accuracy (*p*-values reflect mixed logistic models). HPI, history of present illness.

4 | DISCUSSION

Written reports of clinical history indicative of "possible" FDS should prompt early and accurate referral for more definitive neurodiagnostic testing (e.g., video-EEG monitoring), because diagnostic accuracy ranged from 49% (nonclinicians) to 67% (epileptologists). Accurate identification was challenging, even for the highest performing participants (accuracy = 75%). Machine learning-assisted identification with the FSLS improved the performance of groups containing 66% of readers by 12%–15%, but other readers could not effectively identify when to change their diagnosis based on the FSLS. These online case modules provide unique insights regarding how human-in-the-loop machine learning can integrate into clinical pathways.

4.1 | Diagnostic implications of "possible" FDS

This study focuses on the highest ILAE certainty of FDS that could be achieved by clinical history, which is called "possible."[15] When readers in our performance-based seizure expert groups felt that FDS was "possible," they were correct 63% and 53% of the time for Group 1 and 2, respectively (predictive value). They also identified 70% and 63% of patients with FDS, respectively (sensitivity). That performance was similar to other evaluations that used clinical history.[33]

However, the difference between self-identified formal training and performance-based grouping highlighted that formal training in epilepsy or neurology was not sufficient to be in the highest performing groups. There was insufficient evidence to conclude that 27% (3/11) of epileptologists and 50% (7/14) of neurologists had high enough performance to earn a place in the two top-performing data-driven groups.

The average reader, who may represent a generic clinician without specific formal or practice experience in seizures, had quite poor diagnostic performance. When the 55% of readers with average performance (Group 5) felt FDS was "possible," they were correct in 41% of patients (predictive value), and they only identified 37% of patients with FDS (sensitivity). These poor results mirror video-based studies of nonseizure specialists (e.g.,

emergency physicians), who could have accuracies as low as 45%.[15,33–35]

These limited diagnostic performances indicated that "possible" FDS could prompt referral for further neurodiagnostic testing, but this level of evidence was not high enough to impact treatment decisions. Because shorter time to diagnosis may improve long-term outcomes, patients identified with "possible" FDS should promptly pursue more definitive neurodiagnostic evaluation.[10,11] The ILAE diagnostic criteria for FDS recommends observation of seizures by a seizure specialist with video, EEG, or video-EEG.[14,15,33] Video observation of a seizure by a seizure specialist corresponded with the ILAE certainty category of "clinically established" (predictive value = 92%, sensitivity = 93%).

4.2 | Machine learning-assisted diagnosis

To focus on the goal of early clinical triage of "possible" FDS to more definitive neurodiagnostic evaluation, this study directly addressed how machine learning integrated with clinicians using a human-in-the-loop design by identifying when the FSLS produced a net benefit, no change, or had potential harm.[36] The human-in-the-loop design differs from when machine learning tools aim to replace or compete with clinicians.[37]

Groups including 66% of readers utilized the FSLS effectively as a supportive tool by leveraging its strengths while maintaining critical oversight and clinical reasoning.[29] The group of average readers (Group 5) was able to understand the FSLS enough to improve their accuracy from 44% to 60%. Although that was lower than the 76% accuracy of the FSLS alone, it was a significant improvement ($p = .02$). In addition, the near highest performance group (Group 2) improved their accuracy from 70% to 82% with the FSLS ($p = .04$). That combined performance was numerically, but not statistically, higher than the FSLS alone (76%) or Group 1. In other studies of the integration of artificial intelligence with clinical insights, clinicians who effectively used artificial intelligence had higher artificial intelligence literacy and more experience and expressed a balanced approach toward human–artificial intelligence collaboration.[28,29,38]

In contrast, the remaining groups with 44% of readers did not significantly benefit from FSLS support. Similar to other studies, this lack of benefit may have been due to lack of trust, poor understanding of the FSLS's predictions, or misalignment between the FSLS' recommendations and clinical reasoning.[29] When readers were wrong (and the FSLS was correct), readers were persuaded in only 30%–45% of cases. The highest performance group (Group 1) may not have been convinced because they felt that they had more nuanced knowledge than the FSLS.[39] However, the fair Cohen kappa of 32% between the FSLS and the highest performance group (Group 1) indicated that the FSLS may provide some unique information that could supplement, not replace, their clinical expertise. Even though Group 1 did not recognize those unique insights, the near highest performance group (Group 2) successfully improved their performance by recognizing and effectively utilizing that supplemental information provided by the FSLS.

Readers with intermediate performance may have enough seizure expertise to identify FDS better than the average reader, but the depth of their knowledge was not enough to recognize when to be persuaded by the FSLS. Even though their performance did not improve, these intermediate performance readers tried to use the FSLS more than the higher performers because the intermediate readers changed their impression after viewing the FSLS more often (40% vs. 20%). Readers in all groups recognized that the FSLS supplemented, and did not replace, their diagnostic impression; no reader "misskilled," which is defined by adopting the FSLS's errors uncritically.[29] No reader who viewed a disagreeing FSLS more than twice uniformly changed their impression to match the FSLS. Therefore, the readers attempted to benefit from the FSLS, but many did not recognize when it supplemented their clinical insights. Further exposure to and training on the FSLS may permit these readers to identify how to effectively use the FSLS.

Lastly, we must consider the potential harm of utilizing the FSLS through "deskilling."[29] Deskilling is defined as the loss of clinical reasoning skills.[29] Although the net accuracy of readers in intermediate Group 4 reduced numerically by 4% (post-FSLS difference, $p = .92$), that difference may have corresponded to a shift in the threshold of concern for FDS, where they increased the overall rate at which they predicted FDS, which resulted in 19% increase of the FDS predictive value plus a corresponding 10% decrease of the epilepsy predictive value (Figure S4, Table S7). That shift in the threshold of concern without a substantial increase in overall performance was reflected by a net 1% increase in the area under the receiver operating curve (Table S7).

An alternative potential harm of the FSLS may be perpetuation of harmful historical practices. Initially, we were concerned about the ethics of the FSLS because it included components of patient sex and the presence of psychosocial trauma, including sexual trauma.[40,41] Overemphasis of those factors could have perpetuated sex-driven disparities and stigmatization of victims in health care.[42–45] However, we did not observe worsening performance when the FSLS was provided, which

indicated that these initial concerns were not seen in the data.

There are some limitations to our approach of online case modules and our analyses. Our use of observed performance to group participants highlighted that formal training was not a reliable indicator of how they would utilize the FSLS. Future studies are needed to identify characteristics that may predict which clinicians may benefit most from the FSLS. To promote patient anonymity and recruit more readers with broad clinical background, these case modules excluded seizure videos, which, with specific ictal behaviors, can increase the ILAE certainty of FDS to clinically established.[15] The estimated response rate was <1%, and 28% of readers only observed a single case, which may have contributed to a selection bias. Additionally, the limited number of cases for each reader led to insufficient data to definitively categorize individual readers based on whether they did or did not benefit from the FSLS. For example, one third of epileptologists and half of board-certified neurologists were included in the reference group due to insufficient evidence to put them in higher performing groups. Therefore, we primarily focus on the interpretation of the composition of the groups with sufficient evidence that they were different from the reference groups. We also had insufficient readers from key categories of clinicians who care for patients with FDS, including emergency physicians, emergency nurses, emergency medicine technicians, and primary care providers.[34] This evaluation focuses on reader-associated performance. The additional complexity of addressing case-associated drivers of performance can be addressed in future work. Whereas the proportion of patients with FDS matched patients admitted for VEM, the prevalence of FDS was much lower in non-VEM settings.[46,47] This evaluation also focused on adults with seizures, and the identification of pediatric FDS differs from adult FS.

5 | CONCLUSIONS

The diagnosis of FDS based on written history alone was challenging, even for seizure specialists. The machine learning-based FSLS improved the diagnostic accuracy of groups including 66% of readers by 12%–15% and also improved other diagnostic performance measures. This differential benefit of the FSLS in some readers indicated that efforts to implement the FSLS should focus on targeted groups of clinicians who can effectively understand its benefits and limitations. The predictive value of "possible" FDS was insufficient for a reliable diagnosis; therefore, patients should pursue further neurodiagnostic testing for a more definitive diagnosis (e.g., VEM).

AUTHOR CONTRIBUTIONS

All authors reviewed and approved the final version of the manuscript as well as at least one earlier draft of the manuscript. Wesley T. Kerr was responsible for all aspects of the manuscript from start to finish. Katherine N. McFarlane performed the statistical analysis. Corinne H. Allas created the Qualtrics. Samuel W. Terman contributed to study design. Markus Reuber assisted with design and recruitment of participants, especially those within the United Kingdom. Sung Hyun Seo, Payge Barnard, Adriana Y. Koek, Amir H. Karimi, Elissa H. Patterson, James Castellano, Anto I. Bagić, L. Brian Hickman, Kenneth H. Kutschman, Laurence Knowles, Alistair Wardrope, Piyush Ostwal, Bridget K. MacDonald, Lomalan Reddy, Nicholas J. Beimer, Sarah K. Yaghoubi, Laura A. Kirkpatrick, Danielle R. Carns, Alex J. Israel, Siddhika S. Sreenivasan, Jena S. Grauer, Di Sun, Meagan Watson, Melissa Berry, Najda Robinson-Mayer, and Rachna Reddy assisted with representation from diverse training levels and the interpretation of these results for nonseizure experts. Jamie D. Feusner, Zongqi Xia, Yanshan Wang, Laura A. Strom, Page B. Pennell, and Lubomir Hadjiiski assisted with interpretation of how the computer-aided diagnostic score would integrate with clinical care. John M. Stern, William C. Stacey, Markus Reuber, and Wesley T. Kerr were principal investigators on relevant regulatory bodies (e.g., institutional review boards) and assisted throughout multiple aspects of design.

AFFILIATIONS

[1]Department of Neurology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[2]Department of Neurology, University of California, Los Angeles, Los Angeles, California, USA
[3]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, California, USA
[4]Department of Neurology, University of Michigan, Ann Arbor, Michigan, USA
[5]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[6]Academic Neurology Unit, University of Sheffield, Royal Hallamshire Hospital, Sheffield, UK
[7]Department of Neurology, University of California, San Francisco, San Francisco, California, USA
[8]Department of Anesthesiology, Harbor–University of California, Los Angeles Medical Center, West Carson, California, USA
[9]Oregon Health Science University, School of Medicine, Portland, Oregon, USA
[10]Department of Neurology, University of Colorado, Aurora, Colorado, USA
[11]Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA
[12]Department of Neurology, University of Exeter, Exeter, UK
[13]Department of Neurology, Sheffield Teaching Hospitals National Health Service Foundation Trust, Sheffield, UK
[14]Department of Neurology, Latrobe Regional Hospital, Traralgon, Victoria, Australia

[15]Department of Neurology, Croydon University Hospital, Thornton Health, UK
[16]Department of Neurology, St. George's Hospital, London, UK
[17]Community Epilepsy Service, Bradford Teaching Hospitals National Health Service Foundation Trust, Bradford, UK
[18]Department of Social Work, University of Michigan, Ann Arbor, Michigan, USA
[19]Lake Erie College of Osteopathic Medicine, Erie, Pennsylvania, USA
[20]Department of Pediatrics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[21]Department of Neurology, Emory University, Atlanta, Georgia, USA
[22]Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[23]Department of Psychiatry, Division of Neurosciences and Clinical Translation, University of Toronto, Toronto, Ontario, Canada
[24]Centre for Addiction and Mental Health, Toronto, Ontario, Canada
[25]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[26]Department of Health Information Management, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[27]Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA
[28]Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan, USA

ACKNOWLEDGMENTS

We thank each of our readers for their time and expertise in contributing to this work (see Supplemental Tables for anonymous usernames). We appreciate the contribution of Anuj Patel. This work was supported by the US National Institute of Neurological Disorders and Stroke (NIH K23NS135134, R25NS065723, R25NS089450, U24NS 107158, R01NS033310, P20NS080131, T32GM08042, T90DA022768, R90DA022768, R90DA023422, K12NS 098482), the American Academy of Neurology, the American Brain Foundation, the Epilepsy Foundation, the American Epilepsy Society, the Epilepsy Study Consortium, the William M. Keck Foundation, and the Christina Louise George Trust.

CONFLICT OF INTEREST STATEMENT

W.T.K. has received compensation as an associate editor of *Epilepsia*; writes review articles for *Medlink Neurology*; is a paid consultant for SK Life Sciences, UCB Pharmaceuticals, Jazz Pharmaceuticals, Azurity, Acuta, Ventus, Capsida, Epygenix, Biohaven Pharmaceuticals, the Epilepsy Study Consortium, Cerebral Therapeutics, Neurelis, Noema, EpiTel, QurAlis, Neurona, Neuropace, and Rapport; has collaborative or data use agreements with Eisai, Janssen, Johnson & Johnson, Praxis, Radius Health, and GSK; and has been a site investigator for a trial including UCB Pharmaceuticals and Equilibre Pharmaceuticals. The other coauthors have no relevant disclosures. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

DATA AVAILABILITY STATEMENT

Concurrently with publication of this article, the deidentified data and code from this work will be available on Mendeley Data. The text of the case modules can be made available upon request. Four example cases are available in the Supplemental Materials.

ORCID

Wesley T. Kerr https://orcid.org/0000-0002-5546-5951
Katherine N. McFarlane https://orcid.org/0009-0000-0365-813X
Corinne H. Allas https://orcid.org/0000-0002-7329-480X
Samuel W. Terman https://orcid.org/0000-0001-6179-9467
Markus Reuber https://orcid.org/0000-0002-4104-6705
Sung Hyun Seo https://orcid.org/0000-0001-7723-9317
Payge Barnard https://orcid.org/0009-0009-1314-9887
Adriana Y. Koek https://orcid.org/0000-0002-4806-3743
Amir H. Karimi https://orcid.org/0000-0003-3560-9433
Siddhika S. Sreenivasan https://orcid.org/0000-0002-3353-5031
Jena S. Grauer https://orcid.org/0000-0001-8272-5221
Di Sun https://orcid.org/0000-0001-6578-5142
Meagan Watson https://orcid.org/0000-0002-2977-681X
Elissa H. Patterson https://orcid.org/0000-0002-8892-6435
James Castellano https://orcid.org/0000-0003-1052-8725
Anto I. Bagić https://orcid.org/0000-0002-6284-8336
L. Brian Hickman https://orcid.org/0000-0002-7685-9562
Laurence Knowles https://orcid.org/0000-0003-2316-4676
Alistair Wardrope https://orcid.org/0000-0003-3614-6346
Piyush Ostwal https://orcid.org/0000-0001-6057-2578
Bridget K. MacDonald https://orcid.org/0000-0002-9435-5157
Melissa Berry https://orcid.org/0000-0002-4288-7557
Nicholas J. Beimer https://orcid.org/0000-0002-8054-5769
Laura A. Kirkpatrick https://orcid.org/0000-0002-8259-5997
Danielle R. Carns https://orcid.org/0009-0009-9382-8993
Alex J. Israel https://orcid.org/0000-0001-9429-5089
Jamie D. Feusner https://orcid.org/0000-0002-0391-345X
Zongqi Xia https://orcid.org/0000-0003-1500-2589
Yanshan Wang https://orcid.org/0000-0003-4433-7839

Laura A. Strom https://orcid.org/0000-0002-3297-5908
Page B. Pennell https://orcid.org/0000-0003-0568-2795
John M. Stern https://orcid.org/0000-0002-3549-1642
Lubomir Hadjiiski https://orcid.org/0000-0003-2069-8066
William C. Stacey https://orcid.org/0000-0002-8359-8057

REFERENCES

1. Seneviratne U, Low ZM, Low ZX, Hehir A, Paramaswaran S, Foong M, et al. Medical health care utilization cost of patients presenting with psychogenic nonepileptic seizures. Epilepsia. 2019;60(2):349–57.
2. Tan M, Pearce N, Tobias A, Cook MJ, D'Souza WJ. Influence of comorbidity on mortality in patients with epilepsy and psychogenic nonepileptic seizures. Epilepsia. 2023;64(4):1035–45.
3. Zhang L, Beghi E, Tomson T, Beghi M, Erba G, Chang Z. Mortality in patients with psychogenic non-epileptic seizures a population-based cohort study. J Neurol Neurosurg Psychiatry. 2022;93(4):379–85.
4. Nightscales R, McCartney L, Auvrez C, Tao G, Barnard S, Malpas CB, et al. Mortality in patients with psychogenic nonepileptic seizures. Neurology. 2020;95(6):e643–e652.
5. Kerr WT, Sreenivasan SS, Allas CH, Janio EA, Karimi AH, Dubey I, et al. Title: functional seizures across the adult lifespan: female sex, delay to diagnosis and disability. Seizure. 2021;91:476–83.
6. Hingray C, Popkirov S, Kozlowska K, Pretorius C, Sarudiansky M, el-Hage W, et al. Functional/dissociative seizures: proposal for a new diagnostic label and definition by the ILAE task force. Epilepsia. 2025;66:4162–82.
7. Dickinson P, Looper KJ. Psychogenic nonepileptic seizures: a current overview. Epilepsia. 2012;53(10):1679–89.
8. Kerr WT, Zhang X, Hill CE, Janio EA, Chau AM, Braesch CT, et al. Factors associated with delay to video-EEG in dissociative seizures. Seizure. 2021;86:155–60.
9. Kerr WT, Janio EA, le JM, Hori JM, Patel AB, Gallardo NL, et al. Diagnostic delay in psychogenic seizures and the association with anti-seizure medication trials. Seizure. 2016;40:123–6.
10. Goldstein LH, Robinson EJ, Chalder T, Stone J, Reuber M, Medford N, et al. Moderators of cognitive behavioural therapy treatment effects and predictors of outcome in the CODES randomised controlled trial for adults with dissociative seizures. J Psychosom Res. 2022;158:110921.
11. Walczak TS, Papacostas S, Williams DT, Scheuer ML, Lebowitz N, Notarfrancesco A. Outcome after diagnosis of psychogenic nonepileptic seizures. Epilepsia. 1995;36(11):1131–7.
12. Lenio S, Kerr WT, Watson M, Baker S, Bush C, Rajic A, et al. Validation of a predictive calculator to distinguish between patients presenting with dissociative versus epileptic seizures. Epilepsy Behav. 2021;116:107767.
13. Wardrope A, Howell SJ, Reuber M. Diagnostic features of functional/dissociative seizures in the first presentation of transient loss of consciousness. Epilepsy Behav. 2025;164:110263.
14. Kerr WT, Janio EA, Chau AM, Braesch CT, Le JM, Hori JM, et al. Objective score from initial interview identifies patients with probable dissociative seizures. Epilepsy Behav. 2020;2020(113):107525.
15. LaFrance WC, Baker GA, Duncan R, Goldstein LH, Reuber M. Minimum requirements for the diagnosis of psychogenic nonepileptic seizures: a staged approach: a report from the International League Against Epilepsy Nonepileptic Seizures Task Force. Epilepsia. 2013;54(11):2005–18.
16. Wardrope A, Jamnadas-Khoda J, Broadhurst M, Grünewald RA, Heaton TJ, Howell SJ, et al. Machine learning as a diagnostic decision aid for patients with transient loss of consciousness. Neurol Clin Pract. 2020;10(2):96–105.
17. Chen M, Jamnadas-Khoda J, Broadhurst M, Wall M, Grünewald R, Howell SJL, et al. Value of witness observations in the differential diagnosis of transient loss of consciousness. Neurology. 2019;92:e895–e904.
18. Reuber M, Chen M, Jamnadas-Khoda J, Broadhurst M, Wall M, Grünewald RA, et al. Value of patient-reported symptoms in the diagnosis of transient loss of consciousness. Neurology. 2016;87(6):625–33.
19. Trainor D, Foster E, Rychkova M, Lloyd M, Leong M, Wang AD, et al. Development and validation of a screening questionnaire for psychogenic nonepileptic seizures. Epilepsy Behav. 2020;112:107482.
20. Muthusamy S, Seneviratne U, Ding C, Phan TG. Using semiology to classify epileptic seizures vs psychogenic nonepileptic seizures: a meta-analysis. Neurol Clin Pract. 2022;12(3):234–47.
21. Seneviratne U, Rajendran D, Brusco M, Phan TG. How good are we at diagnosing seizures based on semiology? Epilepsia. 2012;53:e63–e66.
22. Kerr WT, Zhang X, Janio EA, Karimi AH, Allas CH, Dubey I, et al. Reliability of additional reported seizure manifestations to identify dissociative seizures. Epilepsy Behav. 2021;115:107696.
23. Kerr WT, Chau AM, Janio EA, Braesch CT, le JM, Hori JM, et al. Reliability of reported peri-ictal behavior to identify psychogenic nonepileptic seizures. Seizure. 2019;67:45–51.
24. Kerr WT, Janio EA, Braesch CT, le JM, Hori JM, Patel AB, et al. An objective score to identify psychogenic seizures based on age of onset and history. Epilepsy Behav. 2018;80:75–83.
25. Kerr WT, Janio EA, Braesch CT, le JM, Hori JM, Patel AB, et al. Identifying psychogenic seizures through comorbidities and medication history. Epilepsia. 2017;58(11):1852–60.
26. Sobregrau P, Baillès E, Radua J, Carreño M, Donaire A, Setoain X, et al. Design and validation of a diagnostic suspicion checklist to differentiate epileptic from psychogenic nonepileptic seizures (PNES-DSC). J Psychosom Res. 2024;180:111656.
27. Kerr WT, McFarlane KN, Pucci GF, Carns DR, Israel A, Vighetti L, et al. Supervised machine learning compared to large language models for identifying functional seizures from medical records. Epilepsia. 2025;66(4):1155–64.
28. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. N Engl J Med. 2023;388(21):1981–90.
29. Abdulnour RE, Gin B, Boscardin CK. Educational strategies for clinical supervision of artificial intelligence use. N Engl J Med. 2025;393(8):786–97.
30. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang. 2008;59(4):390–412.
31. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48.
32. Hunt L, Jorgensen M. Clustering mixed data. Wiley Interdiscip Rev Data Min Knowl Discov. 2011;1(4):352–61.

33. Hanrahan B, Ghearing G, Urban A, Plummer C, Pan J, Hendrickson R, et al. Diagnostic accuracy of paroxysmal spells: clinical history versus observation. Epilepsy Behav. 2018;78:73–7.
34. Wasserman D, Herskovitz M. Epileptic vs psychogenic nonepileptic seizures: a video-based survey. Epilepsy Behav. 2017;73:42–5.
35. Battaglia G, Sofia V, la Delfa G, Cantalupo G, Didato G, Mostacci B, et al. Distinguishing epileptic seizures from imitators: a video-based evaluation of non-epileptologist physicians. Epilepsy Behav. 2025;166:110369.
36. Andrienko N, Andrienko G, Artikis A, Mantenoglou P, Rinzivillo S. Human-in-the-loop: visual analytics for building models recognising behavioural patterns in time series. IEEE Comput Graph Appl. 2024;44(3):14–29.
37. Kerr WT, McFarlane KN. Machine learning and artificial intelligence applications to epilepsy: a review for the practicing epileptologist. Curr Neurol Neurosci Rep. 2023;23(12):869–79.
38. Angus DC, Khera R, Lieu T, Liu V, Ahmad FS, Anderson B, et al. AI, health, and health care today and tomorrow: the JAMA summit report on artificial intelligence. JAMA. 2025;334(18):1650–64.
39. Brehaut JC, Graham ID, Wood TJ, Taljaard M, Eagles D, Lott A, et al. Measuring acceptability of clinical decision rules: validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. Med Decis Mak. 2010;30(3):398–408.
40. Park Y, Hu J, Singh M, Sylla I, Dankwa-Mullan I, Koski E, et al. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. JAMA Netw Open. 2021;4(4):e213909.
41. Wissel BD, Greiner HM, Glauser TA, Mangano FT, Santel D, Pestian JP, et al. Investigation of bias in an epilepsy machine learning algorithm trained on physician notes. Epilepsia. 2019;60(9):e93–e98.
42. Huang J, Galal G, Etemadi M, Vaidyanathan M. Evaluation and mitigation of racial bias in clinical machine learning models: scoping review. JMIR Med Inform. 2022;10(5):e36388.
43. Robson C, Lian OS. "Blaming, shaming, humiliation": stigmatising medical interactions among people with non-epileptic seizures. Wellcome Open Res. 2017;2:55.
44. McLoughlin C, Lee WH, Carson A, Stone J. Iatrogenic harm in functional neurological disorder. Brain. 2025;148(1):27–38.
45. McLoughlin C, McWhirter L, Pisegna K, Tijssen MAJ, Tak LM, Carson A, et al. Stigma in functional neurological disorder (FND)—a systematic review. Clin Psychol Rev. 2024;112:102460.
46. Lesser RP. Psychogenic seizures. Neurology. 1996;46(6):1499–507.
47. Gumnit RJ, Walczak TS, National Association of Epilepsy C. Guidelines for essential services, personnel, and facilities in specialized epilepsy centers in the United States. Epilepsia. 2001;42(6):804–14.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kerr WT, McFarlane KN, Allas CH, Terman SW, Reuber M, Seo SH, et al. Quantifying the impact of a computer-aided diagnostic score on the clinical diagnosis of functional seizures. Epilepsia. 2026;00:1–12. https://doi.org/10.1002/epi.70069