Received: 10 August 2022 | Accepted: 5 October 2022

DOI: 10.1002/epd2.20010

Epileptic Disorders

ORIGINAL ARTICLE

The importance of discriminative power rather than significance when evaluating potential clinical biomarkers in epilepsy research

Geertruida Slinger | **Remi Stevelink** | **Eric van Diessen** | **Kees P. J. Braun** | **Willem M. Otte**

Department of Child Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Correspondence
Geertruida Slinger, Department of Child Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Room KG 01.310.0, P.O. Box 85090, Utrecht 3508 AB, The Netherlands.
Email: g.slinger-2@umcutrecht.nl

Funding information
Clinical research fellowship from the UMC Utrecht; Research fellowship from the Brain Center Rudolf Magnus (current name: UMC Utrecht Brain Center); The Friends UMC Utrecht/MING Fund

Abstract

Objective: The quest for epilepsy biomarkers is on the rise. Variables with statistically significant group-level differences are often misinterpreted as biomarkers with sufficient discriminative power. This study aimed to demonstrate the relationship between significant group-level differences and a variable's power to discriminate between individuals.

Methods: We simulated normal-distributed datasets from hypothetical populations with varying sample sizes (25–800), effect sizes (Cohen's *d*: .25–2.50), and variability (standard deviation: 10–35) to assess the impact of these parameters on significance and discriminative power. The simulation data were illustrated by assessing the discriminative power of a potential real-case biomarker—the EEG beta band power—to diagnose generalized epilepsy, using data from 66 children with generalized epilepsy and 385 controls. Additionally, we evaluated recently reported epilepsy biomarkers by comparing their effect sizes to our simulation-derived effect size criterion.

Results: Group size affects significance but not discriminative power. Discriminative power is much more related to variability and effect size. Our real data example supported these simulation results by demonstrating that group-level significance does not translate, one to one, into discriminative power. Although we found a significant difference in the beta band power between children with and without epilepsy, the discriminative power was poor due to a small effect size. A Cohen's *d* of at least 1.25 is required to reach good discriminative power in univariable prediction modeling. Slightly over 60% of the biomarkers in our literature search met this criterion.

Significance: Rather than statistical significance of group-level differences, effect size should be used as an indicator of a variable's biomarker potential. The minimal required effects size for individual biomarkers—a Cohen's *d* of 1.25—is large. This calls for multivariable approaches, in which combining multiple variables with smaller effect sizes could increase the overall effect size and discriminative power.

UMC Utrecht Brain Center: Member of ERN EpiCARE.

KEYWORDS
area under the curve, precision medicine, receiver operating curve, sensitivity, specificity

1 | INTRODUCTION

Proper diagnosis of epilepsy and prediction of its course are essential for effective management and patient counseling. Diagnostic and prognostic studies, therefore, aim to identify markers that can support these processes.[1] Estimates of these markers are conventionally reported at the group level and used to infer population-wide conclusions. The emerging field of personalized medicine, however, targets the search for individualized markers—so-called biomarkers[2]—rather than group-averaged differences.

Biomarkers may have different applications but are generally defined as "characteristics that are measured as indicators of physiological processes, pathogenic processes, or biological responses to an exposure or intervention, including therapeutic interventions".[3] In the epilepsy field, quite some progress has been made in the discovery of both diagnostic and prognostic biomarkers.[4] Besides epileptiform EEG activity as an important, but not entirely accurate, biomarker used in the diagnosis of epilepsy,[5,6] various microRNAs,[7–9] proteins,[10,11] metabolites,[12] and immune system components[13,14] have been proposed as biomarkers for the diagnosis of seizures and epilepsy subtypes. MRI markers, mainly network connectivity measures[15,16] and specific EEG (network) features[17–20] have diagnostic value as well and seem to be able to predict epilepsy severity and refractoriness. Genetic markers, including polygenic risk scores (PRS),[21] are also receiving growing interest as biomarkers of epileptogenesis and epilepsy risk, diagnosis, and prognosis (for overviews, see: Refs 22,23).

Despite the large expansion in publications reporting potential biomarkers in the epilepsy field—and beyond—very few biomarkers have yet found their way to clinical practice. Specific disease-related challenges, such as seizures near the time of biomarker sampling and the use of anti-seizure medications (ASMs), might play a role in this, as they can have a direct effect on biomarker levels (e.g., miRNA)[24] and thus distort measurements. Another important contributing factor is that statistical significance is still often incorrectly interpreted as the power to identify personal traits, resulting in low discriminative power.[25] Group-level significance is based on average rather than individual differences, whereas the performance of a biomarker is defined by both the difference and the variation between individuals. Therefore, new markers should be evaluated on their discriminative performance rather than on statistically significant group differences before labelling them as biomarkers.[26]

Key Points

- The quest for epilepsy biomarkers is on the rise as they play an important role in the evolution of precision medicine.
- Statistical significance of mean group differences is a poor indicator of the individual-level discriminative power of a variable.
- Rather than significance, effect size has a strong relationship with discriminative power.
- Individual variables need extremely large effect sizes to have good discriminative power; we should, therefore, focus on combining variables.
- Widespread methodological knowledge of biomarker evaluation among researchers and clinicians might contribute to the further evolution of precision medicine.

This study aimed to reinforce the idea that caution is needed in identifying and reporting differences in markers at the group level as biomarkers that can discriminate between subjects at the individual level. We do not claim that this idea is new,[25] but merely wish to illustrate the importance of this concept for the interpretation and potential usage of epilepsy biomarkers. Therefore, we first assessed the relationship between significant group-level differences and a marker's power to discriminate between individuals in a simulated setting. Secondly, we provide a real data example evaluating a potential biomarker—EEG beta power oscillations—for diagnosing generalized epilepsy in children previously, as published in *Epilepsia*.[27] Thirdly, we evaluated the recent scientific literature on new potential biomarkers in epilepsy.

2 | MATERIALS AND METHODS

2.1 | Data simulation

Biomarkers may quantify different types of outcomes: binary, continuous, or correlational (Figure 1). The logarithmic (log) transformed odds ratio from binary data, as well as Pearson's correlation coefficient (r) from correlational data, is convertible into the same standardized mean difference (SMD) obtained from continuous

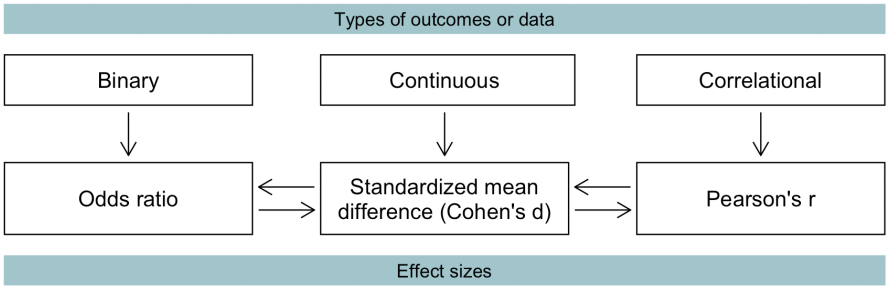

FIGURE 1 Relationship between different effect sizes. Cohen's *d* is the natural logarithm of the odds ratio multiplied by the constant .551 (i.e., $\sqrt{3}/\pi$), and Cohen's *d* is the correlation *r* multiplied by two, divided by $\sqrt{(1-r^2)}$.[28]

data. The SMD, also named Cohen's *d*, is one of the most commonly used effect size measures, indicating how many standard deviations (SDs) two group means differ.[28] Hence, we restricted our simulations to Cohen's *d* outcome.

We performed two different sets of simulations. The first set was used to evaluate the impact of both sample size and effect size on discriminative power. We generated multiple normally distributed datasets sampled from hypothetical populations, reflecting outcomes in a patient and control group (control mean: 100; standard deviation: 15), with varying sample sizes (25–800) and effect sizes (.25–2.50). Equally sized datasets with similar parameters were generated and used as independent validation data to evaluate the discriminative power. The second set of simulations, used to evaluate the relationship between variability and discriminative power, was based on similarly generated datasets but with a fixed mean of the control and patient group (100 and 115, respectively), a fixed sample size (400 subjects per group), and varying SDs (10–35). The simulation code is available via the Zenodo platform (DOI: 10.5281/zenodo.7095386).[29]

We evaluated the group-level differences in outcome between the patient and control groups using the *Z* value. The discriminative power of the simulated biomarkers was evaluated using receiver operating characteristic (ROC) curve analysis[30] (Box 1, Figure 2).

2.2 | Example: EEG beta band power in generalized epilepsy

To illustrate the simulations with real research data, we evaluated a potential epilepsy biomarker derived from one of our group's recently published research projects.[27] This project demonstrated a significant genetic relationship between generalized epilepsy (GE) and background beta power oscillations on resting-state EEG. As this points to a shared biological mechanism underlying background EEG beta band oscillations and the susceptibility for or development of generalized seizures, we hypothesized that altered background beta power oscillations might indicate a prodromal state or be a feature of GE and thus could potentially be a diagnostic GE biomarker.[27] Therefore, we investigated the difference in beta power oscillations in children with GE compared to children without epilepsy, and the accuracy of the beta power oscillations for diagnosing GE.

BOX 1 *Z* value and ROC analysis

Evaluation of statistical significance

The *Z* value can be used to characterize the difference between two groups. With a significance level of .05 and the corresponding 95% confidence interval (CI), *Z* values smaller than −1.96 and greater than1.96 will yield significant results.[31]

Evaluation of predictive power

The receiver operating characteristic (ROC) curve analysis can be used to evaluate discriminative power.[30] An ROC curve is a graphical representation of the true-positive rate (TPR; i.e., sensitivity) against the false-positive rate (FPR; i.e., 1-specificity). The TPR represents the proportion of study subjects correctly classified as patients (TP) out of the total number of patients (TP + FN). Similarly, the FPR is the proportion of subjects incorrectly classified as patients (FP) out of all control subjects (TN + FP). The TPR and FPR can be calculated for every possible threshold value of a biomarker or test. An ROC curve is generated by plotting the TPR and FPR across varying thresholds (Figure 2).[32] The area under the curve (AUC) summarizes the area underneath the entire ROC curve across all possible thresholds and thus provides an overall and combined measure of sensitivity and specificity. The AUC ranges between 0 and 1, where .5 indicates that the model does not perform better than chance. The closer the AUC comes to 1.0, the better a biomarker can discriminate between patients and controls.[30,32]

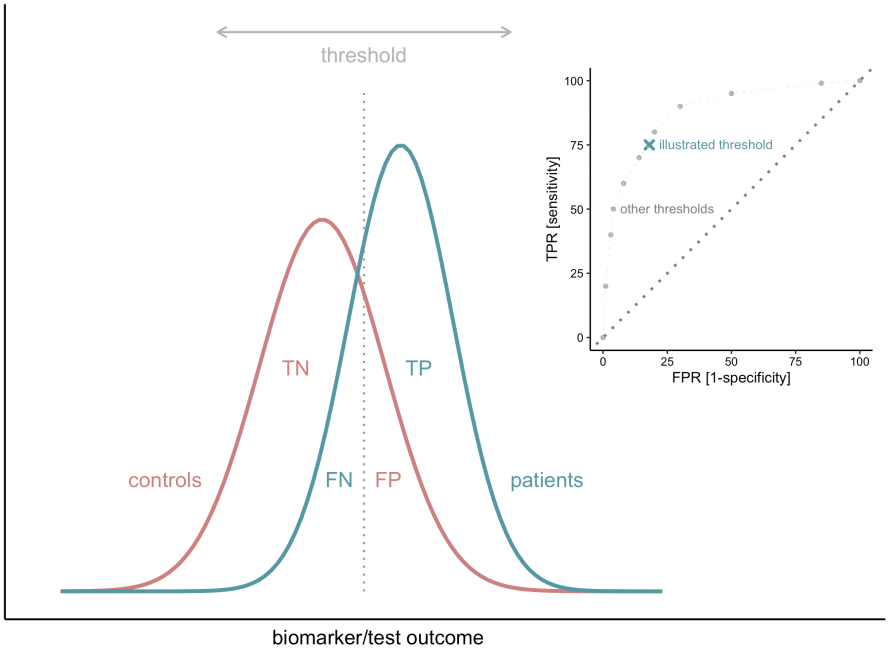

FIGURE 2 Subject classifications at various thresholds constitute the ROC curve. In this illustration, biomarker or test outcomes above the threshold classify subjects as patients (green), while outcomes below the threshold classify subjects as controls (red). Each threshold yields a true-positive, false-negative, true-negative, and false-positive fraction. The true-positive rate (TP/[TP + FN]) and false-positive rate (FP/[FP + TN]) at each threshold constitute the ROC curve. FN, false negatives; FP, false positives; FPR, false-positive rate; TN, true negatives; TP, true positives; TPR, true-positive rate.

2.2.1 | Study cohort and EEG data collection

We retrospectively reviewed children (0–18 years) referred to the outpatient First Seizure Clinic (FSC) of the Wilhelmina Children's Hospital, Utrecht, the Netherlands, between January 2008 and May 2018, after a suspected first seizure. Diagnoses were made, directly after the FSC evaluation or after additional investigations if needed, by an experienced pediatric neurologist according to the epilepsy definition and classification of the International League Against Epilepsy (ILAE).[33,34] For the analyses, we only included children diagnosed with GE (cases) and children in whom the epilepsy diagnosis was discarded (controls).

For each child, we collected demographic data, including sex, age at EEG recording, seizure history (defined as a history of febrile seizures, neonatal convulsions, or acute symptomatic seizures), and other neurological history (defined as a history of or the presence of perinatal asphyxia, congenital or acquired brain lesions, head trauma, central nervous system infections, or migraine). We also collected the raw data of the first routinely performed EEG. All EEGs were recorded with at least 21 scalp electrodes, arranged according to the international 10–20 system (SystemPLUS Evolution; Micromed). Sampling frequency varied between 256 and 2048 Hz. EEGs were exported as raw TRC files and, in case of a higher sampling frequency, down-sampled to 256 Hz. From each EEG, we selected one 15-s eyes-closed resting-state epoch without epileptiform discharges, nonspecific abnormalities, or artifacts. Epoch data were filtered into the beta (13–30 Hz) band. We computed the beta band signal power for the vertex electrode (Cz) according to Smit et al.[35] and Stevelink et al.[27]

The institutional review board approved using the retrospective data for research purposes without explicit informed consent (project numbers 09-353/K and 18-354/C).

2.2.2 | Statistical analysis

We applied a logarithmic transformation to the Cz signal beta power data, referring to it as simply the Cz power. Group differences in the Cz power were assessed using the Mann–Whitney–Wilcoxon test for independent samples. We fitted both univariable and multivariable logistic regression models to evaluate the ability of the Cz power to discriminate between children with GE and children without epilepsy. The multivariable model included the Cz power and four demographic variables—sex, neurological history, age at EEG, and seizure history—of which the first two are known predictors of the diagnosis of pediatric epilepsy.[36] Model predictions were normalized. ROC curve analysis (Box 1, Figure 2) was used for model evaluation. Because of the solely explanatory nature of the data example, we did not perform internal and external model validation.

(A)

Group size	Z value (95% CI)	Significance
25	1.1 (CI: 0.1–2.4)	
50	1.3 (CI: 0.1–3.0)	
100	1.8 (CI: 0.1–3.7)	
200	2.5 (CI: 0.6–4.7)	*
400	3.6 (CI: 1.8–5.4)	***
800	5.0 (CI: 3.1–6.3)	*****

(B)

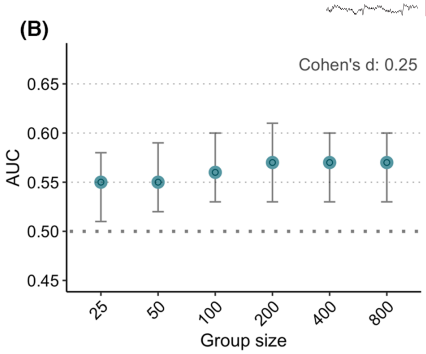

FIGURE 3 Simulated effect of group size on the significance and discriminative power for an effect size (Cohen's *d*) of .25. (A) Relation between group size, *Z* value, and significance. (B) Relation between group size and discriminative power in terms of AUC. The AUC is given as the mean with 95% CI. AUC, area under the curve; CI, 95% confidence interval; *$p < .05$; ***$p \leq .001$; and *****$p \leq .00001$.

All analyses were performed in R statistical software[37] using the psd package (version 2.1.0)[38] and pROC package (version 1.18.0).[39]

2.3 | Effect sizes reported in the literature

To get a sense of biomarker effect sizes obtained in the epilepsy field, we searched PubMed for recent (2019–2021) publications reporting novel individual biomarkers across all areas of epilepsy research. We compared the biomarkers' effect sizes to our simulation-derived effect size criterion. We used the following query: "epilepsy[tiab] AND biomarker[tiab] AND Humans[MESH] AND English[LANG] NOT (meta-analysis[tiab] OR review[tiab]) AND 2019/01/01:2021/12/31[DP]" and only included entries with an abstract and a free full-text document (i.e., open access). We extracted the biomarker type, sample size, effect size, and AUC data from each publication. We calculated the effect size in case it was not explicitly given. From publications reporting multiple biomarkers or one biomarker for various subgroups, we extracted the largest effect size. All effect sizes were converted to Cohen's *d* and absolutized. We characterized the distribution of both the reported effect sizes and sample sizes.

3 | RESULTS

3.1 | Data simulation

3.1.1 | Sample size

Sample size has a strong relationship with group-level significance, as an increase in sample size directly increases the *Z* value. Therefore, large sample sizes give significant differences at the group level, even with a small effect size of Cohen's *d* .25 (Figure 3A). By contrast, sample size hardly has any effect on discriminative power, expressed as the AUC. In the case of an effect size of Cohen's *d* .25, the AUC was .55 (95% CI: .51–.58) with 25 subjects per group and increased to .57 (95% CI: .53–.61) with 200 subjects per group (Figure 3B, Table S1).

3.1.2 | Effect size

In contrast to the sample size, the effect size strongly affects the AUC. The greater the effect size of a biomarker, the better its discriminative power (Figure 4). A single, normally distributed biomarker requires a Cohen's *d* of at least 1.25 to reach an AUC of .80 (Figure 4C), considered as the lower limit of good discrimination.[32,40] Biomarkers with both a sensitivity and specificity of .8 or 80%—instead of combined in the AUC—require an even greater effect size, namely a Cohen's *d* of 1.66 (Figure S1). Effect sizes of 1.25 and 1.66 correspond to odds ratios (OR) of 9.65 (Table S2) and 20.31, respectively. The combined impact of sample size and effect size on the AUC is shown in Figure S2.

3.1.3 | Data variability

The AUC is also highly dependent on data variability, expressed as SD. Since Cohen's *d* is calculated by subtracting the means of the control group and patient group, divided by the pooled SD, an increase of the SD directly leads to a decrease in Cohen's *d*. This translates, as shown in Figure 4, into a decreased AUC (Figure 5).

3.2 | Example: EEG beta band power in generalized epilepsy

A total of 587 children (54.9% boys), with a mean age of 8.8 ± 4.2 years at the first EEG recording, were evaluated

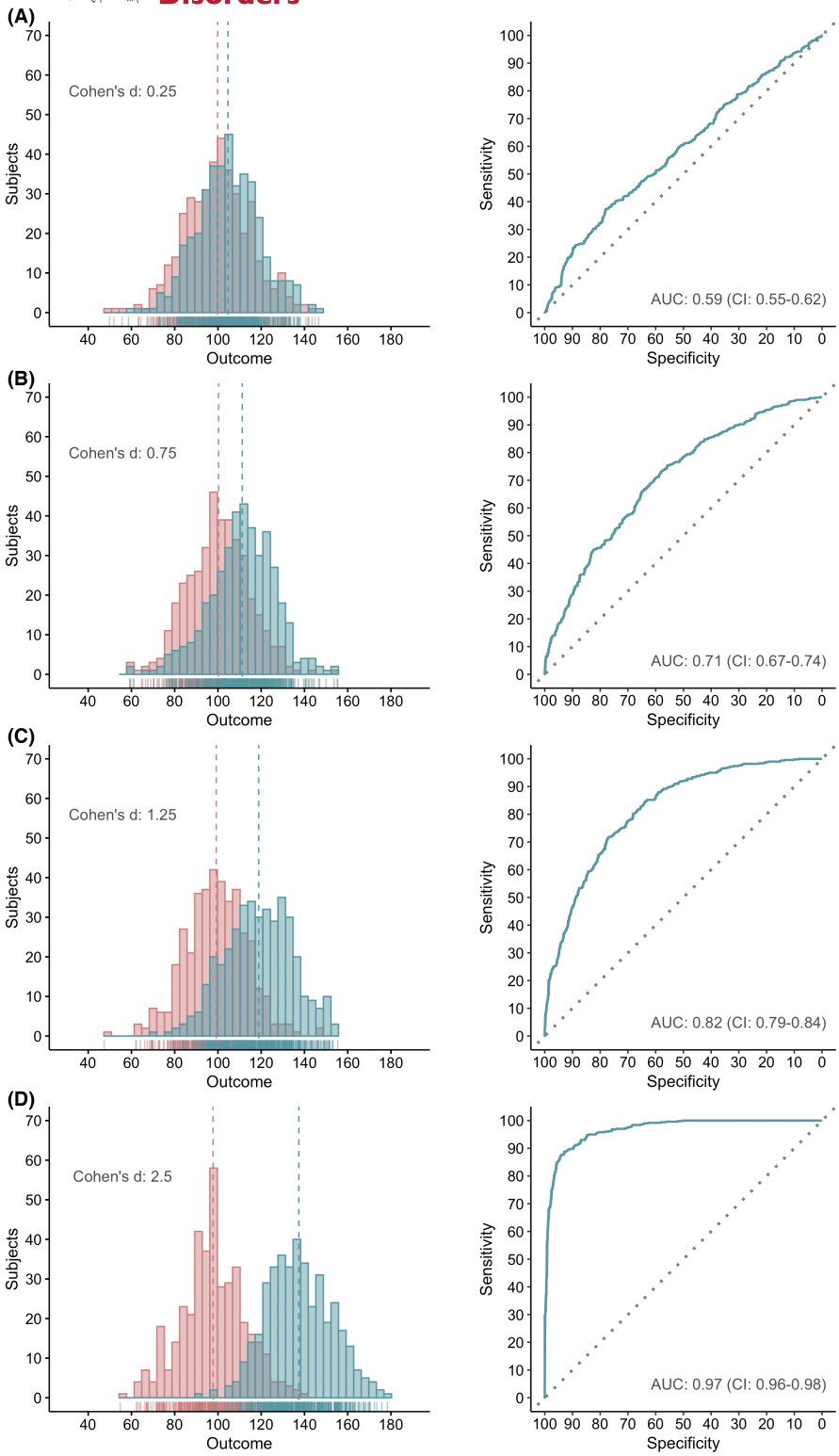

FIGURE 4 Simulated outcome (left) and corresponding receiver operating curves (right) for four different effect sizes (Cohen's *d*) and a fixed group size (400 subjects per group). Left plots: data are presented as histograms with individual data below the histogram as density plots; dotted lines represent group means. Red: control group; green: patient group. AUC, area under the curve; CI, 95% confidence interval.

at the FSC. Previous seizures (other than in the context of epilepsy) occurred in 97 children (16.5%), and 110 children (18.7%) had other neurological problems in their medical history. Of the 587 children, 136 (23.2%) were diagnosed with focal epilepsy and 66 (11.2%) with generalized epilepsy. The remaining 385 (65.6%) children did not have epilepsy and served as controls. For analysis, we only included the controls and children with generalized epilepsy.

The beta band Cz power was significantly higher in children with generalized epilepsy than in children without epilepsy, with a median difference of −.075 (CI: −.14 to −.0064; $p = .034$) (Figure 6A). However, univariable model predictions, with the Cz power as the sole

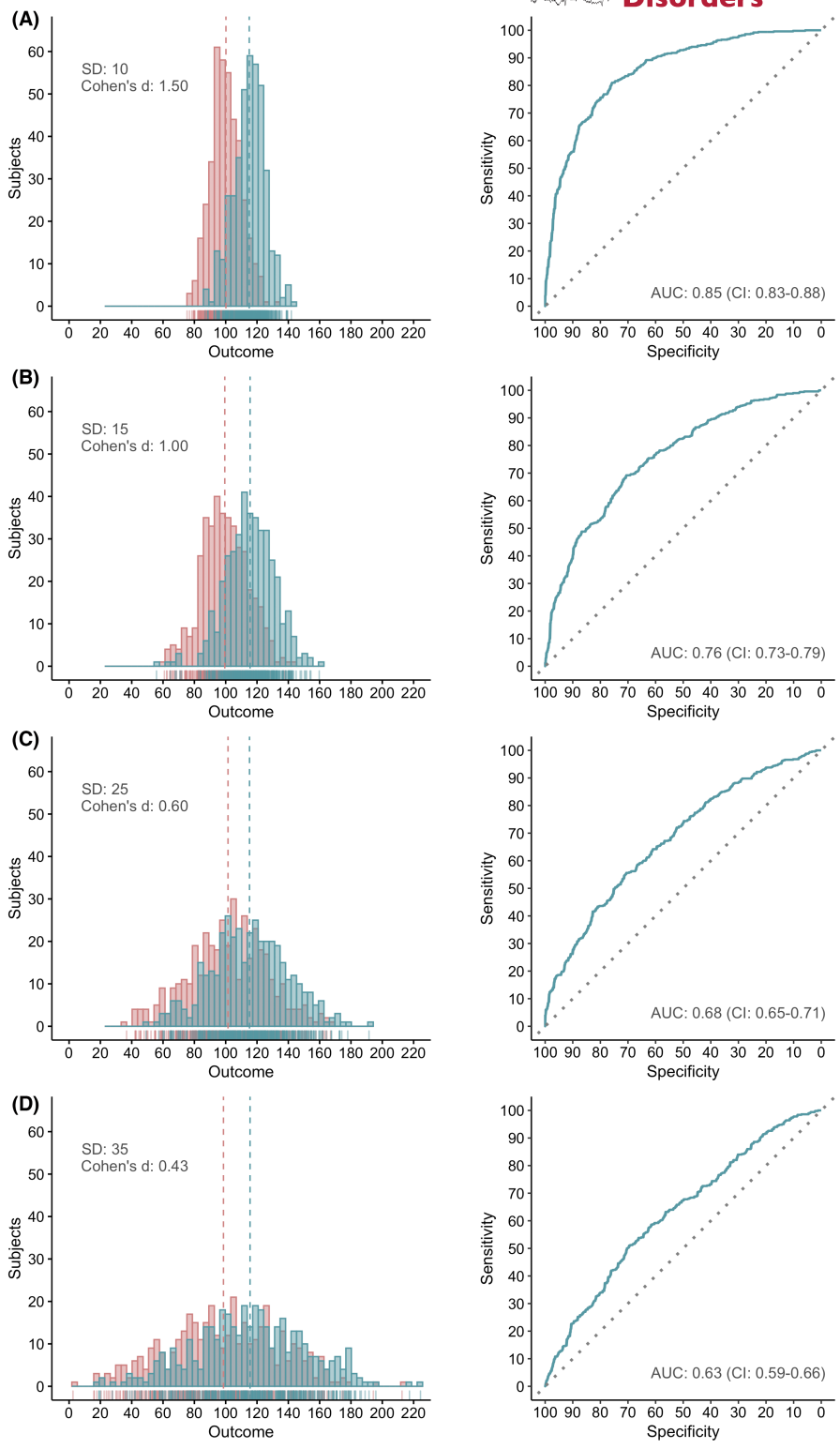

FIGURE 5 Simulated outcome (left) and corresponding receiver operating curves (right) for four different standard deviations and a fixed group size (400 subjects per group). Left plots: data are presented as histograms with individual data below the histogram as density plots; dotted lines represent group means. Effect sizes vary due to varying SDs. Red: control group; green: patient group. AUC, area under the curve; CI, 95% confidence interval.

predictor, for the presence of generalized epilepsy had poor discriminative power, with an AUC of .58 (CI: .51–.65) (Figure 6B,C). Adding the variables, sex, seizure history, age at EEG recording, and neurological history to the model, improved the overall performance to an AUC of .67 (CI: .60–.74) (Figure 6C,D).

3.3 | Effect sizes reported in the literature

The PubMed search yielded 123 results on April 08, 2022. A total of 91 publications were excluded for the following reasons: no report of a new single biomarker

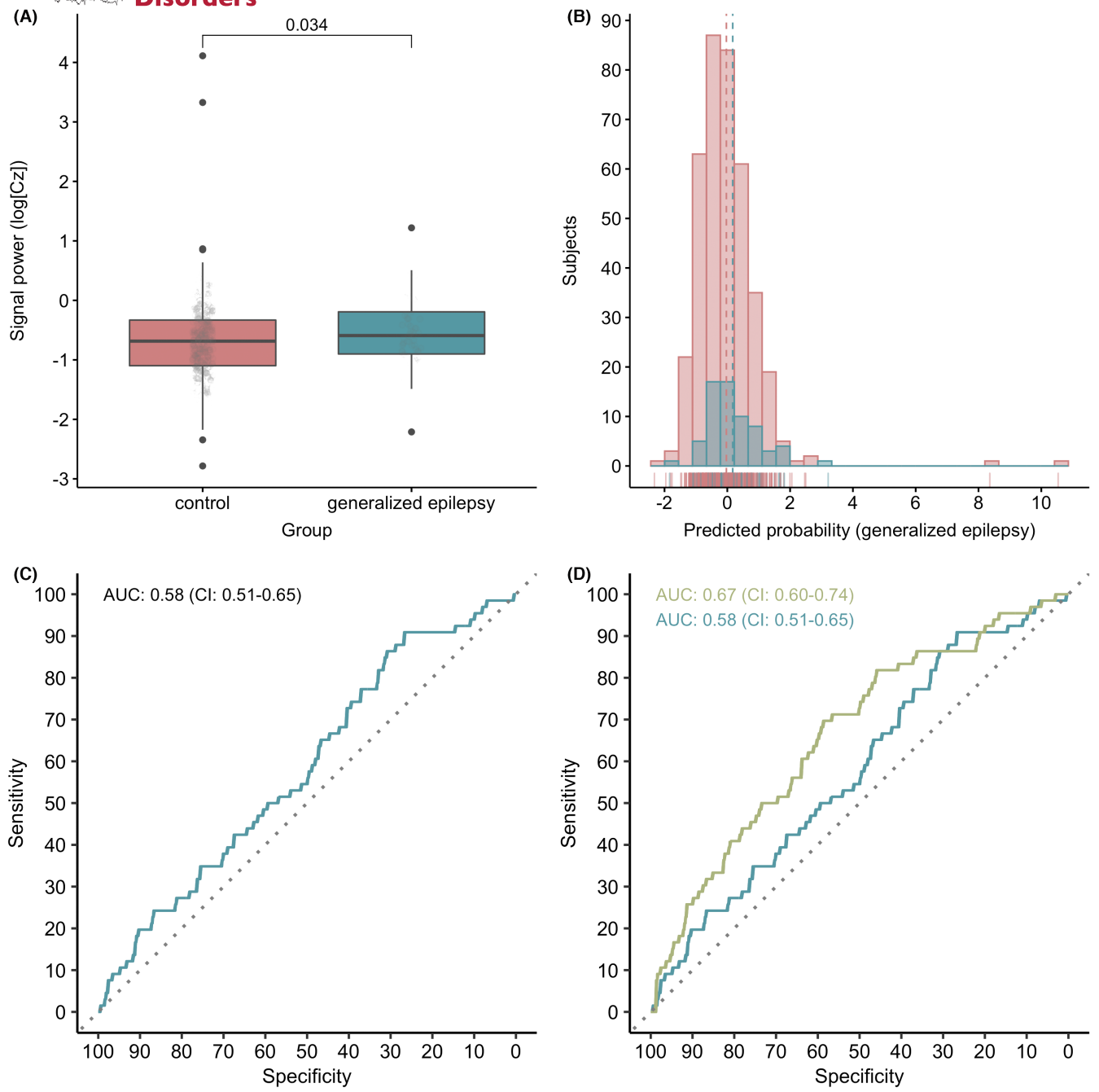

FIGURE 6 Significance versus discriminative power of Cz signal power. (A) Group-level difference in log(Cz) power between children with generalized epilepsy (green) and children without epilepsy (red), as assessed using the Mann–Whitney–Wilcoxon test. Boxes represent medians, 25th percentiles, and 75th percentiles. (B) Normalized predictions of the univariable logistic regression model for epilepsy in children with generalized epilepsy (green) and without epilepsy (red) are presented as histograms with individual data below the histograms as density plots. Dotted lines represent means. (C) The ROC curve of the univariable logistic regression model with an AUC of .58 (CI: .51–.65). (D) ROC curves of the univariable regression model (green) and multivariable regression model (light green); AUC: .67 (CI: .60–.74). AUC, area under the curve; CI, 95% confidence interval.

(n = 25), reported data not useful for analysis (n = 24), research not primarily focused on epilepsy (n = 17), no report of human data (n = 12), no original research (n = 10), case studies (n = 3). Of the 32 included publications, 13 reported a neurophysiological biomarker, while nine reported a molecular and eight an imaging biomarker. The two remaining studies presented a genetic biomarker and a biomarker which did not fit in one of the categories mentioned above. The references and categorization of the publications can be found in Appendix S3. In all publications, the described findings were explicitly labeled as "biomarkers," though in most

cases with caution, expressed with terms such as "potential" and "may be."

Of all included publications, 62.5% reported effect sizes for their biomarkers equal to or greater than the proposed minimal Cohen's *d* of 1.25 (Figure S3A). Since we only extracted the largest effect size from each publication, these numbers represent the best-case scenario. A total of 12 publications reported AUC data for (one of) the individual biomarkers presented (Figure 3B). Sample sizes were generally smaller than 50 subjects per subgroup (i.e., control or patient group) (Figure 3C).

4 | DISCUSSION

This study demonstrates the discrepancy between group-level significance and the individual-level discriminative power of potential biomarkers. In contrast to what is often implied when significant group differences are put into perspective, the statistical significance of mean group differences is a very poor indicator of the utility of a variable as an individual biomarker. Rather than significance, the effect size directly impacts discriminative power and is thus more suitable for evaluating a variable's biomarker potential.

Are group-level differences not important at all? The importance of one (group-level) or the other (individual-level) outcome depends on the research question and the stage of the research. In fact, most biomarker research starts with the search for group-level differences, as this is the basis for estimating the effect size. Therefore, also sample size matters. Although sample size does, as shown, not affect discriminative power, effect size estimates can only be precise and accurate with an appropriate sample size.[41] Both underestimates and overestimates of the effect size can result in a distorted view of a variable's potential to be a biomarker.

Based on the required effect size, many individual variables will probably not qualify as good biomarkers. Moderate to even large effect sizes, traditionally defined as Cohen's *d*'s of .5–1.0,[42] do not translate into sufficient discriminative power if tested in isolation. Our results show that an effect size of 1.25, corresponding to an OR of almost 10, is needed for a variable to reach an AUC value of at least .8. Biomarkers with both a sensitivity and specificity of .8 require an even larger Cohen's *d* of 1.66, which corresponds to an OR greater than 20. We found a relatively large number of biomarkers with a large effect size in our concise literature search. This number might be inflated as we only reviewed publications explicitly reporting a "biomarker" in the title or abstract. Nonetheless, 37.5% of the studies reported a biomarker with an effect size smaller than a Cohen's *d* of 1.25. The minimal required effect size for individual biomarkers promotes the use of multivariable approaches. As illustrated in our data example, combining multiple (clinical) variables with smaller effect sizes could increase the overall effect size and discriminative power. Multivariable approaches also better suit the complexity of the pathophysiology of epilepsy.

Our study has limitations. We only simulated normally distributed data, whereas typical epilepsy outcomes are often distributed according to more complex patterns. Moreover, our simulated training (i.e., the original) and validation datasets had the same distribution parameters, while in a real-case scenario, those datasets are likely to differ—at least to some extent—in distribution as they are collected from independent populations. Hence, our simulation results might be too optimistic. Secondly, we collected our example data retrospectively, which is generally regarded as a suboptimal approach for evaluating a diagnostic biomarker.[30] Nevertheless, we best approximated a prospective study by sampling patients and controls from a suspected, not yet diagnosed population. Additionally, to keep our example as clear as possible, we quantified the discriminative ability of the EEG data models using the same data from which the models were developed.[43] Model validation on independent data most likely would have yielded even worse discriminative power, strengthening our general message to be careful with *p* values and focus on sensitivity, specificity, and the AUC. Lastly, besides traditional ROC analysis, other biomarker evaluation methods exist,[44,45] particularly for assessing a biomarker's added value, clinical utility, or healthcare impact. Although we did not cover these methods here, we are aware that discovering biomarkers with sufficient discriminative power is the first rather than the last step of the biomarker evaluation process, as discriminative power does not necessarily translate into added value in clinical practice or improved outcomes for the patient or healthcare system.[46]

We do not intend to present a pessimistic view of biomarker discovery efforts, as we believe biomarkers will have an invaluable role in personalized medicine in epilepsy.[47] Instead, we aim to promote the use of appropriate biomarker selection methods and increase methodological knowledge. This study might aid with translating published effect sizes into hypothetical AUC values. In line with this, we recommend researchers to always report effect sizes instead of only *p* values.[48] Moreover, we call for the objective reporting of study results without unsupported or unjustified claims on biomarker potential. Growing knowledge and awareness of the methodology of biomarker research on both the authors' and interpreters' side will help move the field

in the right direction and contribute to the further evolution of precision medicine.

ACKNOWLEDGMENTS

GS was supported by the Friends UMC Utrecht/MING Fund and a Research Fellowship from the Brain Center Rudolf Magnus (current name: UMC Utrecht Brain Center). RS was supported by the Friends UMC Utrecht/MING Fund. EvD was supported by a Clinical Research Fellowship from the UMC Utrecht.

CONFLICT OF INTEREST STATEMENT

None of the authors have any conflicts of interest to disclose. We confirm that we have read the journal's position on issues involved in ethical publication and that this report is consistent with those guidelines.

DATA AVAILABILITY STATEMENT

Simulation data scripts are publicly available via the Zenodo platform: https://zenodo.org/record/7095386#.ZBM5iS2iFpQ. Deidentified data may be obtained from a third party and are not publicly available. Request via the corresponding author.

ORCID

Geertruida Slinger https://orcid.org/0000-0002-3656-8795
Remi Stevelink https://orcid.org/0000-0003-0214-7965
Eric van Diessen https://orcid.org/0000-0002-7773-1990
Willem M. Otte https://orcid.org/0000-0003-1511-6834

REFERENCES

1. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research. 2nd ed. Burlington, MA: Jones and Bartlett Publishers, Inc.; 2014.
2. Califf RM. Biomarker definitions and their applications. Exp Biol Med. 2018;243(3):213–21.
3. FDA-NIH Biomarker Working Group. BEST (biomarkers, EndpointS, and other tools) resource [internet]. Silver Spring, MD: Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US); 2016.
4. Pitkänen A, Ekolle Ndode-Ekane X, Lapinlampi N, Puhakka N. Epilepsy biomarkers – toward etiology and pathology specificity. Neurobiol Dis. 2019;123:42–58.
5. Engel J, Bragin A, Staba R. Nonictal EEG biomarkers for diagnosis and treatment. Epilepsia Open. 2018;3(S2):120–6.
6. Smith SJM. EEG in the diagnosis, classification, and management of patients with epilepsy. J Neurol Neurosurg Psychiatry. 2005;76(suppl 2):ii2–7.
7. An N, Zhao W, Liu Y, Yang X, Chen P. Elevated serum miR-106b and miR-146a in patients with focal and generalized epilepsy. Epilepsy Res. 2016;127:311–6.
8. Raoof R, Jimenez-Mateos EM, Bauer S, Tackenberg B, Rosenow F, Lang J, et al. Cerebrospinal fluid microRNAs are potential biomarkers of temporal lobe epilepsy and status epilepticus. Sci Rep. 2017;7(1):3328.
9. Avansini SH, de Sousa Lima BP, Secolin R, Santos ML, Coan AC, Vieira AS, et al. MicroRNA hsa-miR-134 is a circulating biomarker for mesial temporal lobe epilepsy. PLoS One. 2017;12(4):e0173060.
10. Wang R, Zeng GQ, Liu X, Tong RZ, Zhou D, Hong Z. Evaluation of serum matrix metalloproteinase-3 as a biomarker for diagnosis of epilepsy. J Neurol Sci. 2016;367(2):91–7.
11. Kalkan A, Demirel A, Atiş ŞE, Karaaslan EB, Ferhatlar ME, Senturk M. A new biomarker in the differential diagnosis of epileptic seizure: neurogranin. Am J Emerg Med. 2022;54:147–50.
12. Magnusson C, Herlitz J, Höglind R, Wennberg P, Edelvik Tranberg A, Axelsson C, et al. Prehospital lactate levels in blood as a seizure biomarker: a multi-center observational study. Epilepsia. 2021;62(2):408–15.
13. Pollard JR, Eidelman O, Mueller GP, Dalgard CL, Crino PB, Anderson CT, et al. The TARC/sICAM5 ratio in patient plasma is a candidate biomarker for drug resistant epilepsy. Front Neurol. 2013;3:3.
14. Clarkson BDS, LaFrance-Corey RG, Kahoud RJ, Farias-Moeller R, Payne ET, Howe CL. Functional deficiency in endogenous interleukin-1 receptor antagonist in patients with febrile infection-related epilepsy syndrome. Ann Neurol. 2019;85(4):526–37.
15. Pressl C, Brandner P, Schaffelhofer S, Blackmon K, Dugan P, Holmes M, et al. Resting state functional connectivity patterns associated with pharmacological treatment resistance in temporal lobe epilepsy. Epilepsy Res. 2019;149:37–43.
16. Bryant L, McKinnon ET, Taylor JA, Jensen JH, Bonilha L, Bezenac C, et al. Fiber ball white matter modeling in focal epilepsy. Hum Brain Mapp. 2021;42(8):2490–507.
17. van Klink NEC, van't Klooster MA, Leijten FSS, Jacobs J, Braun KPJ, Zijlmans M. Ripples on rolandic spikes: a marker of epilepsy severity. Epilepsia. 2016;57(7):1179–89.
18. Kramer MA, Ostrowski LM, Song DY, Thorn EL, Stoyell SM, Parnes M, et al. Scalp recorded spike ripples predict seizure risk in childhood epilepsy better than spikes. Brain. 2019;142(5):1296–309.
19. Douw L, de Groot M, van Dellen E, Heimans JJ, Ronner HE, Stam CJ, et al. 'Functional connectivity' is a sensitive predictor of epilepsy diagnosis after the first seizure. PLoS One. 2010;5(5):e10839.
20. van Diessen E, Otte WM, Braun KPJ, Stam CJ, Jansen FE. Improved diagnosis in children with partial epilepsy using a multivariable prediction model based on EEG network characteristics. PLoS One. 2013;8(4):e59764.
21. Leu C, Stevelink R, Smith AW, Goleva SB, Kanai M, Ferguson L, et al. Polygenic burden in focal and generalized epilepsies. Brain. 2019;142(11):3473–81.
22. Weber YG, Nies AT, Schwab M, Lerche H. Genetic biomarkers in epilepsy. Neurotherapeutics. 2014;11(2):324–33.
23. Pitkänen A, Löscher W, Vezzani A, Becker AJ, Simonato M, Lukasiuk K, et al. Advances in the development of biomarkers for epilepsy. Lancet Neurol. 2016;15(8):843–56.
24. Surges R, Kretschmann A, Abnaof K, van Rikxoort M, Ridder K, Fröhlich H, et al. Changes in serum miRNAs following generalized convulsive seizures in human mesial

temporal lobe epilepsy. Biochem Biophys Res Commun. 2016;481(1–2):13–8.
25. Loth E, Ahmad J, Chatham C, López B, Carter B, Crawley D, et al. The meaning of significant mean group differences for biomarker discovery. PLoS Comput Biol. 2021;17(11):e1009477.
26. Ma H, Bandos AI, Gur D. On the use of partial area under the ROC curve for comparison of two diagnostic tests. Biom J. 2015;57(2):304–20.
27. Stevelink R, Luykx JJ, Lin BD, Leu C, Lal D, Smith AW, et al. Shared genetic basis between genetic generalized epilepsy and background electroencephalographic oscillations. Epilepsia. 2021;62(7):1518–27.
28. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to meta-analysis. Hoboken, NJ: John Wiley & Sons, Inc.; 2009.
29. Slinger G, Otte WM. The importance of discriminative power rather than significance when evaluating potential clinical biomarkers in epilepsy research (v1.0). Zenodo. 2022. https://doi.org/10.5281/zenodo.7095386
30. Linnet K, Bossuyt PMM, Moons KGM, Reitsma JB. Quantifying the accuracy of a diagnostic test or marker. Clin Chem. 2012;58(9):1292–301.
31. Hazra A. Using the confidence interval confidently. J Thorac Dis. 2017;9(10):4124–9.
32. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology. 2003;229(1):3–8.
33. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al. ILAE official report: a practical clinical definition of epilepsy. Epilepsia. 2014;55(4):475–82.
34. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58(4):512–21.
35. Smit DJA, Wright MJ, Meyers JL, Martin NG, Ho YYW, Malone SM, et al. Genome-wide association analysis links multiple psychiatric liability genes to oscillatory brain activity. Hum Brain Mapp. 2018;39(11):4183–95.
36. van Diessen E, Lamberink HJ, Otte WM, Doornebal N, Brouwer OF, Jansen FE, et al. A prediction model to determine childhood epilepsy after 1 or more paroxysmal events. Pediatrics. 2018;142(6):e20180931.
37. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2021.
38. Barbour AJ, Parker RL. Psd: adaptive, sine multitaper power spectral density estimation for R. Comput Geosci. 2014;63:1–8.
39. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
40. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8(4):283–98.
41. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. Jama. 1994;272(2):122–4.
42. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York: Lawrence Erlbaum Associates; 1988.
43. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003;56(5):441–7.
44. Cook NR. Methods for evaluating novel biomarkers – a new paradigm. Int J Clin Pract. 2010;64(13):1723–7.
45. Cook NR. Quantifying the added value of new biomarkers: how and how not. Diagnostic Progn Res. 2018;2(1):14.
46. Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem. 2012;58(12):1636–43.
47. Josephson CB, Wiebe S. Precision medicine: academic dreaming or clinical reality? Epilepsia. 2021;62:S78–89.
48. Sullivan GM, Feinn R. Using effect size or why the P value is not enough. J Grad Med Educ. 2012;4(3):279–82.

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: Slinger G, Stevelink R, van Diessen E, Braun KPJ, Otte WM. The importance of discriminative power rather than significance when evaluating potential clinical biomarkers in epilepsy research. Epileptic Disord. 2023;25:285–296. https://doi.org/10.1002/epd2.20010

Test yourself

1. What is true about the relationship between sample size and discriminative power?
   A. The larger the sample size, the better the discriminative power
   B. The larger the sample size, the worse the discriminative power
   C. Sample size only impacts discriminative power in cases with small sample size
   D. Sample size does not have an impact on discriminative power
2. How does data variability (e.g., the standard deviation) impact effect size?
   A. Data variability does not have an impact on effect size
   B. An increase of the data variability leads to a decrease of the effect size
   C. A decrease of the data variability leads to a decrease of the effect size
3. Suppose you have found a new variable Y that might aid in discriminating epileptic seizures from vasovagal events. Its discriminative power, expressed as AUC, however, is only 0.62. What might help best to increase the AUC?
   A. Test the performance of variable Y in another, independent population
   B. Combine variable Y with other variables in a multivariable model
   C. Make sure you enter an equal number of seizures and vasovagal events in your analyses

Answers may be found in the supporting information.