Check for updates

Received: 4 April 2026 | Revised: 27 August 2026 | Accepted: 28 August 2026

DOI: 10.1002/epi.70488

Epilepsia®

RESEARCH ARTICLE

Inter-rater reliability on rhythmic and periodic EEG patterns using Standardized Computer-based Organized Reporting of EEG (SCORE): A multicenter study

Elena Tartara[1] | **Harald Aurlien**[2,3] | **Melita Čačić Hribljan**[4] | **Francesca Bisulli**[5] | **Laura Canafoglia**[6] | **Nicolas Gaspard**[7] | **Carla Bentes**[8,9,10] | **Rūta Mameniškienė**[11] | **Elena Ballante**[12] | **Alessandro Amatu**[13] | **Maria Vlachou**[14] | **Lorenzo Ferri**[15] | **Elena Pasini**[15] | **Chiara Pastori**[6] | **Andrea Stabile**[6] | **Charlotte De Maeseneire**[16] | **Lorenzo Ferlini**[16] | **Ana Catarina Franco**[8,9,10] | **Ana Rita Peralta**[8,9,10] | **Hanna-Reetta Lajunen**[17] | **Alessandro Principe**[18,19,20] | **Alexis Arzimanoglou**[21] | **Sándor Beniczky**[22]

Correspondence
Elena Tartara, IRCCS Mondino Foundation, Epilepsy Center, Member ERN EpiCARE, Pavia, Italy.
Email: elena.tartara@mondino.it

Funding information
EU4Health Programme

Abstract

Objective: This multicenter study aimed to evaluate inter-rater reliability (IRR) for rhythmic and periodic patterns (RPPs), background activity, and interictal findings using the Standardized Computer-based Organized Reporting of EEG (SCORE) across 10 European Reference Network (ERN) EpiCARE centers.

Methods: Fifteen neurologists and neurophysiologists, divided into two panels, independently analyzed a total of 277 adult electroencephalography (EEG) recordings, including RPP cases and controls without RPPs, using 5-min EEG epochs. IRR was measured among pairs of raters and quantified using the maximum-likelihood kappa (κML), an unbiased reliability estimator. EEG studies evaluated by at least six raters were included in the analyses ($n = 263$).

Results: The overall identification of RPP's presence vs absence achieved substantial agreement (κML: .61–.66). Reliability for Main Terms 1 and 2 was substantial to almost perfect (κML: .73–.89), matching or exceeding previously published results. Notably, individual RPP types (κML up to .98) and complex modifiers, such as evolving dynamics (κML: .74–.87) and triphasic morphology (κML: .60–.88), achieved high reliability. Although background activity components showed almost perfect consensus (κML: .84–.98), agreement for posterior dominant rhythm was lower (κML: .06–.47). In contrast, the identification of sporadic interictal epileptiform discharges was high (κML: .89).

For affiliations refer to page 14.

© 2026 International League Against Epilepsy.
All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

Significance: The SCORE framework enabled a detailed and comprehensive classification of RPPs, while also capturing background activities and other non-RPP findings in a setting that closely resembles clinical practice. It highlighted complex characteristics for which IRR remains low.

KEYWORDS

EEG, ERN EpiCARE, ictal-interictal continuum, inter-rater reliability, rhythmic and periodic EEG patterns, SCORE

1 | INTRODUCTION

Rhythmic and periodic electroencephalography (EEG) patterns (RPPs) pose a significant interpretative challenge in intensive and neurological care, being linked closely to clinical complexity and prognosis.[1–4] Their importance is heightened because they often lie on the ictal-interictal continuum (IIC), complicating seizure-risk assessment and therapeutic decision-making.[5,6] To address these challenges, the American Clinical Neurophysiology Society (ACNS) developed standardized terminology.[7,8]

Studies of inter-rater reliability (IRR) in RPP identification and classification have highlighted several methodological limitations that hinder cross-study comparability. Early investigations were constrained by small groups of evaluators and limited technical flexibility.[9,10] Later studies expanded the number of raters and incorporated structured training programs, leading to measurable improvements in agreement.[11,12] Most focused on mixed intensive care unit (ICU) populations, although some targeted specific groups,[9,13] thereby limiting generalizability. Furthermore, few previous studies include recordings without RPPs, better reflecting real-world clinical complexities.[9,14]

Methodological heterogeneity also extends to sample duration: brief 10-s segments may overestimate agreement on static features while missing longer-term dynamics,[10,11] whereas longer recordings tended to reduce IRR, likely reflecting greater complexity.[9] Although clear, uniform nomenclature improves the consistency and clinical interpretation of EEG results,[11,13–14] IRR varies considerably: agreement is high ($\kappa > .80$) for broad categories such as the presence of seizures or RPP type, but declines to fair-to-moderate levels (κ .20–.60) for main and minor modifiers such as dynamics, "plus" features, and triphasic morphology, which may influence the association with seizures and clinical decision-making along the IIC. Technical constraints such as fixed montages and filters may have influenced earlier results,[9–12] whereas digital EEG software allowing montage and filter manipulation aided systematic data collection.[15,16]

The standardized reporting platform SCORE (Standardized Computer-based Organized Reporting of EEG) offers a robust framework for objective, reproducible evaluation.[17,18] By integrating ACNS terminology, the International Federation of Clinical Neurophysiology (IFCN) glossary,[19] and International League Against Epilepsy (ILAE) classifications within a hierarchical, context-sensitive set of predefined terms, and by permitting interactive manipulation of montages and filters on recordings of variable duration, SCORE replicates the clinical review station while compelling systematic interpretation not only of general categories but also of the complex characteristics that have shown lower agreement in previous studies.[20] Within the EpiCARE European Reference Network (ERN), this study assessed inter-rater agreement using SCORE-EEG as a standardized framework to harmonize RPP classification across a large volume of EEG recordings from highly specialized centers.

Key points

- First multicenter rhythmic and periodic pattern (RPP)–inter-rater reliability (RPP-IRR) study to combine standardized interactive scoring (Standardized Computer-based Organized Reporting of EEG [SCORE]), 5-min epochs, and unbiased maximum-likelihood kappa (κML).
- RPP detection, seldom tested before, was substantially reliable; type and main features classification remained highly reliable.
- Features long considered unreliable (evolving dynamics, triphasic morphology, plus modifiers) proved highly reliable.
- Frequency, sharpness, fluctuating/static dynamics, and lobar localization showed poor-to-fair reliability and warrant further study.

1.1 | Study objectives

This study aimed to investigate IRR of RPPs using SCORE-EEG across EpiCARE centers, determining

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

agreement among experts on all assessable RPP characteristics, as well as background activity and interictal findings.

2 | MATERIALS AND METHODS

2.1 | Study design and patient population

This study utilized EEG recordings from adult patients collected in routine clinical practice between January 2018 and December 2022, across seven EpiCARE centers, prioritizing recordings with at least 19 electrodes placed according to the International 10–20 System.

EEG was collected by specialists at each center, who also participated in the evaluation process, and structured into case and control datasets. Recordings were selected by direct visual review of the raw EEG traces at each center, independently of the wording of the original clinical report. Inclusion criteria: the collected EEG recordings had to include RPPs of any etiology, and were sourced from adult inpatients hospitalized in critical care, intermediate care, or stroke unit settings. A control group of EEG recordings without RPPs ("control EEGs") included various abnormalities such as slow waves and interictal epileptiform discharges, as well as normal findings; controls were selected on the absence of RPPs, without formal matching for age, etiology, or level of consciousness. A 1:1 ratio of RPP cases to control EEG studies was maintained at inclusion. Clinical characteristics beyond EEG and demographics were not collected, and the ICU vs non-ICU proportion could not be retrieved retrospectively after de-identification.

For each EEG recording, a 5-min epoch was extracted starting from the first observation of the RPPs. These segments were anonymized, assigned a standardized alphanumeric code, comprising a two-letter center identifier followed by sequential numbers, and converted into European Data Format (EDF or EDF+). The recordings were then uploaded to the SCORE-EEG platform following a randomization sequence specific to each center.

EEG recordings were reviewed independently by 15 expert neurologists and neurophysiologists from 10 EpiCARE centers using the EEG reader (NicoletOne) linked to the SCORE-EEG system.

All raters were experienced in ICU-EEG terminology, averaging 15.8 ± 7 years of experience in EEG interpretation and 13.4 ± 6 years in ICU-EEG interpretation (mean ± SD). None of the raters held 2021 ACNS certification; however, most (13 of 15) had completed specialized fellowships in European countries, all at centers with an established continuous EEG (cEEG) monitoring program for critically ill patients. A Summary Table of raters' characteristics is presented as Table S4.

To ensure consistency, all raters attended a 1-h live online session led by author S.B., dedicated to the SCORE-EEG software, with hands-on practice on EEG samples external to the study dataset; the session was recorded and redistributed for asynchronous review.

EEG scoring was conducted from December 2023 to November 2024.

2.2 | SCORE framework and scored variables

EEG recordings were assessed and reported using the Standardized Computer-based Organized Reporting of EEG (or SCORE), developed with international consensus under the auspices of the International Federation of Clinical Neurophysiology (IFCN) and endorsed by the ILAE.[17,18] SCORE uses a context-sensitive interface: raters select clinically relevant features from organized "findings" folders, replacing free-text reporting with predefined terminology to generate standardized reports. SCORE is distributed by Holberg EEG AS in Bergen, Norway, and a free educational edition is available online[21] (https://www.holbergeeg.com/the-score-standard).

The present study used the second version of SCORE (2017), which predates the 2021 ACNS terminology. Although "Unilateral independent" (introduced in 2021) is not available as a default category, SCORE captures this pattern by scoring the same EEG grapho-element in two different location-regions of the same hemisphere. "Polyspike-and-wave" discharges are subsumed under the Main Term 2 category "Spike/polyspike or sharp waves (SW)." Within SCORE, "modulators" denotes the state- and stimulus-dependent features that distinguish spontaneous from stimulus-induced patterns, and "abnormal rhythmic activity" denotes non-paroxysmal pathological rhythmic activity that meets neither the RPP criteria nor those for ACNS rhythmic delta activity. For laterality, "bilateral synchronous" and "diffuse" are not synonymous in SCORE: "generalized" RPPs were operationally defined as "bilateral" + "synchronous" + region, whereas "diffuse" denoted patterns occurring in all or most regions, on both sides, but asynchronously.[18] Raters evaluated the following variables:

- Rhythmic and Periodic Patterns (or RPPs): Main Term 1 (Generalized, Lateralized, Bilateral Independent, Multifocal), Main Term 2 (Periodic Discharges, Rhythmic Delta Activity, Spike/Polyspike or Sharp Waves), morphology, location (mandatory), time-related features, and modulators (spontaneous vs stimulus-induced; optional)
- Background Activity (BA): posterior dominant rhythm (frequency, amplitude, asymmetry, eye-opening

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

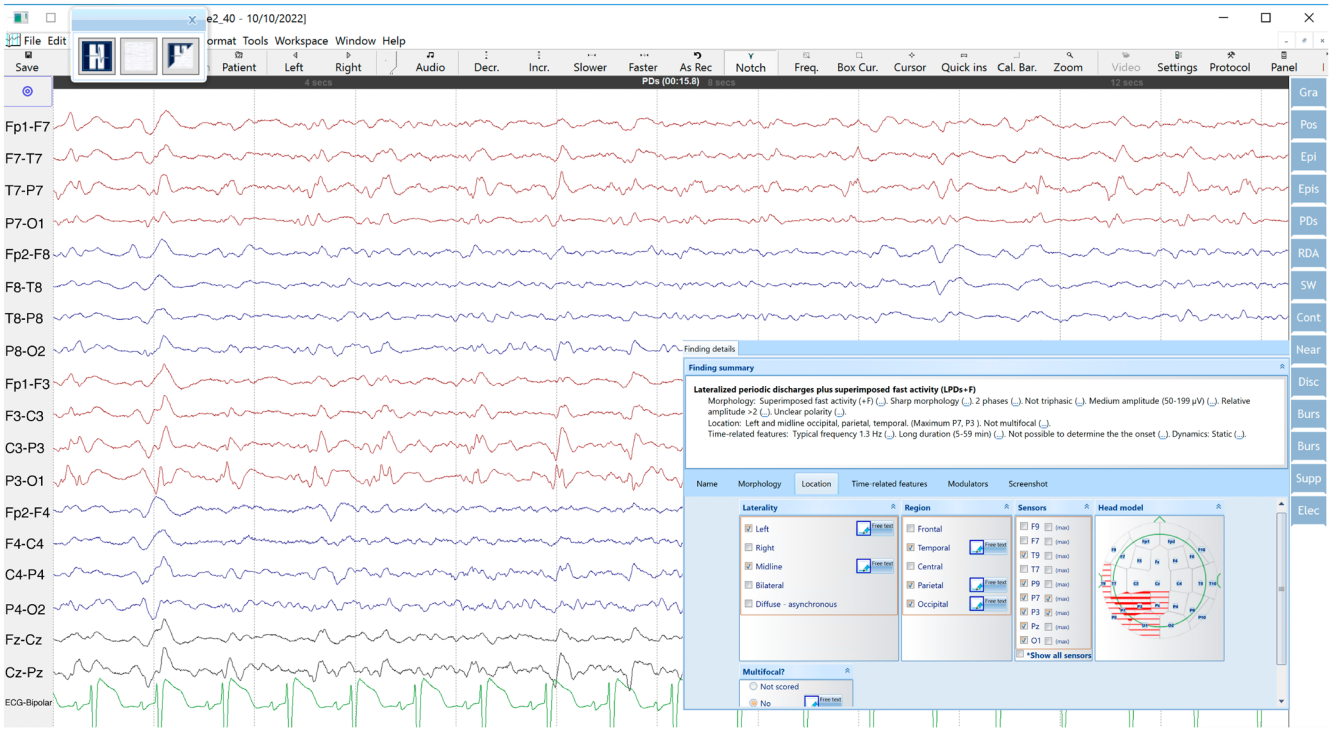

FIGURE 1 Feature extraction and documentation using the Standardized Computer-based Organized Reporting of EEG (SCORE) electroencephalography (EEG) system. EEG showing rhythmic and periodic pattern (RPP) is shown. Marking an abnormality on EEG (black label "PD" - periodic discharges -at the top of the screen) automatically inserts an entry to be characterized in SCORE EEG (blue window). EEG features (including classification of RPP shown here) are scored by clicking on the predefined lists in the software. All scored features are saved automatically in the database. The content in the text box "Finding summary" is generated automatically, as the user scores the observed features.

reactivity, absence, significance), other organized rhythms (morphology, location, time-related features, modulators), special features: continuous background activity, nearly-continuous background activity, discontinuous background activity, burst-suppression, burst-attenuation, suppression, electrocerebral inactivity*

- Interictal Findings (IF): epileptiform interictal activity, abnormal low-frequency activity, abnormal rhythmic activity.

The system enabled montage switching (longitudinal or transversal bipolar, and common average) and adjustment of gain, filters, and time base; a frequency spectrum could be computed on a manually selected segment, whereas continuous spectrograms (quantitative EEG, qEEG) and voltage maps were unavailable. A representative screenshot integrated with the NicoletOne viewer is shown in Figure 1.

*"Electrocerebral inactivity" is retained here as the SCORE special-feature label; the formal ACNS/clinical criteria for electrocerebral inactivity were not assessed within the 5-min research epochs, "suppression" being the operative category for flat backgrounds. This feature was not present in any EEG of the dataset.

2.3 | Statistical analysis

Raters were divided randomly into two panels (seven and eight), each assigned a different randomized EEG list to manage the workload; allocation was fully randomized, without center or expertise stratification.

Four centers contributed more than one rater: two centers contributed two raters each, who were assigned to different panels, whereas two centers contributed three raters each, with two of them assigned randomly to the same panel and the third to the other panel.

Inter-rater reliability (or IRR) was measured among pairs of raters and quantified using the maximum likelihood kappa (κML)[22] as an unbiased estimator of interrater reliability; κML is reported as the primary coefficient throughout the main text, in Table 1, and Tables S1–S3.

Coefficients were interpreted according to the Landis–Koch classification, which ranges from "poor" (<.20) to "almost perfect" (>.80).[23]

The analysis used data from all raters' SCORE-EEG reports. Concordance was assessed hierarchically: first for case (RPP) identification, then for each grapho-element type (RPPs, BA, IF), and, where sample size allowed, their specific characteristics (localization, morphology,

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Inter-rater reliability (IRR) for rhythmic and periodic patterns (RPPs) and their modifiers: Comparison between the present study and previous IRR studies.

Item/variable	Present study, κML (P1→P2)[a]	Gerber 2008[b]	Mani 2012[c]	Gaspard 2014[d]	Westhall 2015[f]	Halford 2015[g]	Foreman 2016[e]	Jing 2023[h]
Study characteristics								
Setting	ICU/non-ICU	ICU (SAH)	ICU	ICU	ICU (post-CA)	ICU	ICU	ICU/EMU / routine EEG
N EEG/epoch duration analysed	263/5 min	58 (10 s, Phase I) 30 (full-length, 20–30 min, Phase II)	~14 (10 s)	37 (10–60 s)	103/>20 min	30 (1 h)	20 (120 s)	50,697 (10 s)
Option to modify display settings	yes	yes	no	no	Yes	yes	yes	yes
N raters/N centres	15/10	7/1	16/11	49/multicenter	4/ 8 sites	8/3	11/ multicenter	30 CNP + 94 others/18
Raters' ACNS training	No	No (2005 prop.)	Yes (ACNS module)	Yes (ACNS slides)	Yes (ACNS module)	Partial (4/8 cert.)	Yes (cert.)	Partial (CN fellows)
Control cases	Yes	Yes	No	No	No	NR	No	Reported as "other"
Overall agreement and Main Terms								
RPP (overall)	.66–.61	.68 (Phase I) .44 (Phase II)	—	—	—	—	—	—
Main Term 1	.86–.78	.49 (Phase I) .42 (Phase II)	.87	.89	—	—	—	—
Generalized	.88–.84							
Lateralized	.87–.78							
Bilateral Independent	.98–.92							
Multifocal	.98–.98							
Unilateral Independent	na							
Main Term 2		.39 (Phase I) .37 *(overall)* (Phase II)	.92	.80 *(overall)*	.57 (including absent)	.38	—	—
PD	.82–.73							
RDA	.89–.85							
SW	.87–.79							

(Continues)

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Item/variable	Present study, κML (P1→P2)[a]	Gerber 2008[b]	Mani 2012[c]	Gaspard 2014[d]	Westhall 2015[f]	Halford 2015[g]	Foreman 2016[e]	Jing 2023[h]
Individual RPP type								
GPD	.81–.83	—	—	—	—	—	.81	.45
LPD	.81–.67							.56
BIPD	.98–.95							.33
GRDA	.93–.94							.20
LRDA	.95–.91							
GSW	.96–.94							
LSW	.95–.87							
Main and minor modifiers								
Duration		.34 *(overall)* (Phase II)	—	—	—	—	—	—
Very brief	na-.33							
Brief	na-1.00							
Intermediate	na-.39							
Frequency		.34 (Phase I) .27 *(overall)* (Phase II)	80% *(agreement)*	.95 *(overall)*		—	.92 *(overall)*	—
Typical frequency	na-.03				.82			
Frequency range	.23–.25				.74 (max)			
Absolute amplitude		—	93% *(abs+rel agr.)*	.94 *(overall)*	.42 (abs, all)	—	.88 *(overall)*	—
20–49 μV	.94–.87							
50–199 μV	.76–.69							
≥200 μV	.85–.84							
Number of phases		—	—	.83 *(overall)*	—	—	.87 *(overall)*	—
1	.92–.94							
2	.13–.30							
3	na-.08							
>3	.58–.73							

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Item/variable	Present study, κML (P1→P2)[a]	Gerber 2008[b]	Mani 2012[c]	Gaspard 2014[d]	Westhall 2015[f]	Halford 2015[g]	Foreman 2016[e]	Jing 2023[h]
Sharpness		—	—	.85 *(overall)*	.73 (overall)	—	.80 *(overall)*	—
Blunt	.85–.96							
Sharply contoured	.64–.38							
Sharp	.19–.05							
Spiky	.62–.50							
Plus modifiers				.34 *(overall)*		—		—
Any	.47–.28	.12 (Phase I) .08 (Phase II)	.25	.19	.17		.56	
+F	.58–.64		.54	.66				
+R	.90–.79		.62	.67				
+S	.91–.93		.16	.82				
+FR	.91–.85							
+FS	.98							
Dynamics		.04 (Phase I) .07 *(overall)* (Phase II)	—	.21 *(overall)*	.19	—	.36 *(variability)*	—
Evolving	.87–.74							
Fluctuating	.18–.26							
Static	.34–.27							
Triphasic morphology	.88–.60	—	—	.58	−.07	—	.33	—
Onset			—	—	—	—	—	—
Gradual	na-.61	.05 (Phase II)						
Sudden	na	.03 (Phase II)						
Amplitude asymmetry – Relative amplitude								
Amplitude asymmetry	na-.63	—	93% *(abs+rel agr.)*	.66 [.65–.67]	—	—	.69 [.47–.84]	—

(Continues)

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Item/variable	Present study, κML (P1→P2)[a]	Gerber 2008[b]	Mani 2012[c]	Gaspard 2014[d]	Westhall 2015[f]	Halford 2015[g]	Foreman 2016[e]	Jing 2023[h]
Synchrony & Multifocality								
Synchrony & multifocality				—	—	—	—	—
Bilateral synchronous	.80-na							
Bilateral independent	.80-na	14% *(BI, raw)*	BI .83					
Multifocal (yes/no)	.94 (P1)	0% *(Mf, raw)*	Mf not assessed					

Note: Coefficients are reported as in each source. Values are proportions unless a % (percent agreement) is shown; 95% CIs are in square brackets. *(overall)* = a single coefficient reported by that study for the whole category, not split by subtype; *na* = estimate not computable (insufficient data) in that panel; — = item not assessed/not applicable in that study; NR = not reported. P1/P2 = Panel 1/Panel 2. Even when considering different metrics, it is noticeable that, in the present study, leveraging the standardized assessment provided by SCORE enables greater granularity in evaluating RPP characteristics.

Abbreviations: ACNS, American Clinical Neurophysiology Society; BI(PD), bilateral independent (periodic discharges); CNP, clinical neurophysiologist; EMU, epilepsy monitoring unit; GPD/GRDA/GSW, generalized periodic discharge/generalized rhythmic delta activity/generalized spike-and-wave; ICU, intensive care unit; LPD/LRDA/LSW, lateralized periodic discharges/lateralized rhythmic delta activity/lateralized spike-and-wave; PD, periodic discharges; RDA, rhythmic delta activity; RPP, rhythmic and periodic patterns; SAH, subarachnoid hemorrhage; SW, spike/sharp-and-wave; +F/+R/+S/+FR/+FS, superimposed fast/rhythmic/sharp waves or spikes, or sharply contoured, and combinations.

[a]Present study: maximum-likelihood κ (κML; Westover et al.[22]). Paired values are listed in fixed Panel 1 → Panel 2 order (not sorted by magnitude); a single value indicates identical Panel 1 and Panel 2 estimates.

[b]Gerber et al.[9]: mean pairwise Cohen's κ (weighted κ for linear-scale items), single center. Phase I (10-s epochs, five readers) and Phase II values are shown throughout (30 full-length recordings, 7 readers).

[c]Mani et al.[12]: Cohen's κ from the Critical Care EEG Monitoring Research Consortium (CCEMRC) certification test (16 raters, 11 centers). Amplitude (4 absolute +2 relative items) and frequency were reported as percent agreement only (no κ). Subtype κ for Main Term 2 (PD .90, RDA .93, SW .93) and the location κ used for synchrony/multifocality (bilateral independent .83; multifocal not represented among items) are from the per-feature analysis; +FR and + FS combinations were not tested. The number of EEG studies is not stated in the article, but is also inferred by Jing et al., [14] who reported the same data.

[d]Gaspard et al.[11]: Gwet's AC1 (categorical) and quadratic-weighted AC2 (ordinal: sharpness, absolute/relative amplitude, frequency, number of phases, evolution), reported as percentages and shown here as proportions. Main Term 2, sharpness, amplitude, frequency, phases, and evolution were reported as single category-level coefficients, not split by subtype.

[e]Foreman et al.[15]: Gwet's AC1 (categorical)/quadratic-weighted AC2 (ordinal), multicentre. Cohort restricted to generalized periodic discharges, so Main Terms, non-GPD subtypes, and synchrony/multifocality are not applicable. "Variability" (whether discharges varied) is shown against Dynamics.

[f]Westhall et al.[13]: median pairwise Cohen's κ across six rater pairs (weighted κ for ordinal terms) on full-length post-cardiac-arrest EEGs. Main Term 1 (location), relative amplitude, number of phases, duration and onset could not be assessed (skewed distributions or excluded after a pilot). Frequency rows give the typical (κ = .82) and maximum (κ = .74) discharge frequency. The study also reported substantial agreement for background continuity (κ = .76) and voltage (κ = .65) and for the prespecified "highly malignant" pattern (κ = .71), not shown here.

[g]Halford et al.[16]: κ for detection of periodic discharges (.38, shown against Main Term 2 PD) and of electrographic seizures (.58, no corresponding row) in 1-h recordings; 8 raters, 3 centers. Pattern modifiers were not assessed.

[h]Jing et al.[14]: pairwise κ for a six-class ictal–interictal continuum scheme (the source additionally reports a higher majority-vote κ); 18 centers. Bilateral-independent, multifocal, spike/sharp-wave subtypes and the modifier set were not separately assessed. Among the large dataset, the absence of EEG patterns of interest (seizures, LPDs, GPDs, LRDA, GRDA) was classified as "other," without specific inclusion of control cases.

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

time-related features). Main Term 1 variables were treated as mutually exclusive, and Main Term 2 variables as non-mutually exclusive.

Sample size estimation was based on Foreman et al.,[15] who reported an IRR of .81 for the main terms "generalized" and "periodic" using Gwet's coefficient. Power analysis for dichotomous features was conducted using the kappaSize package in R. Because the required sample size decreases as the number of raters increases,[24] the six-rater calculation is conservative. Assuming a significance level of .05 and six evaluators, a minimum of 70 EEG assessments per rater was required.

3 | RESULTS

From an initial dataset of 286 EEG recordings, 9 EEG recordings could not be evaluated due to technical issues, and 263 met the inclusion criteria for final analysis, each evaluated by at least six independent raters. The median number of EEG recordings scored by each rater was 131.

The detailed results of the inter-rater agreement analyses are presented in Table S1 (RPP findings), S2 (Background activity and Interictal findings), and Table S3 (raters' IRR), whereas a synoptic comparison of the principal results with the published literature is provided in Table 1.

3.1 | Inter-rater reliability on rhythmic and periodic patterns

The overall IRR for RPPs reporting was substantial and stable across both rater panels (κML: .61–.66). Compared with the panel consensus, individual raters showed moderate-to-almost-perfect IRR (individual κML .51–.93; Table S3). Primary classification categories, specifically Main Term 1 and Main Term 2, demonstrated consistently high IRR: Main Term 1 achieved agreement coefficients ranging from .78 to .86. Similarly, Main Term 2 exhibited substantial to almost perfect agreement for periodic discharges (κML: .73–.82) and reached near-perfect levels for rhythmic delta activity and spike/polyspike and sharp-wave patterns (κML: .79–.89) (Figure 2). The unilateral independent pattern (Main Term 1) was present in only three EEG recordings; therefore estimation of its IRR was not possible.

The individual RPP patterns also exhibited high IRR (κML: .67–.98), except for the rare patterns where the low prevalence of ratings did not allow for IRR calculation (Bilateral

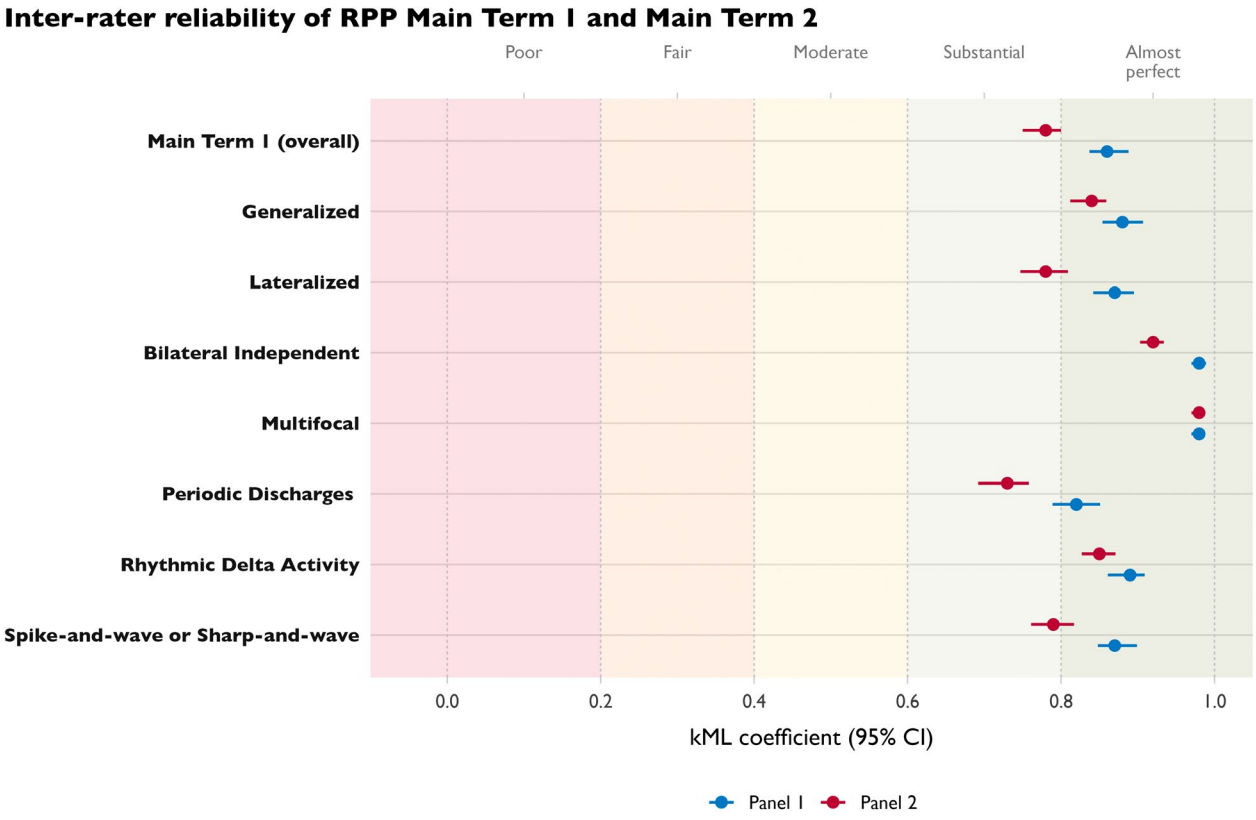

FIGURE 2 The inter-rater reliabilities (IRRs) in the two raters' panels for the classification of rhythmic and periodic patterns (RPPs) Main Term 1 and Main Term 2.

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

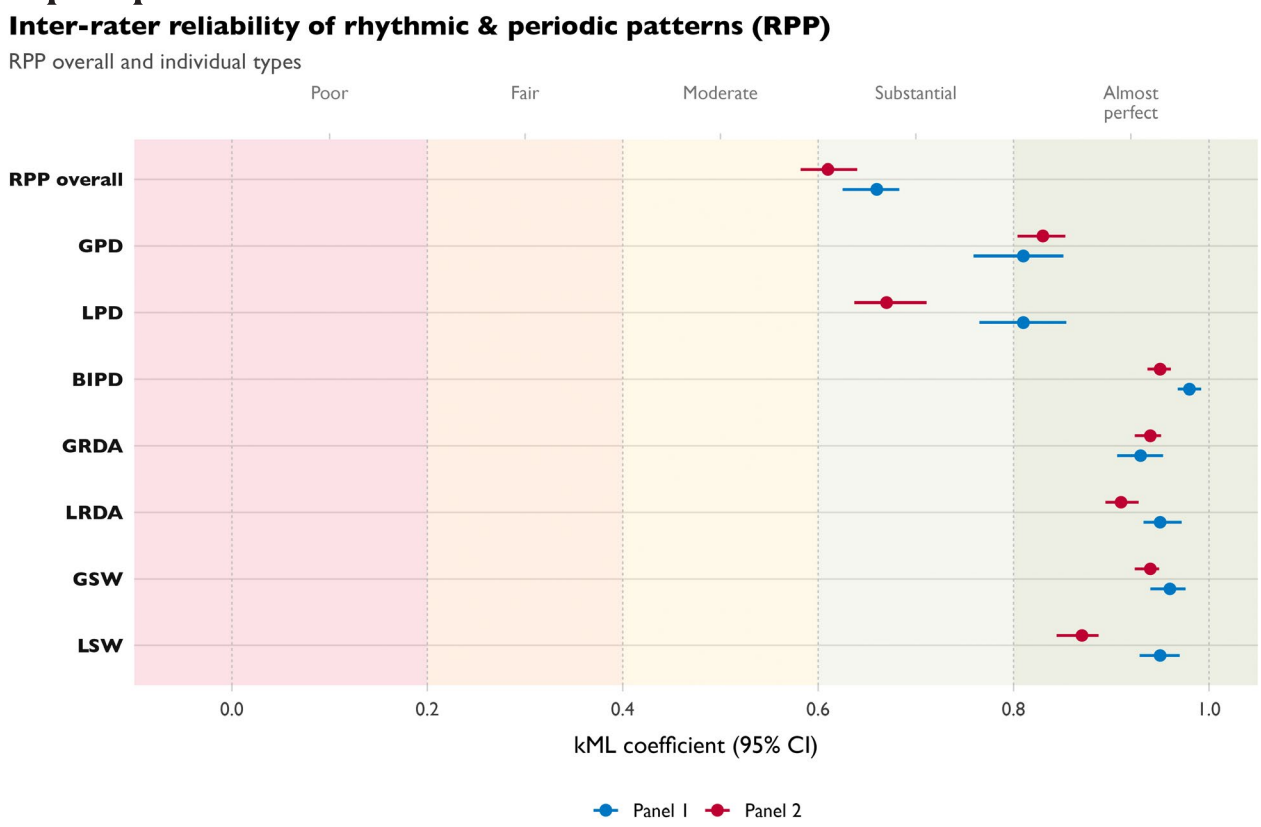

FIGURE 3 The IRR in the two raters' panels for the classification of RPP presence (RPPs overall) and each RPP type. BIPD, bilateral independent periodic discharges; GPD, generalized periodic discharges; GRDA, generalized rhythmic delta activity; GSW, generalized spike/polyspike or sharp-wave; IRR, inter-rater reliability; LPD, lateralized periodic discharges; LRDA, lateralized rhythmic delta activity; LSW, lateralized spike/polyspike or sharp-wave; RPP, rhythmic and periodic pattern.

Independent Rhythmic Delta Activity, BIRDA, four EEG studies, Multifocal Periodic Discharges, MfPD, five EEG studies, Multifocal Rhythmic Delta Activity, MfRDA, three EEG studies, Unilateral Independent Periodic Discharges, UIPD, two EEG studies, and Unilateral Independent Rhythmic Delta Activity, UIRDA, one EEG studies) (Figure 3).

IRR on topographical, morphological, and time-related characteristics was lower. Although hemispheric lateralization (left and right) showed almost perfect agreement (κML .84–.90), regional localization was significantly less reliable; specifically, the temporal region showed poor to moderate agreement (κML .15–.40) and the central region remained within the fair range (κML .25–.32); conversely, bilateral synchronous patterns achieved almost perfect reliability (κML .80, Panel 1).

Among quantitative variables, absolute amplitude was consistently concordant (κML .69–.94), whereas frequency metrics (range and typical frequency) were much lower, ranging from poor to fair (κML .03–.25).

IRR for main modifiers varied significantly with feature complexity. Among morphological modifiers, sharpness descriptors such as "sharp" yielded poor reliability (κML: .05–.19), whereas "blunt" or "spiky" features achieved higher concordance (κML .5 to .96).

Regarding "dynamics," a notable finding was the almost perfect agreement for "evolving" patterns in the first panel (κML: .87), which remained substantial in the second panel (κML: .74). Conversely, agreement was poor to fair for fluctuating (κML: .18–.26) and static patterns (κML: .27–.34).

The IRR for the number of phases was generally low, especially for three-phase patterns, which showed particularly poor or non-calculable agreement (κML .08-NA). In contrast, the assessment of triphasic morphology among minor modifiers achieved a significantly higher level of consensus, ranging from substantial to almost perfect agreement (κML .60–.88).

IRR for superimposed activity indicated a discrepancy between general identification and specific pattern classification. Detection of any superimposed activity (Any +) showed moderate to fair agreement (κML .47 and .28), whereas specific plus modifiers had much higher reliability. Combined patterns achieved the best reliability: +FR close to perfect (κML .85–.91) and +FS nearly perfect (κML .98).

Figure 4 shows the IRR for selected Main and Minor Modifiers. Figures S5–S7 illustrate examples of low and high agreement concerning the characteristics of fluctuation, sharpness (sharp), and triphasic morphology.

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

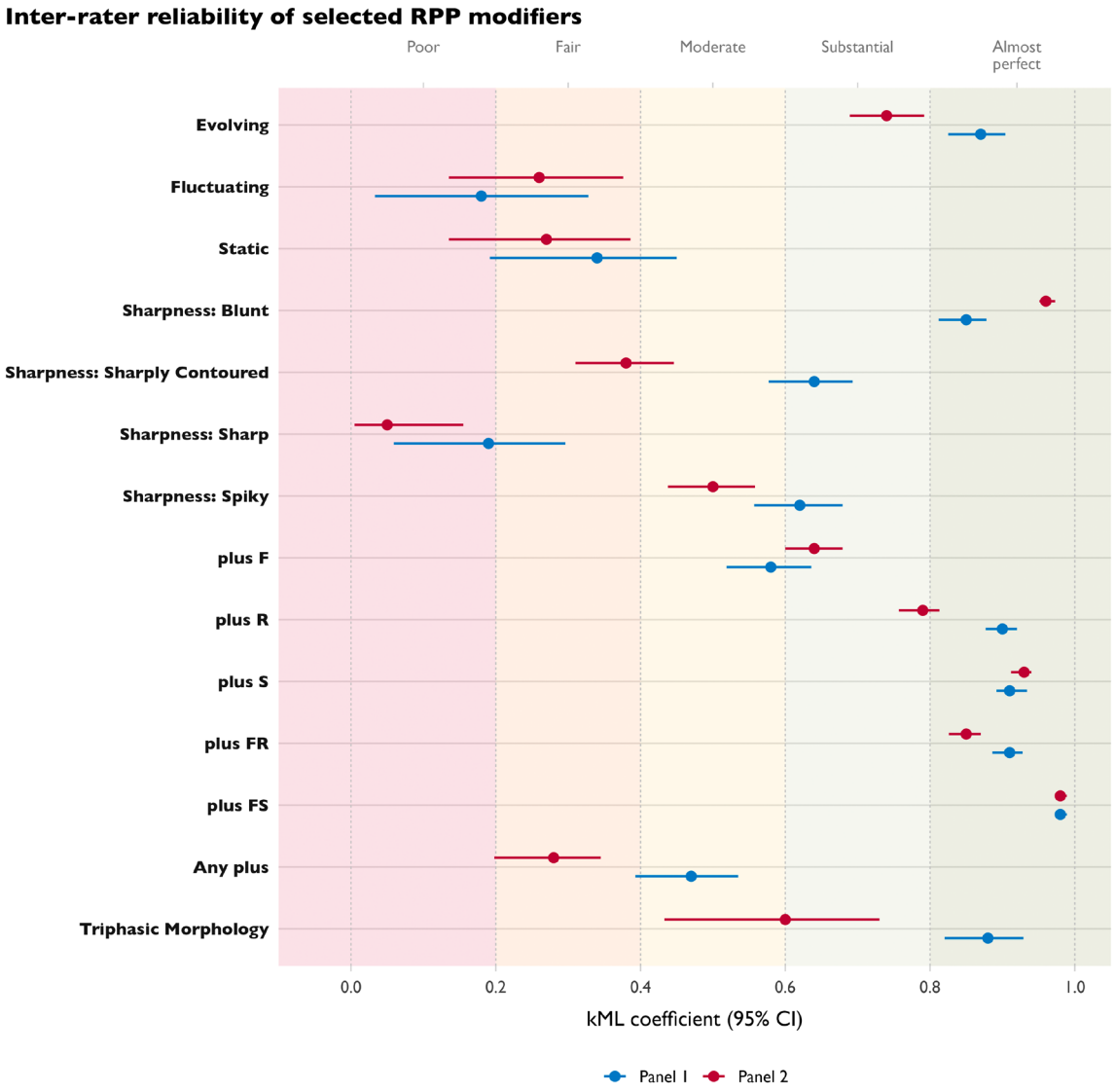

FIGURE 4 The IRRs in the two raters' panels for the classification of selected RPPs, main and minor modifiers. IRR, inter-rater reliability; plus F, plus superimposed fast activity; plus FR, plus superimposed fast and rhythmic activity; plus FS, plus superimposed fast and sharp waves or spikes, or sharply contoured; plus R, plus superimposed rhythmic activity; plus S, plus superimposed sharp waves or spikes, or sharply contoured; RPP, rhythmic and periodic pattern.

3.2 | Inter-rater reliability on background activity and interictal findings

Raters showed very high IRR on the core components of background activity (almost perfect across both panels, κML .84–.98), extending to organized rhythms and several special features. Agreement for both normal and abnormal organized rhythms remained almost perfect (κML: .80–.96). Background continuity was reliably classified: nearly continuous (κML: .85–.90) and discontinuous (κML: .88–.93) backgrounds, together with suppression, burst-attenuation, and burst-suppression patterns (κML: .94–.98); all showed almost perfect agreement across both panels. The main exception was the posterior dominant rhythm, notably lower and ranging from poor in Panel 1 (κML: .09) to moderate in Panel 2 (κML: .47), making it the most controversial element of background evaluation.

Compared with background activity, interictal findings yielded lower, more variable agreement. Interictal findings overall (abnormal low-frequency and epileptiform activities) showed moderate reliability (κML .52, Panel 1; .49, Panel 2), whereas epileptiform interictal activity alone reached almost perfect concordance in Panel 1 (κML .89). IRR for the specific morphology, location, and time-related features of background and interictal findings could not be calculated, owing to the limited evaluations assigned to these variables (46 EEG recordings).

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.3 | Comparability of the two groups of raters (Panel 1 and Panel 2)

Several variables showed overlapping confidence intervals (CIs) and near-identical coefficients across panels, indicating robust, reproducible findings, including overall RPP identification, Main Term 2 patterns (e.g., rhythmic delta activity), generalized pattern types (e.g., generalized periodic discharges [GPD]; generalized rhythmic delta activity [GRDA]), and static dynamics. Others showed non-overlapping 95% CIs, reflecting statistically distinguishable inter-panel differences (e.g., PDs, Spike/Polyspike and Wave, evolving dynamics, and posterior dominant rhythm).

4 | DISCUSSION

In this multicenter EpiCARE study, the SCORE-EEG framework yielded substantial-to-almost-perfect agreement for the core categories of RPPs and of background activity, while exposing a reproducible set of features for which reliability remained poor. The contribution of SCORE here can be read as an operationalization of ACNS terminology: by embedding that terminology within a hierarchical, context-sensitive interface that mandates explicit scoring of every feature, SCORE compelled raters to interrogate the full descriptive apparatus (main terms, morphology, localization, time-related features, modulators, and "plus" modifiers) rather than only the categories habitually captured in free text. This guided granularity, combined with interactive review of 5-min epochs, produced a more complete map of where the ACNS framework is robust and where its application remains difficult.

We acknowledge, however, that the design cannot disentangle the framework's contribution from those of interactive review and longer epochs; reliability reflects the joint influence of terminology, platform, and review conditions.

A distinctive feature of the present work is the assessment of RPP discrimination within a 1:1 case–control design that included recordings without RPPs.

Most prior reliability studies pre-selected EEG samples already known to contain a pattern and assessed only their classification[11,12,15]; the preliminary phase, whether a recording contains an RPP at all, has rarely been examined.[9] By pairing each case with an RPP-negative control (1:1 design), we could quantify this detection step, obtaining substantial agreement for overall RPP identification (κML .61–.66), an advance on the only moderate presence/absence agreement of Gerber et al. for longer recordings, and analogous to the substantial "EEG-wise" reliability demonstrated for interictal-seizure detection.[14]

Beyond detection, the primary classification had a high IRR, matching the almost-perfect findings of Gaspard et al.[11] and Mani et al.[12] "Evolving" dynamics reached substantial-to-almost-perfect reliability, whereas Gaspard et al.[11] obtained only fair agreement for evolution (21%), attributing this to their brief 10-s samples, a limitation our 5-min epochs plausibly overcame. Specific superimposed ("plus") modifiers were highly concordant (κML up to .98), far more so than their global detection ("Any +", κML .28–.47) and surpassing the fast and rhythmic modifiers in Mani et al.[12] Triphasic morphology, historically only fairly reliable (κ .33 in Foreman et al.[15]; 58% in Gaspard et al.[11]), reached substantial-to-almost-perfect levels when scored as a specific minor modifier. Background activity was almost perfect, exceeding the reliability implied by brief-sample encephalopathy grading.[25] Epileptiform interictal activity reached almost-perfect agreement in Panel 1 (κML .89). Although the IFCN criteria for interictal epileptiform discharges were validated mainly in routine, non-ICU recordings,[26,27] this high agreement in a cohort that included critically ill recordings suggests that these morphology-based criteria for the interictal epileptiform graphoelement remain applicable and reliable in the ICU setting, where the interictal-ictal border is otherwise harder to draw. Conversely, several features had low IRR. Typical and range frequency showed poor-to-fair agreement (κML .03–.25), below Mani et al.,[12] who already flagged frequency as among the least reliable descriptors.

No earlier study applied a discrete ±.5 Hz tolerance to frequency; the higher agreement reported elsewhere reflects the metric used, uncorrected percent agreement[12] or weighted kappa/AC2 crediting near-misses,[9,11,15] not lower rater dispersion.

Sharpness behaved similarly: "sharp" yielded poor agreement (κML .05–.19), closely reproducing the slight agreement (κ .16) found for the sharp/spike modifier by Mani et al.[12]

Beyond rater-related noise, three factors are relevant: the protocol did not specify whether the sharpest or the dominant phase should be scored; sharpness duration is defined as measured at the EEG baseline,[8] a convention likewise not specified to raters; and sharpness may inherently carry low signal strength in ambiguous tracings.

A parallel asymmetry affected the dynamics: whereas "evolving" was highly reliable, "fluctuating" (κML .18–.26) and "static" (κML .27–.34) were only poorly to fairly concordant. These categories are distinguished primarily by the degree of "non-evolving change in frequency over time" (the ACNS criterion for a frequency change being ≥.5 Hz), so the poor IRR of frequency itself plausibly reflects, and exacerbates, the unreliability of "fluctuating." It also explains why frequency imprecision spares "evolving" whose directional

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

trend forms a salient gestalt, while undermining the borderline static-fluctuating distinction.[28] Five-minute epochs may be too short to capture slow fluctuations, which may require continuous spectrograms unavailable here, as Gaspard et al.[11] found for dynamic features in brief samples. Regional localization was far less reliable than hemispheric lateralization (κML .84–.90), with the temporal and central regions performing poorly (κML .15–.40), echoing the moderate localization agreement reported by Gerber et al.[9]

The low IRR for specific lobar localization likely reflects the widespread topographical expression of RPPs and the minor role of lobar location in the ACNS classification. Periodic discharges are seldom confined to a single lobe and often show lower amplitudes in a limited number of EEG leads.[29] Rhythmic patterns are generally widespread, prominent anteriorly but extending into posterior regions.[6,30] The absence of detailed regional descriptors in the ACNS lexicon may also contribute.

The most counterintuitive finding was the low reliability of the posterior dominant rhythm (κML .09–.47), an apparently straightforward feature; in pathological, often critically ill backgrounds, it may be attenuated, fragmented, or absent, rendering its presence ambiguous, corroborated by its poor recognition (sensitivity 50%) in Tesh et al.[25]

The experience of the raters, the majority trained at centers with an established continuous EEG monitoring program, likely contributed to the high IRR, but did not resolve the more equivocal descriptors; none held formal ACNS certification, and as certified pairs agree better,[16] this may partly explain residual disagreement on the most demanding features.

Per-rater concordance for RPP identification (Table S3) ranged from moderate to almost perfect (κML .51–.93), with two raters below 70% vs above 79% for the remainder. Retained to avoid circularity and to provide a more realistic picture of the heterogeneity encountered in clinical practice, this variability also suggests SCORE-EEG, already used educationally, could support future targeted training.

Direct numerical comparison with earlier work must be approached with caution, given the heterogeneous statistics (Cohen's κ, Gwet's coefficients, percentage agreement), small or selection-biased samples, and differing ACNS versions and rater experience. We used the maximum-likelihood kappa, unbiased where Gwet's estimator is biased[22] and robust to the prevalence asymmetry underlying the kappa paradox.

Agreement estimates converged across the two independently randomized panels for high-agreement and low-agreement variables alike, indicating that the difficult features reflect genuine properties of the terminology and signal rather than panel-specific noise, as the panels shared no cases, direct cross-panel agreement was not computable.

5 | LIMITATIONS

The cohort consisted exclusively of adults. The case-to-control ratio was 1:1 but we were unable to retrieve the proportion of ICU to non-ICU patients, retrospectively, after de-identification. Controls were selected solely on the basis of the absence of RPP and were not matched for alertness or etiology. The use of 5-min epochs limited our ability to capture longer dynamics; in particular, fluctuating patterns may require longer recordings and continuous qEEG, which were not available in this study. A few descriptors from the SCORE system (IFCN-ILAE terminology) diverge from the ACNS lexicon.

The protocol did not specify the baseline reference for sharpness duration, nor whether the sharpest or dominant phase should be scored. None of the raters held formal 2021 ACNS certification, although the majority (13 of 15) trained at centers with an established cEEG monitoring program.

In addition, two centers contributed multiple raters to a single panel, which may have inflated agreement due to a shared reading culture.

Finally, low-prevalence variables could not be evaluated and all participating centers were located in Europe. SCORE's advantage lies in establishing a common terminological framework for all raters before assessment begins, but the design cannot isolate its contribution from interactive review, epoch length, and the intrinsic clarity of the ACNS/IFCN definitions themselves; any incremental effect of the platform on reliability is therefore not measurable in the present study, and a null platform contribution cannot be excluded.

6 | CONCLUSIONS

Within a large multicenter EpiCARE dataset, reproducible reliability was observed using the structured SCORE-EEG platform for the core categories of rhythmic and periodic patterns, including the rarely examined aspect of RPP detection, as well as for background activity and epileptiform interictal activity; the well-defined ACNS/IFCN terminology may account for this in full, and any additional contribution of the platform was not assessable. The study also identifies, with high granularity, characteristics that remain difficult to score: frequency, sharpness, fluctuating/static dynamics, regional localization, and the posterior dominant rhythm, pointing to

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

targets for terminological refinement, focused training, and future computer-assisted classification.

AUTHOR CONTRIBUTIONS

Elena Tartara: conceptualization, data curation, methodology, project administration, data collection, data reporting, and writing-original draft preparation. **Harald Aurlien**: software development and writing-review and editing. **Melita Čačić Hribljan**: methodology, data collection, data reporting, and writing-review. **Francesca Bisulli**: methodology, data collection, and writing-review. **Laura Canafoglia**: methodology, data collection, data reporting, and writing-review. **Nicolas Gaspard**: methodology, data collection, data reporting, and writing-review. **Carla Bentes**: methodology, data collection, and writing-review. **Rūta Mameniškienė**: methodology, data reporting, and writing-review and editing. **Elena Ballante**: methodology, formal analysis, validation, and writing-review and editing. **Alessandro Amatu**: visualization and writing-review and editing. **Maria Vlachou**: methodology, data collection, data reporting, and writing-review and editing. **Lorenzo Ferri**: methodology, data collection, data reporting, and writing-review and editing. **Elena Pasini**: methodology, data collection, data reporting, and writing-review. **Chiara Pastori**: data reporting, and writing-review. **Andrea Stabile**: methodology, data reporting, and writing-review. **Charlotte De Maeseneire**: data reporting and writing-review. **Lorenzo Ferlini**: data reporting and writing-review. **Ana Catarina Franco**: methodology, data collection, data reporting, and writing-review and editing. **Ana Rita Peralta**: methodology, data collection, data reporting, and writing-review and editing. **Hanna-Reetta Lajunen**: data reporting and writing-review and editing. **Alessandro Principe**: data reporting and writing-review. **Alexis Arzimanoglou**: funding acquisition, project administration, resources, and writing-review and editing. **Sándor Beniczky**: conceptualization, data curation, methodology, project administration, resources, data collection, data reporting, software development, supervision, and writing-review and editing.

AFFILIATIONS

[1]IRCCS Mondino Foundation, Epilepsy Center, Member ERN EpiCARE, Pavia, Italy
[2]Department of Clinical Neurophysiology, Haukeland University Hospital, Bergen, Norway
[3]Natus Medical, Bergen, Norway
[4]Department of Clinical Neurophysiology, University Hospital Copenhagen Rigshospitalet, Copenhagen, Denmark
[5]Department of Biomedical and NeuroMotor Sciences, IRCCS Istituto Delle Scienze Neurologiche di Bologna, Alma Mater Studiorum – University of Bologna, Member ERN EpiCARE, Bologna, Italy
[6]Department of Diagnostic and Technology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Member ERN EpiCARE, Milan, Italy
[7]Université Libre de Bruxelles et Hôpital Universitaire de Bruxelles—Hôpital Erasme, Member ERN EpiCARE, Brussels, Belgium
[8]Reference Centre for Refractory Epilepsies, Hospital de Santa Maria-ULSSM, Member ERN EpiCARE, Lisbon, Portugal
[9]Department of Neurology, Hospital de Santa Maria-ULSSM, Lisbon, Portugal
[10]MUHCRI, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal
[11]Clinic of Neurology and Neurosurgery, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University, Member ERN EpiCARE, Vilnius, Lithuania
[12]Department of Political and Social Sciences, University of Pavia, Pavia, Italy
[13]Anesthesia and Intensive Care Department, Fondazione IRCCS Policlinico san Matteo, Pavia, Italy
[14]Department of Clinical Neurophysiology, Aarhus University Hospital, Member ERN EpiCARE, Aarhus, Denmark
[15]IRCCS Istituto Delle Scienze Neurologiche di Bologna, Member ERN EpiCARE, Bologna, Italy
[16]Neurology Department, Hôpital Universitaire de Bruxelles (HUB), Université Libre de Bruxelles, Brussels, Belgium
[17]Department of Clinical Neurophysiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[18]Epilepsy Program, Hospital del Mar Research Institute, Barcelona, Spain
[19]Department of Neurology, Hospital del Mar, Barcelona, Spain
[20]Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[21]Epilepsy and Neurophysiology of the SJD Barcelona Children's Hospital, ERN Epicare Coordinating Member ERN EpiCARE, Barcelona, Spain
[22]Department of Clinical Neurophysiology, Danish Epilepsy Centre Filadelfia, University of Copenhagen, Member ERN EpiCARE Member, Dianalund, Denmark

ACKNOWLEDGMENTS

We warmly thank the Management Team of the European Reference Network (ERN) for Rare and Complex Epilepsies EpiCARE, funded by the European Commission, for their remarkable daily commitment in

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

coordinating the network, enabling the development of collaborative projects aiming to improve care practices. We are grateful to all individual members and supporting partners of the ERN EpiCARE, and particularly to the members of the Neurophysiology Working Group, for their continued commitment, valuable contributions, and unfailing enthusiasm in advancing our collective missions.

FUNDING INFORMATION

The SCORE platform was partly supported by EU4Health EpiCARE grants.

CONFLICT OF INTEREST STATEMENT

Sándor Beniczky led the international consensus process that developed the SCORE (Standardized Computer-based Organized Reporting of EEG) terminology and SCORE-EEG software (under the auspices of the International Federation of Clinical Neurophysiology, IFCN, and endorsed by the International League Against Epilepsy, ILAE). Harald Aurlien contributed to the software development of SCORE-EEG and is affiliated with Holberg EEG AS (Bergen, Norway), the company that distributes the SCORE-EEG software. Alexis Arzimanoglou received funding from the European Commission to manage ERN (European Reference Network) EpiCARE. The remaining authors have no conflicts of interest relevant to this research activity. 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

We hereby confirm that the data collected and analyzed are available on reasonable request.

ETHICS STATEMENT

The study protocol received approval from the ethics committees of all participating centers (for coordinating center: protocol number 0035772/22, EC Pavia, Italy). We confirm that we have read the Journal's position on ethical publication and affirm that this report is consistent with those guidelines.

ORCID

Elena Tartara https://orcid.org/0000-0003-1656-5420
Harald Aurlien https://orcid.org/0000-0001-6205-6673
Melita Čačić Hribljan https://orcid.org/0000-0002-5400-1805
Francesca Bisulli https://orcid.org/0000-0002-1109-7296
Laura Canafoglia https://orcid.org/0000-0002-5385-761X
Nicolas Gaspard https://orcid.org/0000-0003-1148-6723
Carla Bentes https://orcid.org/0000-0003-2399-7678
Rūta Mameniškienė https://orcid.org/0000-0001-5719-8175
Elena Ballante https://orcid.org/0000-0003-3217-8930
Alessandro Amatu https://orcid.org/0000-0003-4063-9493
Maria Vlachou https://orcid.org/0000-0001-5786-6954
Lorenzo Ferri https://orcid.org/0000-0003-0148-9795
Elena Pasini https://orcid.org/0000-0001-6298-455X
Chiara Pastori https://orcid.org/0000-0002-9121-5087
Andrea Stabile https://orcid.org/0000-0003-0999-215X
Charlotte De Maeseneire https://orcid.org/0009-0008-6659-6640
Lorenzo Ferlini https://orcid.org/0000-0002-6461-8796
Ana Catarina Franco https://orcid.org/0000-0002-4371-9272
Ana Rita Peralta https://orcid.org/0000-0001-9449-431X
Hanna-Reetta Lajunen https://orcid.org/0009-0004-0661-3075
Alessandro Principe https://orcid.org/0000-0002-3694-0892
Alexis Arzimanoglou https://orcid.org/0000-0002-7233-2771
Sándor Beniczky https://orcid.org/0000-0002-6035-6581

REFERENCES

1. Chong DJ, Hirsch LJ. Which EEG patterns warrant treatment in the critically ill? Reviewing the evidence for treatment of periodic epileptiform discharges and related patterns. J Clin Neurophysiol. 2005;22(2):79–91.
2. Gelisse P, Crespel A, Genton P, Jallon P, Kaplan PW. Lateralized periodic discharges: which patterns are interictal, ictal, or peri-ictal? Clin Neurophysiol. 2021;132(7):1593–603. https://doi.org/10.1016/j.clinph.2021.04.003
3. Ruiz AR, Vlachy J, Lee JW, Gilmore EJ, Ayer T, Haider HA, et al. Association of periodic and rhythmic electroencephalographic patterns with seizures in critically ill patients. JAMA Neurol. 2017;74(2):181–8. https://doi.org/10.1001/jamaneurol.2016.4990
4. Snider SB, Fong MWK, Nolan NM, Ruiz AR, Wang W, Laroche S, et al. Clinical and electroencephalographic predictors of seizures and status epilepticus in 12,450 critically ill adults: a retrospective cohort study. Crit Care Med. 2023;51(8):1001–11. https://doi.org/10.1097/CCM.0000000000005872
5. Bosque-Varela P, Machegger L, Lauth W, Tsalouchidou PE, Knake S, Zimmermann G, et al. Ictal–interictal continuum and status epilepticus: two sides of the same coin? A prospective magnetic resonance imaging study. Epilepsia. 2026;67:2377–88. https://doi.org/10.1002/epi.70131
6. Johnson EL, Kaplan PW. Population of the ictal-interictal zone: The significance of periodic and rhythmic activity. Clin Neurophysiol Pract. Volume 2. Elsevier B.V.; 2017. p. 107–18. https://doi.org/10.1016/j.cnp.2017.05.001
7. Hirsch LJ, Fong MWK, Leitinger M, LaRoche SM, Beniczky S, Abend NS, et al. American clinical neurophysiology Society's standardized critical care EEG terminology: 2021 version. J Clin Neurophysiol Off Publ Am Electroencephalogr Soc. 2021;38(1):1–29. https://doi.org/10.1097/WNP.0000000000000806

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

8. Hirsch LJ, Laroche SM, Gaspard N, Gerard E, Svoronos A, Herman ST, et al. American clinical neurophysiology society's standardized critical care EEG terminology: 2012 version. J Clin Neurophysiol. 2013;30(1):1–27. https://doi.org/10.1097/WNP.0b013e3182784729
9. Gerber PA, Chapman KE, Chung SS, Drees C, Maganti RK, Ng Y-T, et al. Interobserver agreement in the interpretation of EEG patterns in critically ill adults. J Clin Neurophysiol. 2008;25:241–9.
10. Ronner HE, Ponten SC, Stam CJ, Uitdehaag BMJ. Inter-observer variability of the EEG diagnosis of seizures in comatose patients. Seizure. 2009;18(4):257–63. https://doi.org/10.1016/j.seizure.2008.10.010
11. Gaspard N, Hirsch LJ, LaRoche SM, Hahn CD, Brandon M. Interrater agreement for critical care EEG terminology. Epilepsia. 2014;55(9):1366–73. https://doi.org/10.1111/epi.12653
12. Mani R, Arif H, Hirsch LJ, Gerard EE, LaRoche SM. Interrater reliability of ICU EEG research terminology. J Clin Neurophysiol. 2012;29(3):203–12. https://doi.org/10.1097/WNP.0b013e3182570f83
13. Westhall E, Rosén I, Rossetti AO, van Rootselaar AF, Wesenberg Kjaer T, Friberg H, et al. Interrater variability of EEG interpretation in comatose cardiac arrest patients. Clin Neurophysiol. 2015;126(12):2397–404. https://doi.org/10.1016/j.clinph.2015.03.017
14. Jing J, Ge W, Struck AF, Fernandes MB, Hong S, An S, et al. Interrater reliability of expert electroencephalographers identifying seizures and rhythmic and periodic patterns in EEGs. Neurology. 2023;100(17):E1737–E1749. https://doi.org/10.1212/WNL.0000000000201670
15. Foreman B, Mahulikar A, Tadi P, Claassen J, Szaflarski J, Halford JJ, et al. Generalized periodic discharges and "triphasic waves": a blinded evaluation of inter-rater agreement and clinical significance. Clin Neurophysiol. 2016;127(2):1073–80. https://doi.org/10.1016/j.clinph.2015.07.018
16. Halford JJ, Shiau D, Desrochers JA, Kolls BJ, Dean BC, Waters CG, et al. Inter-rater agreement on identification of electrographic seizures and periodic discharges in ICU EEG recordings. Clin Neurophysiol. 2015;126(9):1661–9. https://doi.org/10.1016/j.clinph.2014.11.008
17. Beniczky S, Aurlien H, Brøgger JC, Fuglsang-Frederiksen A, Martins-Da-Silva A, Trinka E, et al. Standardized computer-based organized reporting of EEG: SCORE. Epilepsia. 2013;54(6):1112–24. https://doi.org/10.1111/epi.12135
18. Beniczky S, Aurlien H, Brøgger JC, Hirsch LJ, Schomer DL, Trinka E, et al. Standardized computer-based organized reporting of EEG: SCORE—second version. Clin Neurophysiol. 2017;128(11):2334–46. https://doi.org/10.1016/j.clinph.2017.07.418
19. Kane N, Acharya J, Benickzy S, Caboclo L, Finnigan S, Kaplan PW, et al. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. Clin Neurophysiol Pract. 2017;2:170–85. https://doi.org/10.1016/j.cnp.2017.07.002
20. Japaridze G, Kasradze S, Aurlien H, Beniczky S. Implementing the SCORE system improves the quality of clinical EEG reading. Clin Neurophysiol Pract. 2022;7:260–3. https://doi.org/10.1016/j.cnp.2022.07.004
21. The SCORE Standard | Holberg EEG; n.d. Available from: https://www.holbergeeg.com/the-score-standard. Accessed 20 June 2026.
22. Westover AM, Westover TM, Westover MB. Interrater reliability estimation via maximum likelihood for Gwet's chance agreement model. Open J Stat. 2024;14(5):481–91. https://doi.org/10.4236/ojs.2024.145021
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159. https://doi.org/10.2307/2529310
24. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. J Clin Epidemiol. 2012;65(7):778–84. https://doi.org/10.1016/j.jclinepi.2011.10.019
25. Tesh RA, Zahoor A, Banks J, Gallagher K, Eckhardt CA, Sun H, et al. Inter-rater reliability of EEG-based encephalopathy grading. J Clin Neurophysiol. 2026;43(3):237–44. https://doi.org/10.1097/WNP.0000000000001185
26. Jing J, Herlopian A, Karakis I, Ng M, Halford JJ, Lam A, et al. Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms. JAMA Neurol. 2020;77(1):49–57. https://doi.org/10.1001/jamaneurol.2019.3531
27. Kural MA, Duez L, Sejer Hansen V, Larsson PG, Rampp S, Schulz R, et al. Criteria for defining interictal epileptiform discharges in EEG: a clinical validation study. Neurology. 2020;94(20):E2139–E2147. https://doi.org/10.1212/WNL.0000000000009439
28. Isik L, Meyers EM, Leibo JZ, Poggio T. The dynamics of invariant object recognition in the human visual system. J Neurophysiol. 2014;111:91–102. https://doi.org/10.1152/jn.00394.2013
29. Chatrian GE, Shaw CM, Leffman H. The significance of periodic lateralized epileptiform discharges in EEG: an electrographic, clinical and pathological study. Electroencephalogr Clin Neurophysiol. 1964;17(2):177–93. https://doi.org/10.1016/0013-4694(64)90149-X
30. Gaspard N, Manganas L, Rampal N, Petroff OAC, Hirsch LJ. Similarity of lateralized rhythmic delta activity to periodic lateralized epileptiform discharges in critically ill patients. JAMA Neurol. 2013;70(10):1288–95. https://doi.org/10.1001/jamaneurol.2013.3475

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: Tartara E, Aurlien H, Hribljan MČ, Bisulli F, Canafoglia L, Gaspard N, et al. Inter-rater reliability on rhythmic and periodic EEG patterns using Standardized Computer-based Organized Reporting of EEG (SCORE): A multicenter study. Epilepsia. 2026;00:1–16. https://doi.org/10.1002/epi.70488

15281167, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/epi.70488 by Alexis Arzimanoglou - Spanish Cochrane National Provision (Ministerio de Sanidad) , Wiley Online Library on [27/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License