



HAL
open science

The challenge of assessing invasive biomarkers for epilepsy surgery

Nicolas Roehri, Serge Vulliemoz, Stanislas Lagarde

► **To cite this version:**

Nicolas Roehri, Serge Vulliemoz, Stanislas Lagarde. The challenge of assessing invasive biomarkers for epilepsy surgery. *Brain - A Journal of Neurology*, 2024, 147 (8), pp.e52-e54. 10.1093/brain/awae164 . hal-04913246

HAL Id: hal-04913246

<https://hal.science/hal-04913246v1>

Submitted on 27 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/380636481>

The challenge of assessing invasive biomarkers for epilepsy surgery

Article in *Brain* · May 2024

DOI: 10.1093/brain/awae164

CITATION

1

READS

147

3 authors:



[Nicolas Roehri](#)

University of Geneva

66 PUBLICATIONS 1,412 CITATIONS

[SEE PROFILE](#)



[Serge Vulliemoz](#)

Hôpitaux Universitaires de Genève

185 PUBLICATIONS 5,411 CITATIONS

[SEE PROFILE](#)



[Stanislas Lagarde](#)

Assistance Publique Hôpitaux de Marseille

177 PUBLICATIONS 3,466 CITATIONS

[SEE PROFILE](#)

1 LETTER TO THE EDITOR

2 **The challenge of assessing invasive biomarkers for epilepsy** 3 **surgery**

4 Nicolas Roehri,¹ Serge Vulliémoz¹ and Stanislas Lagarde^{1,2,3}

5 **Author affiliations:**

6 1 EEG and Epilepsy Unit, University Hospitals and Faculty of Medicine, University of Geneva,
7 CH-1211, Geneva, Switzerland

8 2 APHM, Timone Hospital, Epileptology and Cerebral Rhythmology, 13005, Marseille, France

9 3 Aix Marseille Univ, INSERM, INS, Inst Neurosci Syst, 13005, Marseille, France

10
11 Correspondence to: Nicolas Roehri

12 EEG and Epilepsy Unit, Division of Neurology, HUG, 4 rue Gabrielle-Perret-Gentil, CH-1211,
13 Genève, Switzerland

14 E-mail: Nicolas.Roehri@unige.ch

15
16 We read with interest the article that has recently been published by Brain entitled “Spike ripples
17 localize the epileptogenic zone best: an international intracranial study” by Shi *et al.*¹ This study
18 examined a substantial dataset comprising over 100 patients from various centres, using an
19 automated algorithm designed to detect interictal spike and ripple events, collectively referred to
20 as 'spike ripple,' in intracranial EEG recordings. The aim of this study is to investigate whether this
21 biomarker outperforms other existing biomarkers in accurately localising the epileptogenic zone.
22 Notably, the authors have integrated data from both spikes and ripples, aligning with our prior
23 study's recommendation, which underscored the complementary nature of spikes and high-
24 frequency oscillations (HFOs)². Such studies are crucial in the search for the most effective
25 interictal biomarker for epileptogenicity.

1 We have, however, some concerns about how this manuscript presents non-statistically significant
2 results, as it could potentially lead to misinterpretation of the study's findings. Additionally, the
3 rate ratio metric and certain statistical analyses used appear inappropriate, which could introduce
4 bias into certain results. We would like to raise the following specific methodological issues.

5
6 **First**, in the abstract and corresponding section of the results, the authors present some findings as
7 evidence despite their lack of statistical significance.

8
9 *“Subjects with curative resection (ILAE 1) had a higher proportion of spike ripple generating brain
10 tissue removed compared to those who were not seizure free (ILAE 2-6; $P = 0.06$, $d = 0.17$).”*

11
12 *“The percentage of ILAE 1 subjects with the majority of spike ripples removed was higher than the
13 percentage of subjects with the majority of spikes (69%, $P = 0.12$), spike-gamma (69%, $P = 0.12$),
14 wideband HFOs (63%, $P = 0.03$), ripples (45%, $P = 0.01$), or fast ripples (36%, $P < 0.001$)
15 removed.”*

16
17 These excerpts illustrate the authors' claims. They first claim that a higher proportion of spike
18 ripples removed in patients with favourable outcomes compared to those with poor outcomes,
19 despite not being statistically significant ($P=0.06$) and a low effect-size ($d=0.17$, equivalent to an
20 area under the curve of 0.585). Their second claim is that a higher proportion of patients with the
21 majority of spike ripples removed compared to spikes or spike-gamma, again without statistically
22 significant evidence ($P=0.12$). We acknowledge that P-values and null-hypothesis significance
23 testing have limitations due to their overly dichotomous nature. However, stating that there is an
24 effect with such P-values and low effect size is clearly misleading. A more accurate conclusion
25 would have acknowledged the absence of statistically significant differences between patients with
26 favourable and unfavourable outcomes regarding the proportion of spike ripple-generating brain
27 tissue removed. Additionally, it should have explicitly stated that the proportion of ILAE 1 subjects
28 with the majority of spike ripples removed was not statistically higher than for spikes and spike-
29 gamma. Furthermore, none of the aforementioned tests related to the percentage of ILAE 1 subjects

1 with the majority of spike ripples removed have been corrected for multiple comparisons, a step
2 that should have been taken given that the results of the spike ripple biomarker have been compared
3 to all the other biomarkers. For instance, applying Bonferroni or FDR correction would render the
4 comparison between spike ripple and wideband HFOs not significant.

5
6 **Secondly**, the event rate ratio (RR) used here to quantify the different biomarkers, while common
7 in the literature, is debatable. In fact, it solely evaluates specificity to the resected volume, lacks
8 clinical applicability and sometimes produces counterintuitive results, which contradict those
9 obtained through other commonly used metrics from the binary classification field. Figure 1
10 illustrates these points with six toy examples. It also shows the classification performance with and
11 without considering the imbalance between resected and non-resected channels.

12
13 An ideal biomarker is not only specific but also sensitive^{3,4}. The hope of finding a biomarker in the
14 current context is either to better delineate the epileptogenic zone (EZ) to improve surgical outcome
15 or to predict the surgical outcome. A biomarker with perfect specificity but poor sensitivity would
16 only delineate a portion of the EZ, and removing partially the EZ would not lead to seizure freedom.
17 For instance, if spike ripples are found only in one-third of the electrodes implanted in the EZ and
18 nowhere else, its sensitivity would be 30% but its RR would equal 1 (Figure 1 A, Case 1 vs Case
19 2). This may account for the lack of statistical difference in the RR of the spike ripples between
20 patients with favourable and unfavourable outcomes. It is plausible that the most epileptogenic
21 areas, i.e., with the highest event rates, were removed in the latter group, in line with the intention-
22 to-treat principle. Alternatively, the authors' hypothesis posits that a high RR in patients with poor
23 outcome could arise from incomplete coverage of the EZ (Figure 1 D of the original article). If this
24 were true, any invasive biomarker would be inefficient in such scenarios, as acknowledged by the
25 authors. Moreover, clinicians had sufficient other information from both non-invasive and invasive
26 investigations to propose surgery, thereby lowering the likelihood of this hypothesis in most of the
27 patients.

28
29 Even if a biomarker were able to differentiate between the two patient groups based on its RR, it
30 remains unclear how the RR would influence the clinical decision making. A biomarker with an

1 RR close to zero or negative could indicate the need to enlarge the resection, only if the event rate
2 is high outside the resected volume (Figure 1 A, Case 3 and 4). However, it may not aid in
3 delineation or ensure seizure freedom, as its sensitivity has not been evaluated. In other words,
4 removing the remaining area with high event rates may not suffice, as this biomarker may not be
5 present in all regions of the EZ. This risk of high specificity but low sensitivity was already
6 acknowledge for fast ripples^{2,5,6}, which greatly limited their clinical use. It is possible that removing
7 the majority of spike ripples is necessary but not sufficient to guarantee seizure freedom, again
8 mirroring earlier research on fast ripples⁷. In some cases, the RR could also be negative, even if all
9 channels with the highest rates are within the resected volume (i.e., perfect separation resected vs
10 non-resected channels, balanced accuracy equals 1). This may occur due to the widespread
11 distribution of events (e.g., epileptic spikes) with high rates within the resected volume but lower
12 rates outside the resected volume (Figure 1 A, Case 5). The RR is thus negatively biased by the
13 extent of the event, could underestimate its performance, and could contradict other binary
14 classification metrics.

15
16 **Finally**, the authors seem to have employed statistical tests designed for independent groups on
17 dependent groups. The methods section clearly states that spike ripples, spikes, spike-gamma and
18 HFO were “detected on all data from all sites” (sites A, B, C and D).

19
20 *Right-tailed two-proportion z-test* is used to determine if the proportions of categories in two group
21 variables significantly differ from each other. However, this test requires that the two groups are
22 independent, yet in this article, the proportions being tested come from the same group of patients
23 but are characterised by different biomarkers, thus violating the assumption of independence. We
24 believe that the McNemar’s test would have been more suitable, as it is designed to compare
25 proportions between two dependent populations with paired samples. Similarly, while *Cliff’s d-*
26 *value* quantifies effect size related to the Mann-Whitney U test or Wilcoxon rank-sum test for
27 independent populations, it is inadequate for comparing different biomarkers in patients with
28 favourable outcomes. Cohen’s d effect size calculated for paired sample (if the assumption of
29 normality is met) would have been more adequate. With appropriate tests, the previously non-

1 significant results could have potentially reached significance, as they would have been able to
2 quantify more subtle increases in performance.

3
4 In conclusion, due to the previously discussed limitations regarding the statistical methodology, it
5 appears challenging to fully support the authors' claims regarding the differences in spike ripples
6 between patients with favourable and unfavourable outcomes, and their assertion that spike ripples
7 are a superior biomarker compared to spikes, spike gamma, and HFOs. The study certainly adds
8 valuable information to this important field of research but presenting results as evidence despite
9 their lack of statistical significance and low effect size is misleading, particularly as these findings
10 may linger and influence future studies. This could be detrimental to the research domain and,
11 ultimately, the clinical practice. Even more crucially, the examples depicted in Figure 1 underscore
12 the challenge of identifying an optimal metric for assessing the performance of biomarkers for
13 epilepsy surgery. It is essential to emphasise the importance of utilising metrics that account for
14 both sensitivity and specificity/precision, as well as the inherent imbalance between positive (i.e.,
15 small EZ) and negative classes (i.e., larger non-EZ), which could be patient dependent. Further
16 research and discussion on quantifying the performance of biomarkers will be pivotal for advancing
17 this research field and enhancing its clinical applicability.

18

19 **Data availability**

20 The code used to produce Figure 1 is available here
21 [https://gitlab.unige.ch/vulliemoz_group/toy_examples/-](https://gitlab.unige.ch/vulliemoz_group/toy_examples/-/blob/main/Event_Rate_Ratio_vs_Accuracy)
22 [/blob/main/Event_Rate_Ratio_vs_Accuracy](https://gitlab.unige.ch/vulliemoz_group/toy_examples/-/blob/main/Event_Rate_Ratio_vs_Accuracy).

23

24 **Competing interests**

25 The authors report no competing interests.

26

1 Funding

2 No funding was received towards this work.

3 References

- 4 1. Shi W, Shaw D, Walsh KG, et al. Spike ripples localize the epileptogenic zone best: an
5 international intracranial study. *Brain*. Published online February 7, 2024.
6 doi:10.1093/brain/awae037
- 7 2. Roehri N, Pizzo F, Lagarde S, et al. High-frequency oscillations are not better biomarkers
8 of epileptogenic tissues than spikes. *Ann Neurol*. 2018;83(1):84-97. doi:10.1002/ana.25124
- 9 3. Engel J, Pitkanen A, Loeb JA, et al. Epilepsy biomarkers. *Epilepsia*. 2013;54(SUPPL.4):61-
10 69. doi:10.1111/epi.12299
- 11 4. Pitkänen A, Ekolle Ndode-Ekane X, Lapinlampi N, Puhakka N. Epilepsy biomarkers –
12 Toward etiology and pathology specificity. *Neurobiol Dis*. 2019;123(2017):42-58.
13 doi:10.1016/j.nbd.2018.05.007
- 14 5. Roehri N, Bartolomei F. Are high-frequency oscillations better biomarkers of the
15 epileptogenic zone than spikes? *Curr Opin Neurol*. 2019;32(2):213-219.
16 doi:10.1097/WCO.0000000000000663
- 17 6. Cuello-Oderiz C, von Ellenrieder N, Sankhe R, et al. Value of ictal and interictal
18 epileptiform discharges and high frequency oscillations for delineating the epileptogenic
19 zone in patients with focal cortical dysplasia. *Clin Neurophysiol*. 2018;129(6):1311-1319.
20 doi:10.1016/j.clinph.2018.02.003
- 21 7. van 't Klooster MA, van Klink NEC, Zweiphenning WJEM, et al. Tailoring epilepsy surgery
22 with fast ripples in the intraoperative electrocorticogram. *Ann Neurol*. 2017;81(5):664-676.
23 doi:10.1002/ana.24928
- 24 8. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental
25 evaluation tool in clinical medicine. *Clin Chem*. 1993;39(4):561-577. doi:ROC; Receiver-
26 Operating Characteristic; SDT; Signal Detection Theory
- 27

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Figure legend

Figure 1 Visualisation of six toy examples used to compare the event rate ratio and balanced accuracy. For each case, dots represent intracranial channels (visually arranged on a 2D sheet for clarity), colour-coded by their event rates. The resected volume (RV) is outlined by a red circle (panel A). Bar plots display the event rate ratio (RR) and balanced accuracy (bACC) for each case (panels B and C, respectively). Sensitivity and specificity for each case are depicted in the receiver operating characteristic (ROC) space (panel D). Sensitivity represents the percentage of correctly predicted resected channels, while specificity represents the percentage of correctly predicted non-resected channels. bACC, the average of sensitivity and specificity, is calculated considering two optimisation approaches (opti. in panel C): optimisation 1 does not account for the proportion of resected over non-resected channels, while optimisation 2 does⁸. Cases 1 and 2 exhibit high event rates within the resected volume and zero outside, with Case 2 having a larger RV than Case 1. Despite both Cases having a rate ratio and specificity values of 1, the bACC and sensitivity is lower in Case 2. Cases 3 and 4 show scenarios where the RR is negative, because either the events are present at the border of the RV or there are two event generating zones, one inside and a stronger one outside the RV. The event distribution would suggest the need for resection enlargement in patients with poor outcomes. In such situations, bACC, sensitivity and specificity vary depending on optimisation approaches. Optimisation 1 clearly overestimates the performance of the metric, emphasizing the importance of considering the proportion of resected over non-resected channels. In Case 5, the RR is negative but the bACC, sensitivity and specificity values equal 1. This means that the RR is negatively biased by the extension of the event even though the RV contains the channels with highest rates. Case 6 is similar to Case 5, except that the RV is larger making the RR positive.

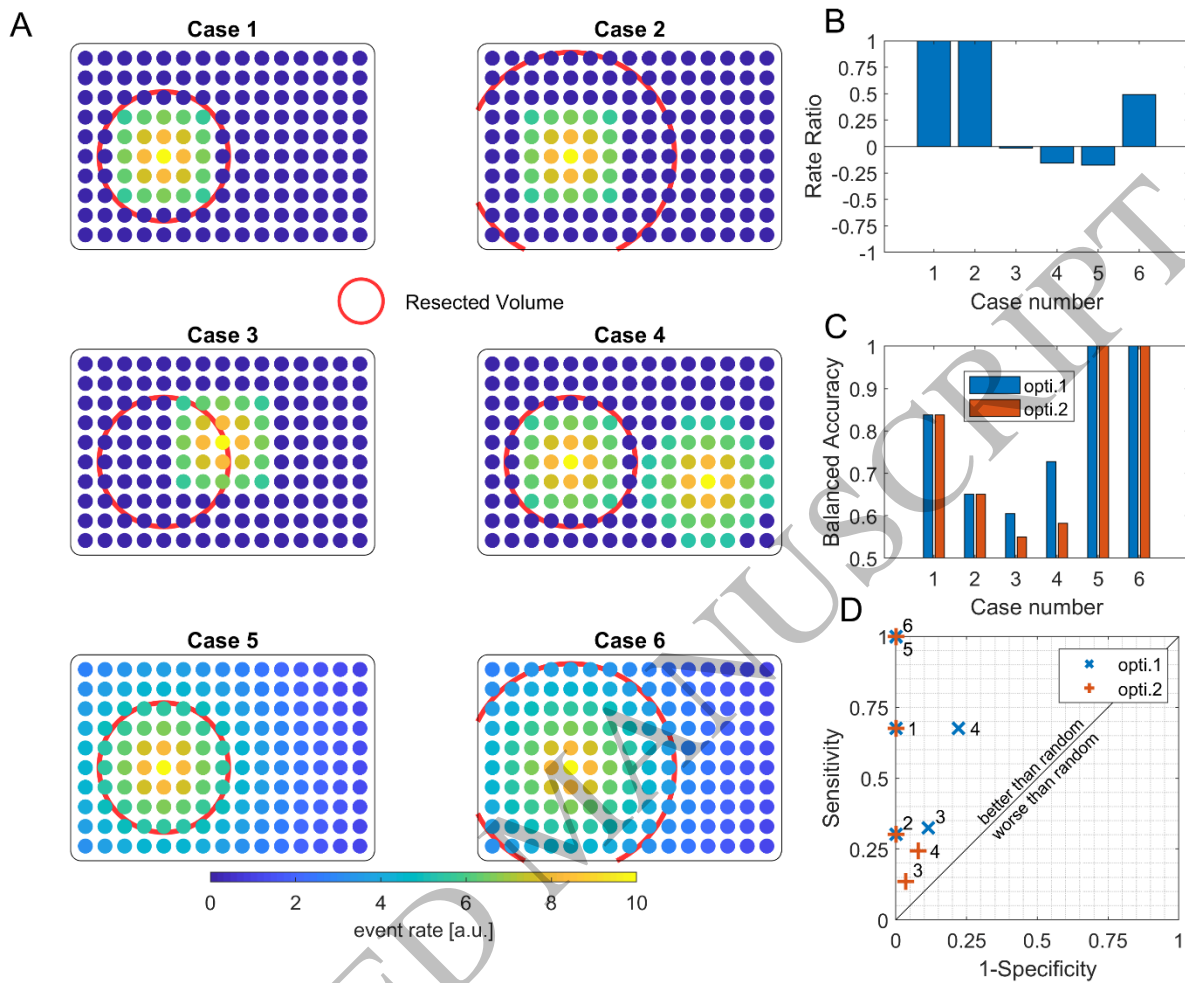


Figure 1
160x132 mm (x DPI)

1
2
3