Prediction of the clinical and naming status after anterior temporal lobe resection in patients with epilepsy

Elise Roger [a,*], Laurent Torlay [a], Sonja Banjac [a], Chrystèle Mosca [b], Lorella Minotti [b], Philippe Kahane [b], Monica Baciu [a]

[a] Univ. Grenoble Alpes, CNRS LPNC UMR 5105, 38000 Grenoble, France
[b] Univ. Grenoble Alpes, Grenoble Institute of Neuroscience, Synchronisation et modulation des réseaux neuronaux dans l'épilepsie' & Neurology Department, 38000 Grenoble, France

ABSTRACT

By assessing the cognitive capital, neuropsychological evaluation (NPE) plays a vital role in the perioperative workup of patients with refractory focal epilepsy. In this retrospective study, we used cutting-edge statistical approaches to examine a group of 47 patients with refractory temporal lobe epilepsy (TLE), who underwent standard anterior temporal lobectomy (ATL). Our objective was to determine whether NPE may represent a robust predictor of the postoperative status, two years after surgery. Specifically, based on pre- and postsurgical neuropsychological data, we estimated the sensitivity of cognitive indicators to predict and to disentangle phenotypes associated with more or less favorable outcomes. Engel (ENG) scores were used to assess clinical outcome, and picture naming (NAM) performance to estimate naming status. Two methods were applied: (a) machine learning (ML) to explore cognitive sensitivity to postoperative outcomes; and (b) graph theory (GT) to assess network properties reflecting favorable vs. less favorable phenotypes after surgery. Specific neuropsychological indices assessing language, memory, and executive functions can globally predict outcomes. Interestingly, preoperative cognitive networks associated with poor postsurgical outcome already exhibit an atypical, highly modular and less densely interconnected configuration. We provide statistical and clinical tools to anticipate the condition after surgery and achieve a more personalized clinical management. Our results also shed light on possible mechanisms put in place for cognitive adaptation after acute injury of central nervous system in relation with surgery.

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The central role of neuropsychology in epilepsy has been historically rooted, notably in the context of neurosurgery for refractory epilepsy. Major discoveries such as Penfield's homunculus [1], the functional specialization of the hippocampal–temporal lobe complex in memory [2,3], or the functional lateralization of certain cognitive functions as observed via the Wada test [4] or surgical callosotomy and "split brain" patients [5] have been documented in this particular context [6]. Neuropsychology and neurosurgery have thus been – and still remain – reciprocally beneficial in understanding of neurocognitive mechanisms [7].

Cognitive impairment as assessed by the neuropsychological assessment is, in fact, the most common comorbidity in patients with epilepsy, whether generalized or focal [8]. It is now

well-established that focal epilepsy is a systemic neurological pathology disrupting large scale functional and structural brain networks which is accompanied by long-term persistent cognitive symptoms (e.g., [9–12] for recent studies). The degree of cognitive impairment reportedly worsens over time, suggesting that chronic epileptic insult (i.e., the recurrence of seizures) causes it, as evidenced by the term "epileptic dementia" introduced in the early 19th century to characterize progressive cognitive decline [13]. Cognitive impairments are actually occurring very early. Frequently, they are already detected by the time of epilepsy's onset. It is estimated that 70% of adult patients with newly diagnosed epilepsy have at least one proven cognitive deficit well before the introduction of the medication [14].

Despite the reported progressive deterioration, the intensity of cognitive symptoms generally remains mild to moderate in focal epilepsy [15]. Regarding the most common form of focal epilepsy in adults, the temporal lobe epilepsy (TLE), interictal lasting language and memory deficits of varying degrees of severity, are commonly noticed [16–19]. The temporal lobe territories are indeed

known to support language and memory networks (Duff & Brown-Schmidt [20] even propose that this is the "language-and-memory interface"; see also [21], which may explain the vulnerability of these functions in TLE). However, in addition to language and memory, the temporal lobe is also involved in large-scale functional circuitry and other cognitive processes, potentially inducing various cognitive deficits (executive functioning, social cognition, or face recognition [22,23]). Recent studies have, in this respect, stressed a variety of cognitive phenotypes in patients with TLE [24,25], moderating the conjunction of language-memory deficits traditionally associated with patients with TLE. Undoubtedly, there is no single cognitive profile associated with TLE, but a panel of phenotypes whose expression depends on many factors [26].

Neuropsychological evaluation (NPE) is an essential tool for clinical diagnosis and management of patients with epilepsy presenting with suboptimal cognitive functioning. Routine objective screening is indeed essential considering that there is a substantial underestimation of cognitive deficits in epilepsy when their assessment is based solely on subjective complaints (this relates to language, memory, and executive performance; [14] for an estimate). As an indicator of cognitive status before invasive neurosurgery, NPE can also help to detect, localize, and lateralize brain dysfunctions and to formulate hypotheses about the epileptogenic networks involved. Consistent with other studies [27,28], we have indeed previously demonstrated the usefulness of NPE in diagnosing epilepsy localization and the demonstrated sensitivity of certain neuropsychological indices to epilepsy lateralization and localization [29]. Even more, NPE performed in epilepsy surgery as an objective control of preoperative cognitive status can be used to prognosticate postsurgical cognitive and functional outcome, allowing for timely initiation of tailored cognitive remediation if necessary [30].

However, although several studies have sought to identify clinical and cognitive factors that are useful in predicting patients' postoperative status ([31,32], for two recent studies on large patient samples), there is still a lack of benchmark about the most relevant cognitive indices (i.e., in terms of sensitivity, reliability, and validity) for predicting postoperative outcomes. The main objective of this study was therefore to provide information regarding this important consideration.

To this end, we first identified relevant presurgical neuropsychological predictors of (1) clinical and (2) naming long-term outcomes after temporal lobe resection (2 years after anterior temporal lobectomy; ATL) in patients with drug-resistant TLE. Clinical outcome was determined using the Engel (ENG) score [33,34] which quantifies the success of neurosurgery in the context of epilepsy [35]. Concerning the postoperative naming outcome, we focused on picture naming ability assessed by the DO80 [36] (a French equivalent of the Boston Naming Test: BNT [37]). Despite some psychometric limitations (mentioned in the discussion section), this task presents substantial clinical advantages. Firstly, lexical access involved in object naming is a major determinant for the quality of life and the return to work activities of patients, after surgery (e.g., [38] in patients with low-grade glioma). Secondly, the naming test and picture naming in particular is a gold standard [39], widely used in temporal epilepsy [17]. Lastly, ATL can produce cognitive "side effects" and naming deficits in particular were shown to persist (persistent dysnomia can be noted in up to 60% of cases depending on the study [40], for a systematic review; in one third of left temporal patients [41], for a review using weighted average estimate of naming decline). For these reasons and in this context, it is one of the most important cognitive indicators to predict.

We performed state-of-the-art and data-driven analyses. We used machine learning (ML) algorithms to identify latent relationships of our predictors and make predictions regarding selected outcomes. We applied model-agnostic methods to facilitate the interpretation of ML results (black box insight [42]) and cognitive phenotypes. Finally, and as a complementary approach to identify factors related to the different outcomes, we modeled the presurgical neuropsychological networks of patients whose surgery was effective versus less effective. On these networks we performed graph theory (GT) analyses to define phenotypes and characterize fundamental differences regarding the preoperative neuropsychological architecture according to postoperative outcome. Graph theory applied to neuropsychological scores indeed offer an optimal viewing angle, adapted to the current vision of the cognitive system as a scaffold of interactive links between cognitive domains [43–45].

2. Material

2.1. Population

Forty-seven patients with unilateral and drug-resistant TLE diagnosed in accordance with the clinical criteria described in ILAE committee report [46] were included in the study. Patient inclusion criteria were as follows: (1) older than 18 years of age; (2) had undergone standard ATL resection; (3) had complete data for all variables of interest. Patients with neurological comorbidities such as stroke, tumor, or neurodegenerative disease and/or a prior neurosurgery were systematically excluded.

2.2. Clinical and neuropsychological evaluations

Presurgical assessments including a comprehensive neurological examination and a NPE for each patient were performed between 2014 and 2019, with the same protocol. Neuropsychological evaluations were conducted by a specialized neuropsychologist in the epilepsy unit of the Grenoble Alpes University Hospital. Clinical information and performance on 32 neuropsychological indices providing a comprehensive overview of cognitive functioning before neurosurgery were collected (Table S1 includes all collected data; see Appendix S1 for details and explanations about clinical and neuropsychological variables). A postoperative evaluation was also systematically performed at two years postoperatively (M = 2.06 years after neurosurgery; SD = 0.22). Two main postoperative factors: **(ENG)** the Engel score and **(NAM)** the postoperative object naming score accounting for the baseline (i.e., the preoperative naming performance) – were selected to predict long-term neurosurgery outcomes. We determined two respective classes for each of these postoperative factors, namely:

- (**ENG+**) **optimal postoperative clinical outcome**: the Engel score of I, no seizures observed 2 years after surgery;
- (**ENG−**) **suboptimal postoperative clinical outcome**: the Engel score of II-IV: persistence of more or less frequent and more or less severe seizures 2 years after surgery;
- **(NAM+) optimal postoperative naming outcome**: the naming score at 2 years after surgery did not decline or even improved compared to the presurgical baseline;
- **(NAM−) suboptimal postoperative naming outcome**: the naming score at 2 years after surgery has decreased compared to the presurgical baseline.

Note that to define if the change of individual NAM performance was meaningful, we calculated a Reliable Change Index (RCI) according to Chelune's methodology ([47] and see also [48] for a description) and based on the DO80 naming scores. RCIs below −1.28 (α = 0.1 or 90% one-sided confidence interval) were

considered as a reliable NAM decline (NAM−). Otherwise, patients' performance was considered to be without significant NAM decline (and therefore categorized in the NAM**+** condition).

Patients belonging to the different classes (ENG−/ENG+ and NAM−/NAM+) were matched by age, education, duration of epilepsy, number of AEDs, age at surgery, and distance of postoperative evaluations from surgery date (Table S2 for statistical reports). Table 1 below summarizes the proportion of patients associated with the different classes based on the main and binarized clinical variables.

3. Method

We mirrored two types of complementary methods to address both the postoperative clinical outcome (ENG) and the postoperative naming outcome (NAM): a predictive machine learning approach (ML), followed by a network analysis using GT. A graphical and general outline of the method is presented below (Fig. 1). Codes and data are openly available: https://github.com/ltor-lpnc/neuropsy-2021.git, for the online repository.

3.1. Predictive ML approach

We used a similar approach we previously employed [29] to estimate the ability of neuropsychological scores to predict the postoperative status and by identifying the most discriminating scores and their interactions. To this end, we developed several Machine Learning workflows trying each time three different algorithms belonging to a different family: (a) a classical Support Vector Machine (SVM) algorithm [49] with a Radial Basis Function (RBF), (b) XGBoost algorithm [50] and (c) a logistic regression with penalty L2. To avoid sub-optimal results and because we had no reason to privilege a particular algorithm, different algorithms were evaluated. Practically, a binary classification has been conducted on the full dataset ($n = 47$ patients; no missing values) with a repeated cross-validation (CV) scheme including a feature selection step in each training set. The prediction performance of the most relevant and stable features was then measured once again with the same CV scheme. We then applied model-agnostic methods to get an insight about the features and their interactions in the prediction.

3.1.1. Binary classification

Our purpose was to predict two kinds of outcomes separately (ENG and NAM) with the help of neuropsychological indices. In order to make robust predictions, we restricted the number of features to 9 main neuropsychological indices (as previously used in Roger et al., 2020) to take into account: (1) the patients' sample size and (2) the curse of dimensionality [51]. We have preliminar-

ily standardized the 9 features of interest as required by some algorithms.

In practice, we trained a model to try to assign correctly a patient to its own group (ENG+/ENG− or NAM+/NAM−). We used a classical 5-fold CV scheme repeated 100 times. The CV was stratified, i.e., samples were randomly chosen in order to get always the same ratio of classes that existed in the original dataset. We chose the balanced accuracy (BAcc; [52]) as a measure of performance to quantify the quality of predictions. The BAcc has the advantage of dealing with imbalanced datasets and to offer an easy to use metrics in a clinical perspective. The BAcc formula is as follows:

$$BAcc = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right)$$

The best value is 100% and the worst 0%. BAcc is equal to the arithmetic mean of sensitivity (true positive rate) and specificity (true negative rate). For balanced datasets, the score is equal to classical accuracy.

3.1.2. Feature selection

In clinical perspective, it is interesting to simplify the model to the most relevant neuropsychological indices (i.e., to perform feature selection) in order to make easier and usable the interpretation of the ML approach. Penalized linear models are popular in order to get sparse solutions. Thus, we applied L2-norm on our dataset. Practically, feature selection was done in each training set with the default threshold implemented in scikit-learn v. 0.24 (i.e. mean of the features importance [53]). Selected features were used then to train the algorithm before measuring the performance with the held-out fold. This was repeated 500 times (5-fold cross validation repeated 100 times) to get a good estimate of feature stability and performance. In addition, we estimated the stability which is a very important aspect of results reproducibility. It is indeed crucial to know whether by randomly repeating the feature selection, we can trust the machine learning workflow and scientific conclusions. To this end and as metrics, we opted for the selected features' frequencies and the stability measure $\widehat{\Phi}$ proposed by Nogueira et al. [54].

3.1.3. Black box insight

Model-agnostic methods are unspecific and offer interpretability to any machine learning algorithm [55]. We chose Partial Dependence Plot (PDP) to see if the relationship between the target and a feature or two features is linear, monotonic or more complex. A PDP allows this by showing the marginal effect that one or two features have on the prediction. This algorithm builds the model by averaging the features, except the feature of interest F, and measures the changes in prediction for different values of F.

Considering the dataset reduced to the selected features, we performed a random partition (90% for the training set). We

Table 1
Summary of the proportion of patients represented in the subgroups according to the main clinical variables.

		TLEs	Gender		Hemisphere		Handedness		Sclerosis		Duration		Severity		Thymic disorders	
		(n)	(GEN)		(HEM)		(HDS)		(HS)		(DUR)		(SEV)		(THY)	
		47	Male	Female	LTLE	RTLE	Left	Right	Yes	No	<15years	>15years	Moderate	High	Yes	No
Clinical outcome	ENG+	30 (63.8)	15 (50)	15 (50)	14 46.7)	16 (53.3)	5 (16.7)	25 (83.3)	18 (60)	12 (40)	12 (40)	18 (60)	11 (36.7)	19 (63.3)	4 (13.3)	26 (86.7)
	ENG−	17 (36.2)	5 (29.5)	12 (70.5)	10 (58.8)	7 (41.2)	2 (11.7)	15 (88.3)	8 (47)	9 (53)	2 (11.8)	15 (88.2)	2 (11.8)	15 (88.2)	3 (17.6)	14 (82.4)
Cognitive outcome	NAM+	28 (59.6)	11 (39.3)	17 (60.7)	12 (42.8)	16 (57.2)	4 (14.3)	24 (85.7)	14 (50)	14 (50)	9 (32.1)	19 (67.9)	9 (32.1)	19 (67.9)	4 (14.3)	24 (85.7)
	NAM−	19 (40.4)	9 (47.4)	10 (52.6)	12 (63.2)	7 (36.8)	3 (15.8)	16 (84.2)	12 (63.2)	7 (36.8)	5 (26.3)	14 (73.7)	4 (21)	15 (79)	3 (15.8)	16 (84.2)

ML workflow

Binary classification

- 🎯 Identifying the predictive power of NPE indicators
- ↔ *Model performance: Balanced Accuracy (BAcc)*

Feature selection

- 🎯 Select the most sensitive NPE indicators
- ↔ *Model performance & stability of selected features*

Model interpretation

- 🎯 Interpreting how NPE indicators help in prediction
- ↔ *Black box insight: Nomograms & PDPs*

GT workflow

Patients networks

- 🎯 Measuring the proximity of patients' cognitive profiles in relation to clinical variables
- ↔ *Proximity measure, community detection & Simple Matching Coefficient (SMC)*

Cognitive networks

- 🎯 Detect relationships between neuropsychological scores in patients
- ↔ *Similarity measure, community detection, global (network level: modularity, density..) and local (node level: strength, clustering coefficient) parameters*

Clinical (ENG) & Naming (NAM) outcomes

Fig. 1. Overview of the methodology. Outline diagram of the method with main objectives and operational measures related to each workflow. ML = Machine Learning; GT = Graph Theory; NPE = Neuropsychological evaluation; ENG = Engel score at two years post-neurosurgery; NAM = Naming score at two years post-neurosurgery. The whole pipeline was applied for ENG and then for NAM with exactly the same methodology.

obtained ten plots: 4 1D plots (one by feature) and 6 2D plots (one by pair). Details and results about the PDPs are presented in supplementary material (Appendix S2).

In a similar way to the PDPs, nomograms are a pictorial representation of a model and provide complementary and useful information about the prediction. Clinical nomograms use variables or features to graphically depict a prognostic model that generates the probability of an event, in our case the clinical and the naming postoperative status. We developed clinical nomograms based on a logistic regression applied on the ML selected cognitive indices, using the rms R package. The quality of the validation was assessed by bootstrapping and measured by the area under the curve (AUC) of a receiver operating curve (ROC).

3.2. Networks and GT approach

Graph theory is a well-established theory using many mathematical methods to study networks. We present here a dual approach based on GT to take a broader view of the relationships in-between (1) patients (networks of individuals) and then (2) neuropsychological indices (cognitive networks). Graphs are defined as a set of nodes connected by edges which represent a relation between the connected nodes. Since we were interested in similarity/proximity relations between the nodes, our networks were composed of weighted edges that are not directed (undirected weighted graphs). The whole workflow was realized with iGraph library [56].

3.2.1. Patients' networks
3.2.1.1. General overview.
We first considered patients as nodes to identify similarities/dissimilarities between their neuropsychological profiles. The link between nodes (i.e., the edge) was obtained from the neuropsychological indices selected in the ML feature selection analyses. We therefore respectively constructed two independent networks: one based on the neuropsychological tests selected for the surgery outcome (ENG patients' network); and a second based on the neuropsychological tests selected for the naming outcome (NAM patients' network). Patients with similar characteristics tend to cluster together and community analyses applied on the patients' networks inform about the shared charac-

teristics by comparing the detected communities to the clinical data.

3.2.1.2. Graph construction.
As with other studies (e.g., [24]), a proximity measure (based on Euclidean distance) was used to construct the patients' networks. To avoid a chance that a variable, here a neuropsychological index, creates unrealistic greater inter-sample differences, we standardized the dataset. Then we implemented the proximity measure as:

$$\text{Proximity}(Node_A, Node_B,) = \frac{\lceil distance_{max} \rceil - distance(Node_A, Node_B)}{\lceil distance_{max} \rceil - distance_{min}}$$

where:

- $\lceil distance_{max} \rceil$ is the ceiling of the maximal standardized Euclidean distance observed between two nodes in the dataset;
- $distance_{min}$ is the minimal standardized Euclidean distance observed;
- $distance(Node_A, Node_B)$ the standardized Euclidean distance between our two nodes/patients of interest.

The proximity maximal value is 1 and it corresponds to the minimal distance and the minimal value is close to zero and it corresponds to the maximal distance. This standardized Euclidean distance was calculated with the neuropsychological features selected in the ML workflow each for the ENG (4 features) and then for the NAM condition (4 features also), independently.

We obtained in this way two complete weighted graphs (ENG patients' graph and NAM patients' graph), with no thresholding assumption. Every pair of nodes is connected but with edges of different weights (weighted graphs), according to the proximity. A higher weight means a smaller distance between the patients, this way a community of patients detected in the graph reveals a proximity of their neuropsychological profiles composed with the 4 selected indices.

3.2.1.3. Graph statistics.
Based on the respective patients' graphs, we used a state-of-the-art community detection to detect clusters of patients: the Louvain/Multilevel algorithm [57,58]. Communities appear in the weighted network if there are groups of nodes

with strong internal connections and weak external (between groups) connections. For both graphs we detected exactly two communities whose membership has been compared to binary clinical variables (such as hemispherical laterality, epilepsy severity, etc.; the variables described in Table 1) with a Simple Matching Coefficient (SMC). Simple Matching Coefficient is a simple and intuitive way to give the ratio of coincidence between the binary labels: 0% means that the labels have nothing in common and 100% that they have identical sequences. Simple Matching Coefficient is the ratio of both mutual presences and mutual absence with the length of the binary sequence.

3.2.2. Cognitive networks
3.2.2.1. General overview.
We have adopted a second network's perspective for which neuropsychological indices constitute the nodes of the graphs. The relationship between nodes is, in this case, the similarity of the performance obtained by patients. Two indices, i.e., two nodes, have a strong link if their sequences of scores have a very similar pattern in the population with TLE. In order to get the most informative network, we included the whole neuropsychological dataset (32 neuropsychological indices; see Appendix S1).

3.2.2.2. Graph construction.
We used a classical Pearson correlation coefficient to measure the similarity. Based on positive correlations only, we modeled weighted graphs including edges weighted by the correlation's strength. Given the moderate number of patients, we used a severe p-value threshold ($p < .005$) adjusted for multiple comparisons with Holm correction [59]. This procedure allowed us to obtain 4 sparse graphs or networks according to the postsurgery condition (ENG+/ENG−; NAM+/NAM−).

3.2.2.3. Graph statistics.
We applied both global and local measurements on the cognitive graphs. Community detection was performed with Multilevel/Louvain algorithm. We reported the (global) modularity index showing how strongly separated the different clusters are from each other. Density was also reported. This measure shows how connected the network is compared to a full graph by calculating the ratio of the number of edges and the number of possible edges. Finally, we estimated the average degree which is simply the average number of edges per node in the graph.

Regarding the local measures computed at the nodal level, we deployed a bootstrap strategy to make inferences about the populations. 1000 resampling iterations were done on each patient sub-group. Each time, a graph was built and two complementary centralities were measured on every node: the strength and the clustering coefficient. The first represents the sum of the edge weights of the adjacent edges for each node and the later quantifies how close its neighbors are to being a complete graph. In other words, clustering coefficient measures the local density in a network, the tendency to form a highly connected neighborhood.

4. Results

4.1. Predictive power and validity of the neuropsychological indices

Among the three different algorithms tested on our dataset, the best performance was obtained each time by the linear algorithm: the L2 logistic regression. The following results are thus obtained using this algorithm.

4.1.1. Prediction of the clinical outcome
We obtained a very high classification performance (*BAcc*) for the prediction of the clinical outcome (ENG score): mean performance = 82.5%±13.7%; precision = 93.1% ± 11.8%; recall = 77.3% ± 16.7% (see Appendix S2 for the distribution of performances and Fig. 2 Panel A for the confusion matrix). The feature selection approach (cf. Fig. 2 Panel B) shows a very good stability of $\widehat{\Phi}$ = 88% with k = 4 features selected (in terms of selection frequencies: VMI = 99%, AMI = 98.2%, NAM = 94% and TMT = 81.6%). The prediction of ENG based on the 4 selected features is further improved: *BAcc* = 84.4% ± 3.7%, with a remarkable level of prediction for the class corresponding to a worse postoperative clinical prognosis in particular (ENG− = 89.9%; Fig. 2 Panel A). The nomogram below shows the influence of the selected features in the prediction and can be used as a tool to predict a patient's risk of postoperative clinical deficit (individual level; Fig. 2 Panel C and associated captions provide detailed explanations). The PDP visualizations (Appendix S2) also show, as a complement, how the values of selected features change the prediction and how they interact.

4.1.2. Prediction of the naming outcome
Classification performance for predicting the naming outcome (NAM score) with all the cognitive indices is moderate: mean *BAcc* performance = 59.1% ± 14.6%; precision = 68.5% ± 16%; recall = 60.2% ± 21.2% (see Appendix S2 for the distribution of performances and Fig. 3 Panel A for the confusion matrix). However, we observed a clear improvement of the performance when the classification was restricted to the selected features (*BAcc* = 65.3% ± 3.4%) and in particular for the NAM class – representing postoperative naming decline (NAM− = 68.4%; Fig. 3 Panel A). The feature selection workflow has selected k = 4 cognitive features, in a relatively stable manner ($\widehat{\Phi}$ = 53.6%; frequencies: AMI = 98.4%, NAM = 93.4%, SFL = 81.6% and VCI = 62.4%; Fig. 3 Panel B). The nomogram below (NAM nomogram; Fig. 3 Panel C) shows the influence of the selected features in the prediction of the risk of postoperative naming decline (see also Appendix S2 for the PDPs). The example of Patient 15 was reused to illustrate the practical application of nomogram and to provide a concrete reading grid for the prediction.

4.2. Neuropsychological networks

#### 4.2.1. Networks of the clinical outcome
4.2.1.1. Patients' network.
The Louvain community detection algorithm applied on the ENG patients' network (individuals' network built from the 4 features selected by the ML prediction; see Fig. 2) detects 2 main communities of patients. The neuropsychological profiles, respectively associated with each of these two communities, are strongly similar to those presented by patients (SMC HEM = 0.77) with left TLE (LTLE) or right TLE (RTLE). In particular, patients with left hemisphere temporal epilepsy (i.e., LTLE), are mostly those who have an ENG− outcome and therefore present similar neuropsychological profiles on the selected indices. In a less clear-cut way, other variables such as gender or epilepsy severity show a moderate rate of coincidence of labels with detected communities of 60% or more (SMC GEN = 0.62; SMC SEV = 0.6). Simple Matching Coefficients of the other clinical variables (HDS, THY, NAM, and HS) are closer to chance (Fig. 4 Panel A).

4.2.1.2. Cognitive networks.
Regarding the networks generated on cognitive indices (i.e., the cognitive networks), the network related to ENG− is more fragmented (6 communities and a modularity index of 0.66) than the one associated with ENG+ (4 communities and a modularity index of 0.29; Fig. 4 Panel B). Appendix S3 shows the evolution of the modularity index as a function of the graph thresholding. The ENG− graph is less densely interconnected (den-

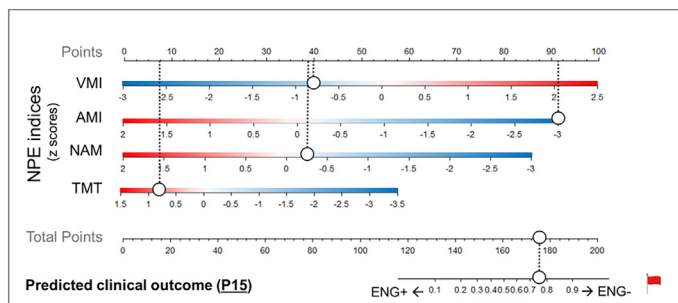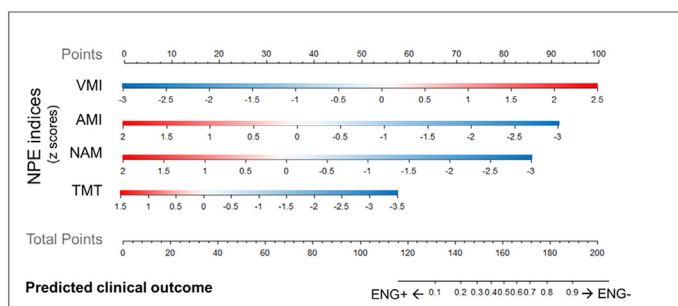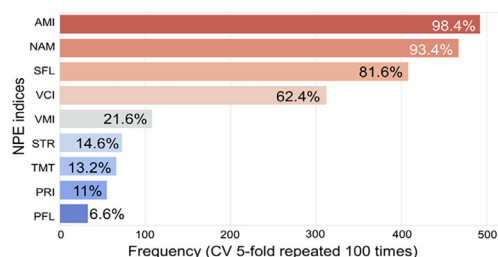## A | ENG Classification
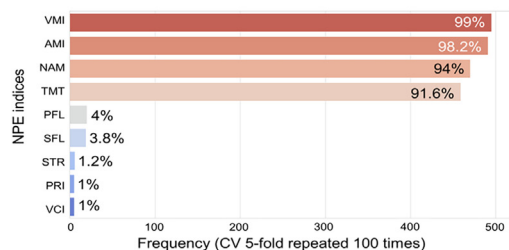
C | ENG Nomogram

Fig. 2. Prediction of postoperative clinical outcome. Panel A. Confusion matrix of average classification performance (*BAcc*) for each ENG class, before and after feature selection (FS). Panel B. Frequency and stability of features in the classification. Throughout the iterations, the 4 features are almost systematically selected (>90% of the time). Panel C. Nomogram intended to predict the postoperative clinical outcome (i.e., the risk of unsuccessful surgery as assessed with the ENG score; AUC = 85.2%, 1000 bootstrap). For its use, the first steps are to (a) position the z scores obtained by a given patient on the respective neuropsychological indices and to (b) report the number of points associated with each of these z scores, with the use of the top axis. For example, a random patient – Patient 15 (P15; bottom right) presenting the following z scores: VMI = −0.67; AMI = −3.47; NAM = −0.39; TMT = 0.91 – the points obtained are, respectively: 40 + 92 + 38 + 7.5. They are then added together (Total points = 177.5) to approximate the probability of being seizure free after the neurosurgery (bottom axis). According to the nomogram, P15 has a risk >75% of continuing to have seizures postoperatively. The tool can be used to draw the clinician's attention (red flag) if the risk of a poor postoperative clinical outcome is high.
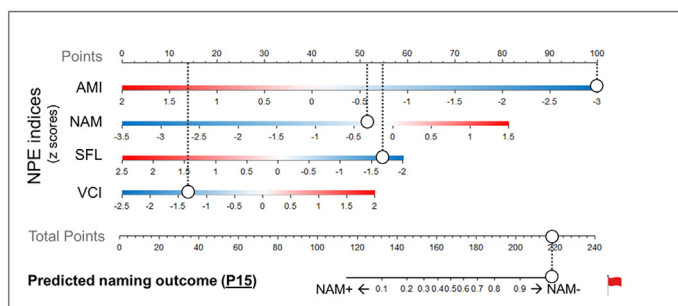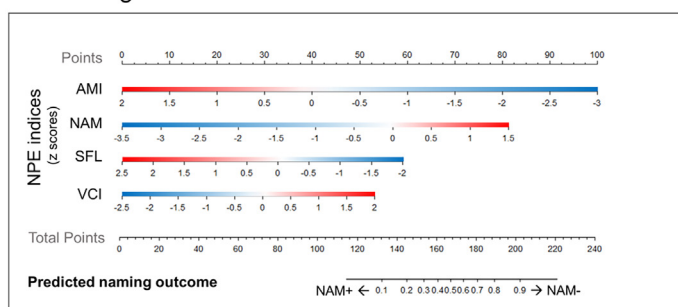
A | NAM Classification

C | NAM Nomogram

Fig. 3. Prediction of postoperative naming outcome. Panel A. Confusion matrix of the average classification performance (*BAcc*) of each NAM class, before and after feature selection (FS). Panel B. Feature frequency and stability in classification. VCI was less frequently selected than the other 3 cognitive features, but still contributed significantly to the prediction. Panel C. Nomogram designed to predict postoperative naming outcome and in particular the risk of experiencing a significant decline in naming abilities following neurosurgery (AUC = 72%, 1000 bootstrap). The procedure for using the nomogram is the same as that detailed in the legend of Fig. 2 Panel C. Following the nomogram, the cognitive profile of P15 on the relevant indices to predict NAM leads to a 100% risk of significant postoperative decline (red flag).
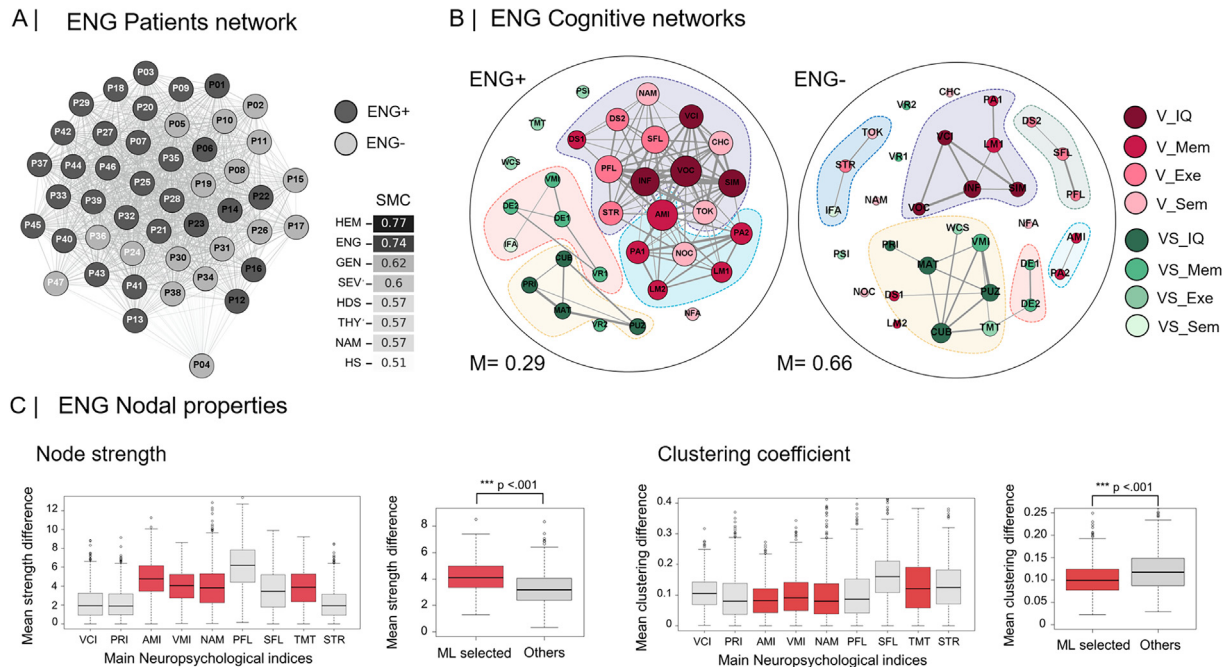
Fig. 4. Main properties of the networks associated with clinical outcome (ENG score). Panel A. Network of patients with TLE based on similarity of neuropsychological profiles on the 4 indices selected in the ENG prediction. White labels correspond to the first community of patients, black labels to the second. The nodes color corresponds to the affiliation to the ENG groups. Community 1 is more strongly represented by ENG+ patients, community 2 by ENG− patients. These communities also match with other clinical variables (SMCs), in particular the epilepsy-related hemisphere (see main text for further details). Panel B. Communities and modularity index (M) associated with each of the ENG cognitive networks. The size of the nodes is proportional to the centrality in the network (node strength). There is a clear separation of verbal (V) versus visuo-spatial indicators (VS) for ENG+ as well as a well-structured internal composition, consistent with the (theoretically) assessed functions. It is less marked for ENG−, for which these indicators are less distinctly separated into specialized communities. Panel C. ENG+/ENG− absolute differences of node strength and clustering coefficient (through 1000 iterations).

sity = 0.05 versus 0.21 for ENG+). The mean degree associated to ENG+ is indeed largely higher than that of the ENG− condition ($d = 6.57$ versus $d = 1.63$, respectively).

At the local level, there is a significant and systematic difference in node strength and clustering coefficient between ENG+ and ENG− networks (for all nodes; Appendix S3). The absolute ENG+/ENG− difference in node strength is significantly more important for the ML selected indices than the non-selected features ($t = 18.525$, df = 1997.7, $p < .001$), meaning that the overall centrality of ML nodes changes more between the ENG+ and ENG− conditions than that of the other nodes. Concerning the local centrality (clustering coefficient), the absolute ENG+/ENG− difference is significantly lower for the ML selected indices compared to the others ($t = −9.1225$, df = 1916.2, $p < .001$; Fig. 4 Panel C). The ML selected features have thus a distinct and specific pattern of connectivity compared to the unselected features.

4.2.2. Networks of the naming outcome
4.2.2.1. Patients' network.
The community detection performed on the NAM patients' network has identified 2 main communities of patients (Fig. 5 Panel A). As for ENG, the hemisphere involved in epileptic seizures shows a remarkable SMC of 70% with the detected communities. Based on the cognitive indices selected for the prediction of NAM (see Fig. 3), patients with LTLE show a similar neuropsychological profile to that of the community predominantly represented by NAM− (the reciprocal is also valid for RTLE and NAM+). To a lesser extent, manual laterality, presence of hippocampal sclerosis, gender, and thymic score present an SMC with the communities greater than or equal to 60% (SMC HDS = 0.64; SMC HS = 0.62; SMC GEN = 0.6; SMC THY = 0.6).

4.2.2.2. Cognitive network.
The cognitive networks of NAM+ and NAM− have the same number of communities ($n = 5$), but NAM− is slightly more modular and less dense ($M = 0.37$; density = 0.11) than NAM+ ($M = 0.28$; density = 0.2; Fig. 5 Panel B and see Appendix S3 for the modularity coefficient as a function of the graph thresholds). The average degree of connection of NAM− ($d = 3.38$) is also lower than that of NAM+ ($d = 5.81$).

Nodal strength and clustering coefficient of NAM+ and NAM− conditions are significantly different ($p < .001$), for almost all nodes (Appendix S3). We observe the same general pattern of difference in terms of centrality as observed for the ENG cognitive networks. We find a more modest but significant difference between the nodal strength of the ML selected nodes versus the unselected ones ($t = 2.23$, df = 1932.4, $p = 0.01$). This difference is related to greater changes between NAM+ and NAM− for ML selected nodes. The clustering coefficient difference is significantly lower for the ML selected nodes, compared to the others ($t = −14.94$, df = 1990.6, $p < .001$; Fig. 5 Panel C).

5. Discussion

The main objective of this data-driven study was to determine whether neuropsychological scores can be useful predictors of postoperative long-term outcome after temporal lobe resection (ATL in patients with drug-resistant TLE), and if so, to identify indicators that are the most relevant. We assessed postsurgical outcome through two important markers reflecting either long-term clinical (ENG) or naming (NAM) result of the neurosurgery. The findings of this study clearly demonstrate the validity of some NPE composite scores in predicting clinical and naming outcome and detecting patients at risk (red flag).
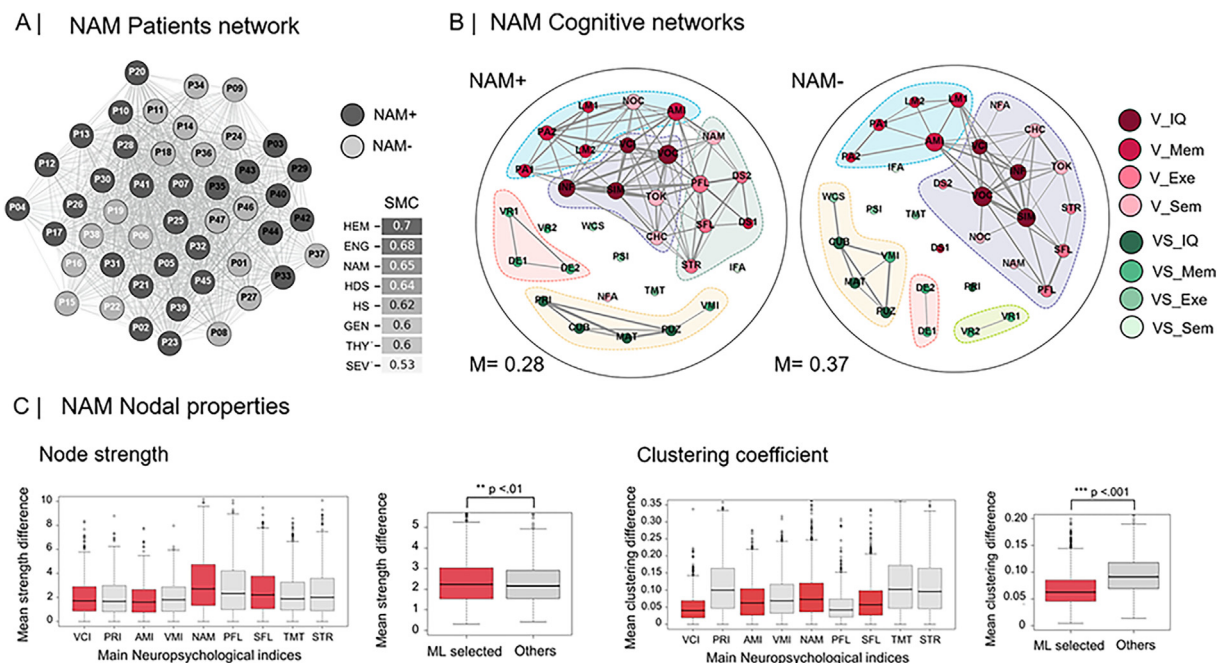
Fig. 5. Main properties of the networks associated with naming outcome (NAM score). Panel A. Network of patients with TLE based on similarity of neuropsychological profiles on the 4 indices selected in the NAM prediction. White labels correspond to the first community of patients, black labels to the second. The nodes color corresponds to the affiliation to the NAM groups. Community 1 is more strongly represented by NAM+ patients, community 2 by NAM− patients. These communities also match with other clinical variables (SMCs), in particular the epilepsy-related hemisphere (see main text for further details). Panel B. Communities and modularity index (M) associated with each of the NAM cognitive networks. The size of the nodes is proportional to the centrality in the network (node strength). There is a clear separation of verbal (V) versus non-verbal (VS) indicators for both of the NAM networks (NAM+ and NAM−). However, the verbal component of NAM− is much less interconnected, while the NAM+ component has highly interconnected specialized submodules (i.e., a "functional" internal congruence between cognitive indices). Panel C. NAM+/NAM− differences of node strength and clustering coefficient (through 1000 iterations).

Reliable prediction of long-term clinical outcome (ENG) is achieved as a combination of preoperative scores from 4 neuropsychological indices in particular – VMI, AMI (visual and auditory memory index of the WMS-IV [60]), NAM (DO80 naming task [36], and TMT (Trail Making Test B-A score [61]). Using these specific predictors, 8 out of 10 patients were correctly classified and the performance further increases to accurately identify 9 out of 10 patients among those who will have a less favorable clinical outcome (i.e., not seizure-free at 2 years after the surgery; ENG−). The power of these NPE indices to target at-risk profiles is therefore excellent (Fig. 2), implying that the initial cognitive state is clinically decisive. Concomitant observation of the entire cognition-brain-clinical sphere in this respect will help improve our understanding of the mechanisms underlying this relationship and will help identify its mediators. It should be noted, however, that we binarized the Engel clinical variable to distinguish complete/optimal surgical success (seizure-free: type I) from partial surgical success or therapeutic failure (persistent but more occasional seizures: types II–III; or no noteworthy improvement: type IV [33]; Table S1). Future work considering different scenarios and refining this categorization is needed to separate patients with occasional seizures after surgery (i.e., substantial improvement in their clinical condition) from patients with no clinical improvement, for example.

Long-term postoperative naming outcome (NAM) is however less easily predictable from preoperative cognitive scores. By relying on the four robustly selected neuropsychological indices – AMI (auditory memory index of the WMS-IV [60]), NAM (DO80 naming task [36]), SFL (semantic verbal fluency [61]), and VCI (verbal comprehension index of WAIS IV [62]) – about 2/3 of the patients can be correctly classified. The predictive power of this formula is nevertheless improved for the prediction of at-risk profiles of patients

that show a significant degradation of the naming efficiency at 2 years of follow-up (NAM−; Fig. 3). The moderate level of prediction obtained for NAM can be related to the method of coding used to define the improvement or deterioration of performance. With the objective of limiting the omission of at-risk patients (red flag profiles), we have chosen a rather permissive threshold (RCI, 10% unilateral), favoring the inclusion of patients without strong decline in the NAM− group. A more restrictive RCI of 95% (cut-offs below −1.65 SD) and/or other empirically based techniques for identifying the change from baseline (e.g., standardized regression-based change scores, changes estimated from standardized rather than raw scores [63]) may potentially improve the predictive power of neuropsychological indices for estimating postoperative naming change. In addition, the RCI is based on parameters related to task's performance distributions (in particular, the estimation of the standard deviation [64]). The overall internal consistency and test–retest stability of the naming task are good, making it a valid and psychometrically reliable task [65]. However, this task has limited discriminatory power [66] and is particularly sensitive to individuals with moderate to severe naming disorders (due to a "ceiling effect"; [67]). The performances of the normative sample as well as of the patients included in this study (see Appendix S2) show indeed skewed or bimodal distributions. Since truncated or non-normal distributions may bias the interpretation of change scores [68], applying additional correction on distributions or to RCI estimate (e.g., [69]) could result in (1) a more accurate approximation of change and therefore (2) a better predictive performance for NAM+/NAM− conditions.

Overall, this study highlights the high sensitivity and validity of specific neuropsychological tests in the context of ATL in temporal epilepsy. Specifically, the tests derived from the Wechsler scales

appear relevant and valid in these patients who may present subtle cognitive disorders (as mentioned in the Introduction; see also Table S1 for the average z scores), difficult to estimate with less sophisticated and comprehensive testing. We previously reported their clinical efficiency in a study aiming to lateralize epilepsy in drug-resistant patients [29] and the robustness of the WMS compared with other tests has also been highlighted by other studies in patients with epilepsy [70]. Furthermore, these neuropsychological indices are associated with the functional disturbances in resting-state brain connectivity exhibited by patients with TLE [11], confirming their clinical efficiency and potentially explaining why they are selected as good predictors of postsurgical status after temporal lobectomy. Thus, these tests appear specifically relevant and fine grained to target epilepsy-related disorders following ATL surgery. A future step may be to determine whether some sub-indicators of these composite indices are better predictors, thereby refining our observations and gaining a better understanding of the processes at work.

Recent studies have also tackled the prediction of postoperative naming performance of patients with refractory TLE. Busch and colleagues [31], for example, performed the predictions based on 10 predictors a priori associated with the change in naming performance following temporal lobe resection (clinical variables selected on the basis of evidence available in the literature, such as: sex, education, age at surgery, age at epilepsy onset, duration of epilepsy, side of surgery, etc.). They observed very good prediction performances of naming scores (6–12 months after surgery) and two clinical factors were particularly important for predicting the decline: the age at epilepsy onset (also found in [71–73]) and the side of surgery (consistent with [74,75]). By employing an individual graph approach (patients' graphs) and in accordance with the body of work on material-specific hemispheric specialization, we have similarly observed in this study that the hemisphere involved in epilepsy and ATL resection was the most important clinical variable (>70% matching with the respective neuropsychological profiles associated with ENG or NAM; Figs. 3-4, Panels A). Thus, the development of combined models, including relevant information from both neurological and neuropsychological examinations, could be beneficial in portraying "red flag" profiles. Moreover, the inclusion of biomarkers of postsurgical outcome in the model, such as resting-state fMRI GT measures [76] or an estimate of the resection volume, could also improve predictions. Although ATL is a standard procedure, the extent of resection is commonly larger in RTLE than LTLE [77]. Variations in the amount of tissue resected and/or in surgical techniques (e.g., standard ATL, Spencer-type lobectomies, lobectomies sparing the superior temporal gyrus, the hippocampus or the lateral neocortex, selective amygdalohippocampectomy, and lesionectomies) did not appear to have a major effect on cognitive outcome – except for naming [41] – and thus remain a factor to consider in future predictions.

In addition, prediction of a broader range of neuropsychological indices assessing problematic cognitive functions in the daily lives of patients with TLE is essential. Naming latency instead of naming score could be considered since it might be more sensitive in identifying finer-grained but troublesome everyday language impairments [38,78,79]. Others naming-related measures could also be highly sensitive to TLE and ATL and thus be used to predict postoperative language status in future works. In particular, the evaluation of the naming of specific categories, in auditory modality (on verbal input such as definition), or in a more natural and ecological context as in spontaneous speech ([40], for a systematic review). The decline in episodic memory (or learning performance) is another factor that should be aimed to be predicted given the high occurrence/incidence of memory disorders [41] and as a memory decrease is often observed in association with language deficits in patients with TLE [80]. Recently, Busch and collaborators

[32] reported reliable prediction of postoperative verbal memory decline by including the side of surgery, the baseline memory score, and the educational level or the hippocampal resection (depending on the memory score to be predicted). These results are promising and a prediction of varied cognitive scores – including functions more traditionally associated with the right hemisphere (or "non-dominant" hemisphere, such as visuo-spatial cognition abilities) – and using as an input a wide spectrum of neuropsychological indices would also be valuable for the development of comprehensive neuropsychological tools to assist clinicians in preoperative decision making and patient counseling.

Some sophisticated methods/algorithms deriving from artificial intelligence are not easy to comprehend (black box models) and thus prevent clinical translatability. The estimation of the classification/prediction performance is essential to determine the relevance of the model tested but does not describe the existing relationships. Considering the "accuracy versus interpretability trade-off" [81], statistical or even visualization techniques now allow to estimate the features importance, the features interaction or the model internals (structure and/or weights of learned attributes [55]). Interpretable surrogate models such as Partial Dependancy Plots (PDPs) provide, for example, threshold scores modulating prediction and reveal relationships between features (Appendix S2 and [29] for a previous application in patients with TLE). In the same way, nomograms allow to (1) directly and graphically visualize the influence of the predictors and; (2) perform individual predictions by locating a given patient. Overall, these tools facilitate the progress towards personalized neuropsychology (i.e., precision neuropsychology).

Furthermore, GT, applied to neuropsychological performance, makes explicit the interdependence between cognitive domains and has become a valuable tool in the study of systemic pathology such as epilepsy. In the past few years, efforts have been made to model the cognitive maps of patients with epilepsy through networks (in pediatric focal and generalized epilepsies [82]; or in TLE presenting diverse cognitive phenotypes [83,84]). Examined in comparison with those of control populations, epileptic cognitive networks systematically show a significant disruption, resulting in interconnectivity decrease and a more severe and disorganized modular fragmentation. Our study carried out in the peri-operative context shows that this ultra-modular and sparse organization of connections is particularly exaggerated in patients presenting a poor long-term clinical or naming outcome after surgery. Optimal cognitive functioning thus emanates from the synergy between indicators (or cognitive processes) and sustained support of an entire cognitive network after surgical injuries is essential. It can be translated into graph language by a higher density of the network and a lower modularity, as observed for conditions with favorable postsurgical outcomes (ENG+ and NAM+). Interestingly, the connectivity of the ML selected NPE indices is particularly different depending on the postsurgical outcome, which may explain the greater predictive power of these specific features. Memory scores such as AMI and general intellectual abilities (e.g., VCI and associated subtests) are important connector hubs in the cognitive landscape of patients with TLE (i.e., they are highly integrated nodes; Fig. 4-4, Panels B). Their connectivity, or centrality within the whole network, undergoes significant changes between the ENG+/ENG− and NAM+/NAM− conditions (Fig. 4-4, Panels C). They would indeed play a protective role against pre/postsurgical decline, echoing the concept of cognitive reserve [85].

Finally, the development and the application of techniques used in this study requires special precautions. In particular the sample size must be sufficient and as homogeneous as much as possible to detect robust patterns. The main pitfall in ML analyses, which is related to small samples, is the risk of overfitting. There are how-

ever several ways to control the reliability of the prediction that we applied in this study to ensure the generalizability of the results: adopting a multi-algorithm approach, performing CVs, iterating and estimating the stability of selected features are examples [86]. Regarding GT analyses, the method and thresholding of the matrices as well as the number of network nodes influence the results (discussed for example by [87–89] in the case of brain networks). To limit these biases, different solutions can be adopted, such as comparing the results obtained with those of random networks [90], estimating the evolution of metrics as a function of the threshold (see Appendix S3 in our case) or applying corrections to multiple comparisons problem. Optimally, results generalizability should be systematically assessed using a fully independent sample. This last point is far from trivial in clinical practice, given the high variability of neuropsychological practices across time, practitioners and/or clinical sites [91].

6. Conclusion

This data-driven research provides support that neuropsychology can serve as a relevant predictor for both clinical and long-term naming outcomes after traditional temporal lobectomy in patients with epilepsy. In this context, machine learning and GT methods are complementary and mutually beneficial. They offer insightful screening of (1) the profiles likely to experience a suboptimal postsurgical outcome; and (2) the cognitive landscape of the concerned patients. This study promotes evidence-based neuropsychology and emphasizes the concept of neuropsychological "canvas", which emphasizes the role of interactions between cognitive functions at the origin of the phenotypic expression of cognition. The prediction visualizations and techniques proposed in this research attempt to bridge the gap between fundamental and clinical research by providing concrete applications (e.g., nomograms). Beyond the diagnosis, the yielded tools and evidence can guide neuropsychologists in setting up an early or even anticipated cognitive rehabilitation ("cognitive prehabilitation") and thus envisage the most appropriate neuropsychological follow-up in case of a risk profile (red flags).

Funding

Ethical Statement

Patients provided written informed consent to participate in the study, which was approved by the local ethics committee (CPP: 09-CHUG-14/ANSM (ID RCB) 2009-A00632-55).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yebeh.2021.108357.

References

[1] Penfield W, Jasper H. Epilepsy and the functional anatomy of the human brain. Oxford, England: Little, Brown & Co.; 1954.
[2] Milner B. Memory and the medial temporal regions of the brain. Biol Mem 1970;23:31–59.
[3] Scoville WB, Milner B. Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry 1957;20:11–21.
[4] Wada J. A new method of determining the side of cerebral speech dominance: a preliminary report on the intracarotid injection of sodium amytal in man. Igaku Seibutsugaki 1949;14:221–2.
[5] Gazzaniga MS. Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? Brain 2000;123:1293–326. https://doi.org/10.1093/brain/123.7.1293.
[6] Novelly RA. The debt of neuropsychology to the epilepsies. Am Psychol 1992;47:1126–9. https://doi.org/10.1037/0003-066X.47.9.1126.
[7] Hermann B, Loring DW, Wilson S. Paradigm shifts in the neuropsychology of epilepsy. J Int Neuropsychol Soc JINS 2017;23:791–805. https://doi.org/10.1017/S1355617717000650.
[8] Brissart H, Maillard L. Neuropsychologie des épilepsies de l'adulte: Approche clinique et pratique. De Boeck Supérieur; 2018.
[9] Campos BM, Coan AC, Beltramini GC, Liu M, Yassuda CL, Ghizoni E, et al. White matter abnormalities associate with type and localization of focal epileptogenic lesions. Epilepsia 2015;1:125–32. https://doi.org/10.1111/epi.12871.
[10] Rodríguez-Cruces R, Velázquez-Pérez L, Rodríguez-Leyva I, Velasco AL, Trejo-Martínez D, Barragán-Campos HM, et al. Association of white matter diffusion characteristics and cognitive deficits in temporal lobe epilepsy. Epilepsy Behav 2018;79:138–45. https://doi.org/10.1016/j.yebeh.2017.11.040.
[11] Roger E, Pichat C, Torlay L, David O, Renard F, Banjac S, et al. Hubs disruption in mesial temporal lobe epilepsy. A resting-state fMRI study on a language-and-memory network. Hum Brain Mapp 2020;41(3):779–96. https://doi.org/10.1002/hbm.24839.
[12] Voets NL, Zamboni G, Stokes MG, Carpenter K, Stacey R, Adcock JE. Aberrant functional connectivity in dissociable hippocampal networks is associated with deficits in memory. J Neurosci 2014;34:4920–8. https://doi.org/10.1523/JNEUROSCI.4281-13.2014.
[13] Helmstaedter C, Witt J-A. Epilepsy and cognition – A bidirectional relationship? Seizure 2017;49:83–9. https://doi.org/10.1016/j.seizure.2017.02.017.
[14] Witt J-A, Helmstaedter C. Should cognition be screened in new-onset epilepsies? A study in 247 untreated patients. J Neurol 2012;259:1727–31. https://doi.org/10.1007/s00415-012-6526-2.
[15] Hoppe C, Elger CE, Helmstaedter C. Long-term memory impairment in patients with focal epilepsy. Epilepsia 2007;48:26–9.
[16] Alessio A, Pereira FRS, Sercheli MS, Rondina JM, Ozelo HB, Bilevicius E, et al. Brain plasticity for verbal and visual memories in patients with mesial temporal lobe epilepsy and hippocampal sclerosis: an fMRI study. Hum Brain Mapp 2013;34:186–99.
[17] Bartha-Doering L, Trinka E. The interictal language profile in adult epilepsy. Epilepsia 2014;55:1512–25. https://doi.org/10.1111/epi.12743.
[18] Jaimes-Bautista AG, Rodríguez-Camacho M, Martínez-Juárez IE, Rodríguez-Agudelo Y. Semantic processing impairment in patients with temporal lobe epilepsy. Epilepsy Res Treat 2015;2015:1–8. https://doi.org/10.1155/2015/746745.
[19] Metternich B, Buschmann F, Wagner K, Schulze-Bonhage A, Kriston L. Verbal fluency in focal epilepsy: a systematic review and meta-analysis. Neuropsychol Rev 2014;24:200–18. https://doi.org/10.1007/s11065-014-9255-8.
[20] Duff MC, Brown-Schmidt S. The hippocampus and the flexible use and processing of language. Front Hum Neurosci 2012;6. https://doi.org/10.3389/fnhum.2012.00069.
[21] Tracy JI, Boswell SB. CHAPTER 31 - Mesial temporal lobe epilepsy: A model for understanding the relationship between language and memory. In: Stemmer B, Whitaker HA, editors. Handb. Neurosci. Lang., San Diego: Elsevier; 2008, p. 319–28. https://doi.org/10.1016/B978-008045352-1.00031-8.
[22] Bora E, Meletti S. Social cognition in temporal lobe epilepsy: A systematic review and meta-analysis. Epilepsy Behav 2016;60:50–7. https://doi.org/10.1016/j.yebeh.2016.04.024.
[23] Lomlomdjian C, Múnera CP, Low DM, Terpiluk V, Solís P, Abusamra V, et al. The right hemisphere's contribution to discourse processing: A study in temporal lobe epilepsy. Brain Lang 2017;171:31–41. https://doi.org/10.1016/j.bandl.2017.04.001.
[24] Elverman KH, Resch ZJ, Quasney EE, Sabsevitz DS, Binder JR, Swanson SJ. Temporal lobe epilepsy is associated with distinct cognitive phenotypes. Epilepsy Behav 2019;96:61–8. https://doi.org/10.1016/j.yebeh.2019.04.015.
[25] Reyes A, Kaestner E, Ferguson L, Jones JE, Seidenberg M, Barr WB, et al. Cognitive phenotypes in temporal lobe epilepsy utilizing data- and clinically-driven approaches: Moving toward a new taxonomy. Epilepsia 2020;61:1211–20.
[26] Chauvière L. Potential causes of cognitive alterations in temporal lobe epilepsy. Behav Brain Res 2020;378:112310. https://doi.org/10.1016/j.bbr.2019.112310.
[27] Frank B, Hurley L, Scott TM, Olsen P, Dugan P, Barr WB. Machine learning as a new paradigm for characterizing localization and lateralization of

neuropsychological test data in temporal lobe epilepsy. Epilepsy Behav EB 2018;86:58–65. https://doi.org/10.1016/j.yebeh.2018.07.006.

[28] Keary TA, Frazier TW, Busch RM, Kubu CS, Iampietro M. Multivariate neuropsychological prediction of seizure lateralization in temporal epilepsy surgical cases. Epilepsia 2007;48:1438–46. https://doi.org/10.1111/j.1528-1167.2007.01098.x.

[29] Roger E, Torlay L, Gardette J, Mosca C, Banjac S, Minotti L, et al. A machine learning approach to explore cognitive signatures in patients with temporo-mesial epilepsy. Neuropsychologia 2020;142:107455. https://doi.org/10.1016/j.neuropsychologia.2020.107455.

[30] Baxendale S, Wilson SJ, Baker GA, Barr W, Helmstaedter C, Hermann BP, et al. Indications and expectations for neuropsychological assessment in epilepsy surgery in children and adults. Epileptic Disord 2019;21:221–34. https://doi.org/10.1684/epd.2019.1065.

[31] Busch RM, Hogue O, Kattan MW, Hamberger M, Drane DL, Hermann B, et al. Nomograms to predict naming decline after temporal lobe surgery in adults with epilepsy. Neurology 2018;91:e2144–52. https://doi.org/10.1212/WNL.0000000000006629.

[32] Busch RM, Hogue O, Miller M, Ferguson L, McAndrews MP, Hamberger M, et al. Nomograms to predict verbal memory decline after temporal lobe resection in adults with epilepsy. Neurology 2021;97:e263–74. https://doi.org/10.1212/WNL.0000000000012221.

[33] Engel J. Update on surgical treatment of the epilepsies. Summary of the Second International Palm Desert Conference on the Surgical Treatment of the Epilepsies (1992). Neurology 1993;43:1612–7. https://doi.org/10.1212/wnl.43.8.1612.

[34] Engel J. Seizures and epilepsy. Oxford University Press; 2013.

[35] Tonini C, Beghi E, Berg AT, Bogliun G, Giordano L, Newton RW, et al. Predictors of epilepsy surgery outcome: a meta-analysis. Epilepsy Res 2004;62:75–87. https://doi.org/10.1016/j.eplepsyres.2004.08.006.

[36] Metz-Lutz MN, Kremin H, Deloche G, Hannequin D, Ferrand I, Perrier D, et al. La dénomination orale d'images chez l'adulte - Test standardisé : effets du niveau de scolarité, de l'âge, du sexe. Glossa 1991.

[37] Kaplan E, Goodglass H, Weintraub S. The Boston Naming Test 1983.

[38] Moritz-Gasser S, Herbet G, Maldonado IL, Duffau H. Lexical access speed is significantly correlated with the return to professional activities after awake surgery for low-grade gliomas. J Neurooncol 2012;107:633–41. https://doi.org/10.1007/s11060-011-0789-9.

[39] Gisbert-Muñoz S, Quiñones I, Amoruso L, Timofeeva P, Geng S, Boudelaa S, et al. MULTIMAP: Multilingual picture naming test for mapping eloquent areas during awake surgeries. Behav Res Methods 2021;53:918–27. https://doi.org/10.3758/s13428-020-01467-4.

[40] Ives-Deliperi VL, Butler JT. Naming outcomes of anterior temporal lobectomy in epilepsy patients: A systematic review of the literature 2012:5.

[41] Sherman EMS, Wiebe S, Fay-McClymont TB, Tellez-Zenteno J, Metcalfe A, Hernandez-Ronquillo L, et al. Neuropsychological outcomes after epilepsy surgery: Systematic review and pooled estimates. Epilepsia 2011;52:857–69. https://doi.org/10.1111/j.1528-1167.2011.03022.x.

[42] Molnar C. Interpretable machine learning. Leanpub 2018.

[43] Garcia-Ramos C, Lin JJ, Kellermann TS, Bonilha L, Prabhakaran V, Hermann BP. Graph theory and cognition: A complementary avenue for examining neuropsychological status in epilepsy 2016:7.

[44] Königs M, Verhoog EM, Oosterlaan J. Exploring the neurocognome: neurocognitive network organization in healthy young adults. Cortex 2021;143:12–28. https://doi.org/10.1016/j.cortex.2021.06.011.

[45] Tosi G, Borsani C, Castiglioni S, Daini R, Franceschi M, Romano D. Complexity in neuropsychological assessments of cognitive impairment: A network analysis approach. Cortex 2020;124:85–96. https://doi.org/10.1016/j.cortex.2019.11.004.

[46] Wieser HG. Mesial temporal lobe epilepsy with hippocampal sclerosis. Epilepsia 2004;45:695–714. https://doi.org/10.1111/j.0013-9580.2004.09004.x.

[47] Chelune GJ. Assessing reliable neuropsychological change. Predict. Forensic Neuropsychol. Sound Stat. Pract., Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2003, p. 123–47.

[48] Martino J, Gomez E, de Lucas EM, Mato D, Vázquez-Bourgon J. Intraoperative identification and preservation of verbal memory in diffuse gliomas: A matched-pair cohort study. Neurosurgery 2018;83:1209–18. https://doi.org/10.1093/neuros/nyx617.

[49] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97. https://doi.org/10.1007/BF00994018.

[50] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc. 22Nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., New York, NY, USA: ACM; 2016, p. 785–94. https://doi.org/10.1145/2939672.2939785.

[51] Bellman RE. Dynamic programming. Princeton University Press; 1957.

[52] Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. 2010 20th Int. Conf. Pattern Recognit., 2010, p. 3121–4. https://doi.org/10.1109/ICPR.2010.764.

[53] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[54] Nogueira S, Sechidis K, Brown G. On the stability of feature selection algorithms 2018.

[55] Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning 2016.

[56] Csardi G, Nepusz T. The igraph software package for complex network research. 2006.

[57] Aynaud T, Blondel VD, Guillaume JL, Lambiotte R. Multilevel local optimization of modularity. Graph Partition 2013:315–45. https://doi.org/10.1002/9781118601181.ch13.

[58] Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008;2008:P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

[59] Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6:65–70.

[60] Wechsler D. Test MEM-IV échelle clinique de mémoire de WECHSLER-quatrième édition- Psychologie clinique - ECPA 2012. https://www.ecpa.fr/psychologie-clinique/test.asp?id=1987.

[61] Godefroy O, GREFEX. Fonctions exécutives et pathologies neurologiques et psychiatriques : Evaluation en pratique clinique. Marseille: DE BOECK UNIVERSITE; 2008.

[62] Wechsler D. Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV). 2011.

[63] Martin R, Sawrie S, Gilliam F, Mackey M, Faught E, Knowlton R, et al. Determining reliable cognitive change after epilepsy surgery: development of reliable change indices and standardized regression-based change norms for the WMS-III and WAIS-III. Epilepsia 2002;43:1551–8. https://doi.org/10.1046/j.1528-1157.2002.23602.x.

[64] Maassen GH. The standard error in the Jacobson and Truax Reliable Change Index: The classical approach to the assessment of reliable change. J Int Neuropsychol Soc 2004;10:888–93. https://doi.org/10.1017/S1355617704106097.

[65] Strauss E, Sherman N and AAPD of P and CNEMS, Spreen O. A compendium of neuropsychological tests: administration, norms, and commentary. Oxford University Press; 2006.

[66] Harry A, Crowe SF. Is the Boston Naming Test still fit for purpose? Clin Neuropsychol 2014;28:486–504. https://doi.org/10.1080/13854046.2014.892155.

[67] Pedraza O, Sachs BC, Ferman TJ, Rush BK, Lucas JA. Difficulty and discrimination parameters of Boston Naming Test items in a consecutive clinical series. Arch Clin Neuropsychol 2011;26:434–44. https://doi.org/10.1093/arclin/acr042.

[68] Gavett BE, Ashendorf L, Gurnani AS. Reliable change on neuropsychological tests in the Uniform Data Set. J Int Neuropsychol Soc 2015;21:558–67. https://doi.org/10.1017/S1355617715000582.

[69] Christensen L, Mendoza J. A method of assessing change in a single subject: An alteration of the RC index 1986. https://doi.org/10.1016/S0005-7894(86)80060-0.

[70] Soble JR, Eichstaedt KE, Waseem H, Mattingly ML, Benbadis SR, Bozorg AM, et al. Clinical utility of the Wechsler Memory Scale—Fourth Edition (WMS-IV) in predicting laterality of temporal lobe epilepsy among surgical candidates. Epilepsy Behav 2014;41:232–7. https://doi.org/10.1016/j.yebeh.2014.10.014.

[71] Bell B, Hermann B, Seidenberg M, Davies K, Cariski D, Rosenbek J, et al. Ipsilateral reorganization of language in early-onset left temporal lobe epilepsy. Epilepsy Behav 2002;3:158–64. https://doi.org/10.1006/ebeh.2002.0322.

[72] Davies KG, Risse GL, Gates JR. Naming ability after tailored left temporal resection with extraoperative language mapping: increased risk of decline with later epilepsy onset age. Epilepsy Behav EB 2005;7:273–8. https://doi.org/10.1016/j.yebeh.2005.05.016.

[73] Yucus CJ, Tranel D. Preserved proper naming following left anterior temporal lobectomy is associated with early age of seizure onset. Epilepsia 2007;48:2241–52. https://doi.org/10.1111/j.1528-1167.2007.01211.x.

[74] Gargaro AC, Sakamoto AC, Bianchin MM, Geraldi CdeVL, Scorsi-Rosset S, Coimbra ÉR, et al. Atypical neuropsychological profiles and cognitive outcome in mesial temporal lobe epilepsy. Epilepsy Behav 2013;27:461–9. https://doi.org/10.1016/j.yebeh.2013.03.002.

[75] Janecek JK, Winstanley FS, Sabsevitz DS, Raghavan M, Mueller W, Binder JR, et al. Naming Outcome After Left or Right Temporal Lobectomy in Patients with Bilateral Language Representation by Wada Testing. Epilepsy Behav EB 2013;28(1):95–8. https://doi.org/10.1016/j.yebeh.2013.04.006.

[76] Doucet GE, Rider R, Taylor N, Skidmore C, Sharan A, Sperling M, et al. Presurgery resting-state local graph-theory measures predict neurocognitive outcomes after brain surgery in temporal lobe epilepsy. Epilepsia 2015;56:517–26. https://doi.org/10.1111/epi.12936.

[77] Schmeiser B, Wagner K, Schulze-Bonhage A, Mader I, Wendling A-S, Steinhoff BJ, et al. Surgical treatment of mesiotemporal lobe epilepsy: which approach is favorable? Neurosurgery 2017;81:992–1004. https://doi.org/10.1093/neuros/nyx138.

[78] Bell BD, Seidenberg M, Hermann BP, Douville K. Visual and auditory naming in patients with left or bilateral temporal lobe epilepsy. Epilepsy Res 2003;55:29–37. https://doi.org/10.1016/S0920-1211(03)00110-4.

[79] Hamberger MJ, Seidel WT. Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. J Int Neuropsychol Soc 2003;9:479–89. https://doi.org/10.1017/S135561770393013X.

[80] Tramoni-Negre E, Lambert I, Bartolomei F, Felician O. Long-term memory deficits in temporal lobe epilepsy. Rev Neurol (Paris) 2017;173:490–7. https://doi.org/10.1016/j.neurol.2017.06.011.

[81] Freitas AA. Comprehensible classification models: a position paper. ACM SIGKDD Explor Newsl 2014;15(1):1–10. https://doi.org/10.1145/2594473.2594475.

[82] Kellermann TS, Bonilha L, Lin JJ, Hermann BP. Mapping the landscape of cognitive development in children with epilepsy. Cortex 2015;66:1–8. https://doi.org/10.1016/j.cortex.2015.02.001.

[83] Garcia-Ramos C, Lin JJ, Kellermann TS, Bonilha L, Prabhakaran V, Hermann BP. Graph theory and cognition: A complementary avenue for examining neuropsychological status in epilepsy. Epilepsy Behav 2016;64:329–35. https://doi.org/10.1016/j.yebeh.2016.02.032.

[84] Kellermann TS, Bonilha L, Eskandari R, Garcia-Ramos C, Lin JJ, Hermann BP. Mapping the neuropsychological profile of temporal lobe epilepsy using cognitive network topology and graph theory 2017:20.

[85] Hermann B, Seidenberg M, Lee E-J, Chan F, Rutecki P. Cognitive phenotypes in temporal lobe epilepsy. J Int Neuropsychol Soc 2007;13:12–20. https://doi.org/10.1017/S135561770707004X.

[86] Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nat Methods 2016;13:703–4. https://doi.org/10.1038/nmeth.3968.

[87] De Vico Fallani F, Richiardi J, Chavez M, Achard S. Graph analysis of functional brain networks: practical issues in translational neuroscience. Philos Trans R Soc B Biol Sci 2014;369:20130521. https://doi.org/10.1098/rstb.2013.0521.

[88] Fornito A, Zalesky A, Breakspear M. Graph analysis of the human connectome: Promise, progress, and pitfalls. NeuroImage 2013;80:426–44. https://doi.org/10.1016/j.neuroimage.2013.04.087.

[89] Sporns O. Contributions and challenges for network models in cognitive neuroscience. Nat Neurosci 2014;17:652–60. https://doi.org/10.1038/nn.3690.

[90] Becq GJ-PC, Barbier EL, Achard S. Brain networks of rats under anesthesia using resting-state fMRI: comparison with dead rats, random noise and generative models of networks. J Neural Eng 2020;17:045012. https://doi.org/10.1088/1741-2552/ab9fec.

[91] Vogt VL, Äikiä M, del Barrio A, Boon P, Borbély C, Bran E, et al. Current standards of neuropsychological assessment in epilepsy surgery centers across Europe. Epilepsia 2017;58:343–55. https://doi.org/10.1111/epi.13646.