

## Supplementary: Node abnormality load predicts chances of seizure recurrence after epilepsy surgery

Nishant Sinha, MSc<sup>1,2\*</sup>; Yujiang Wang, PhD<sup>2,3</sup>; Nádia Moreira da Silva, MSc<sup>2</sup>; Anna Miserocchi, MD<sup>3</sup>; Andrew W. McEvoy, FRCS<sup>3</sup>; Jane de Tisi, BS Hons<sup>3</sup>; Sjoerd B. Vos, PhD<sup>3,4,5</sup>; Gavin P. Winston, PhD<sup>3,5,6</sup>; John S. Duncan, FRCP<sup>3,5</sup>; Peter N. Taylor, PhD<sup>2,3\*</sup>

<sup>1</sup>Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

<sup>2</sup>Computational Neuroscience, Neurology, and Psychiatry Lab, Interdisciplinary Computing and Complex BioSystems Group, School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom

<sup>3</sup>NIHR University College London Hospitals Biomedical Research Centre, UCL Institute of Neurology, Queen Square, London, United Kingdom

<sup>4</sup>Centre for Medical Image Computing, University College London, London, United Kingdom

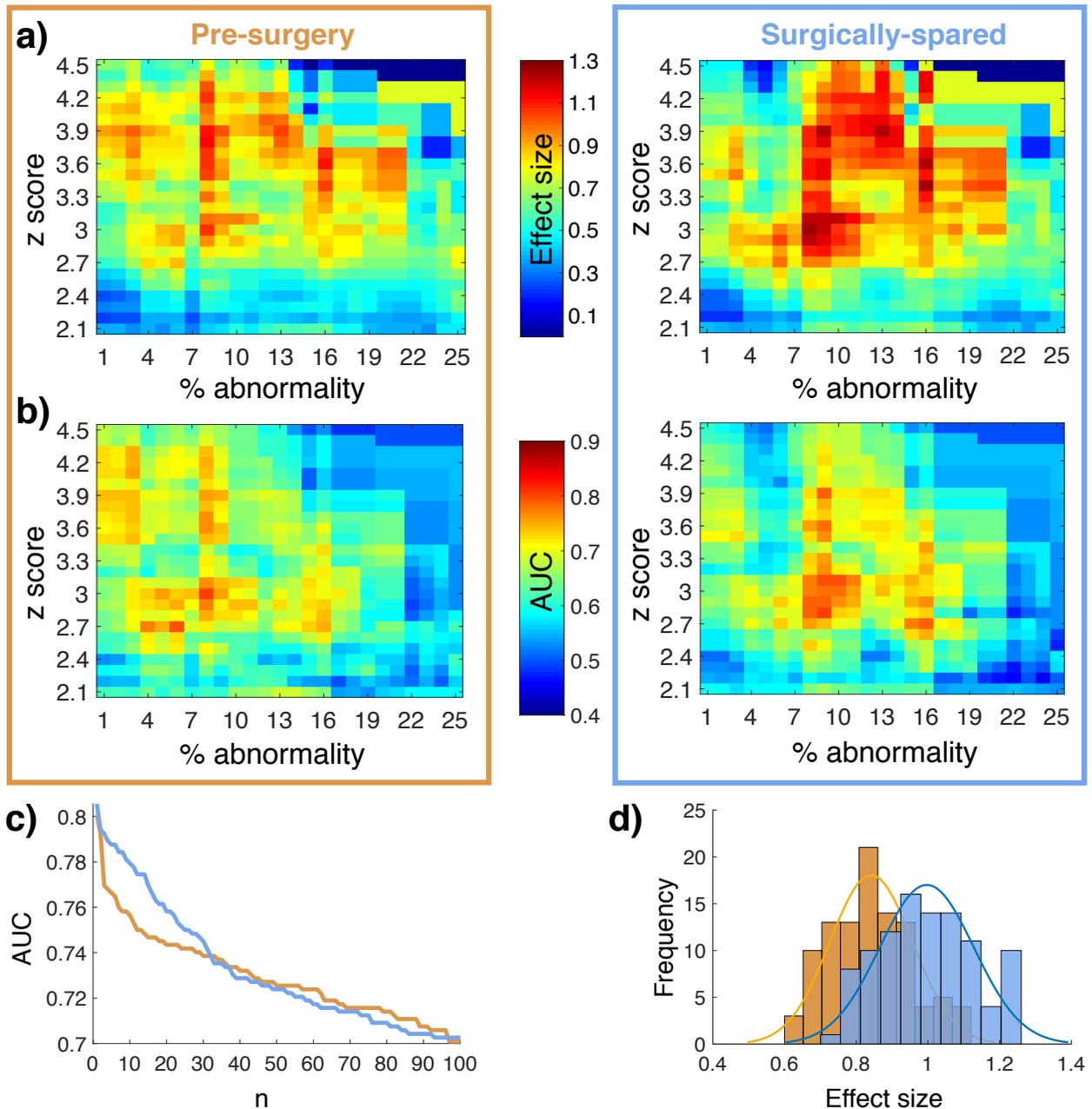
<sup>5</sup>Epilepsy Society MRI Unit, Chalfont St Peter, United Kingdom

<sup>6</sup>Department of Medicine, Division of Neurology, Queen's University, Kingston, Canada

\*Corresponding author: nishant.sinha89@gmail.com (NS); peter.taylor@newcastle.ac.uk (PNT)

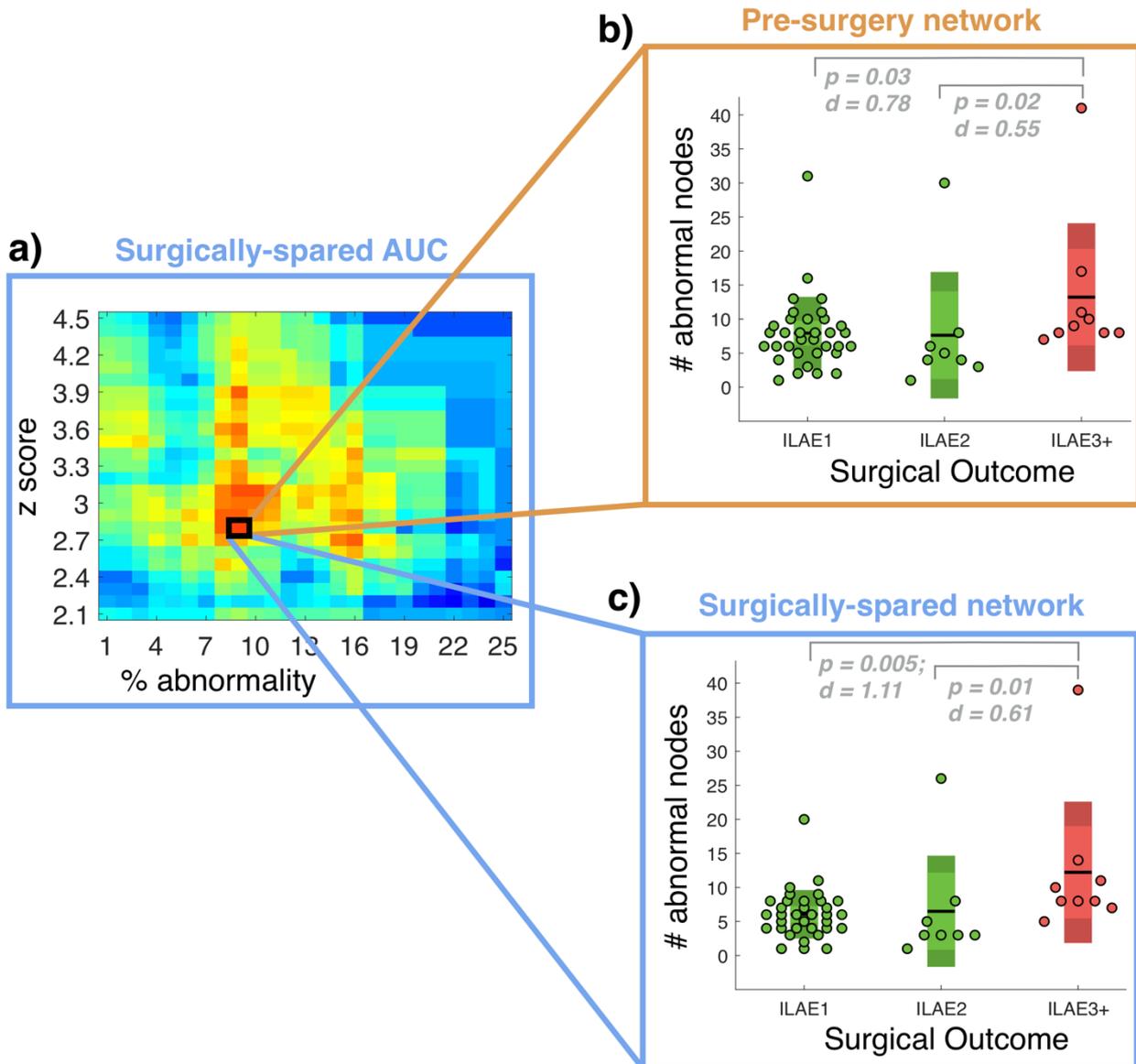
Address: Urban Sciences Building, 1 Science Square, Newcastle Upon Tyne, NE4 5TG, Tyne and Wear, UK

## Supplementary Figures



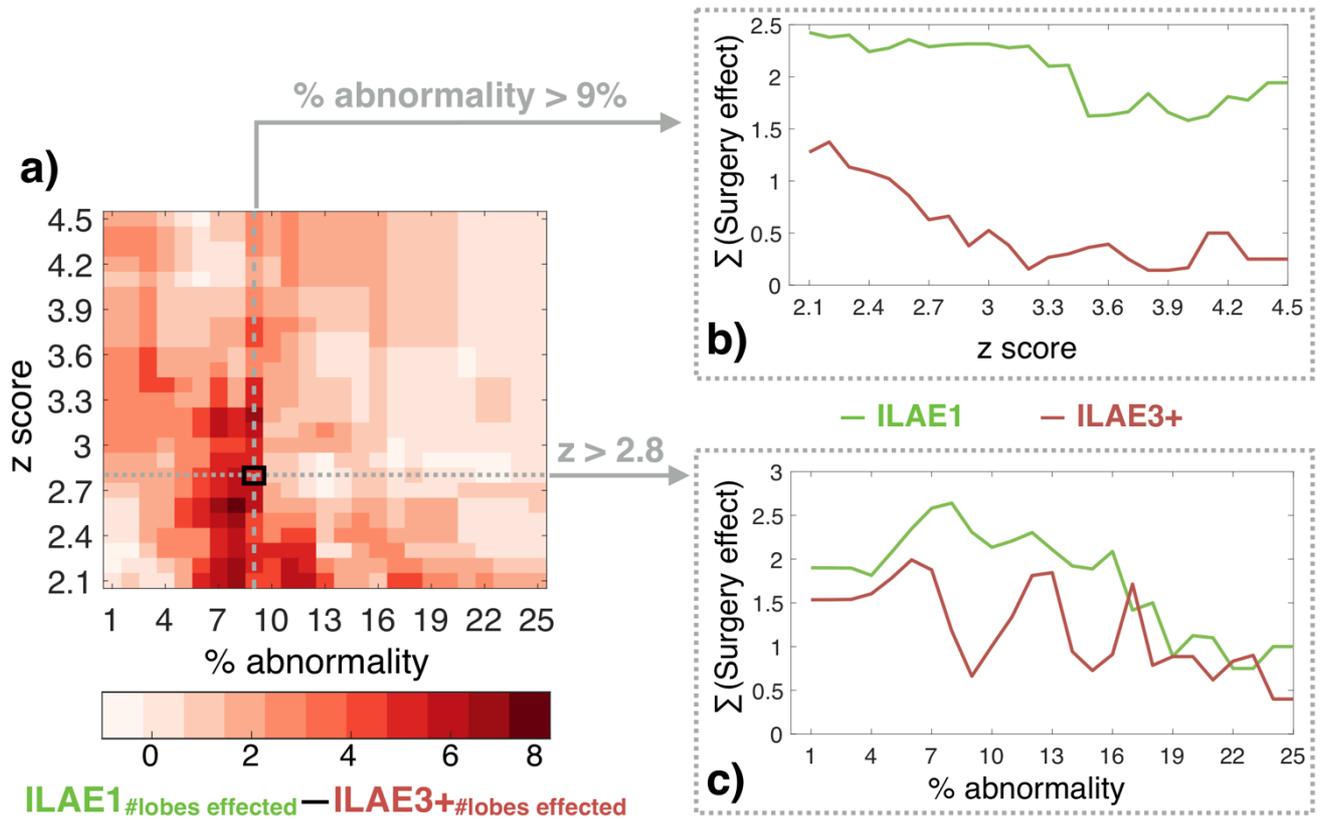
**Figure S1: Association between node abnormality and surgical outcomes in pre-surgery and surgically-spared network is consistent across thresholds. a)** Effect size (d-score) for node abnormality between ILAE 1 and ILAE 3+ groups in pre-surgery and surgically-spared networks. Positive effect size indicates higher node abnormality in ILAE 3+ patients compared to ILAE 1 patients. Medium to large positive effects size, colour coded in red, are evident across a large range of thresholds. **b)** AUC quantifying the discriminatory value of node abnormality is shown at every point on the threshold grid for pre-surgery and surgically-spared networks. **c)** 100 highest AUCs sampled from the threshold grid of pre-surgery and surgically-spared networks in

panel **b)** are plotted against each other. Relatively higher AUCs are apparent in surgically-spared networks. **d)** Histogram of effect size sampled from pre-surgery and surgically-spared networks corresponding to the threshold values of 100 highest AUCs of surgically-spared network.

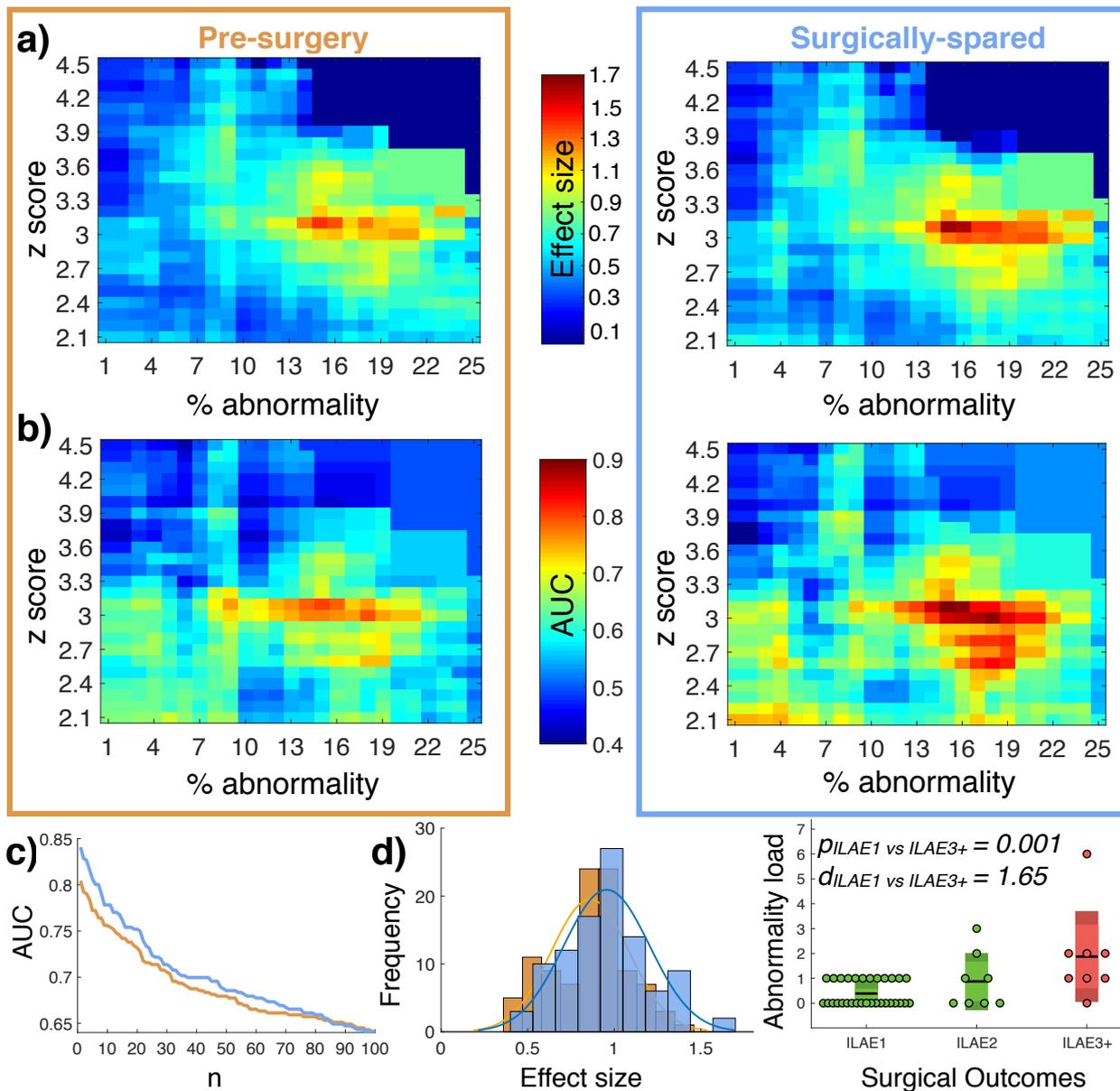


**Figure S2: Surgically-spared networks are more discriminatory than the pre-surgery networks.**

**a)** Corresponding to the threshold of highest AUC, results are shown for pre-surgery and surgically-spared networks. **b)** Node abnormality computed from pre-surgery network discriminates ILAE1 from ILAE3+ with an effect size of  $d = 0.78$  [95%CI 0.04, 2.1] and ILAE 2 from ILAE 3+ with an effect size of  $d = 0.55$  [95%CI -0.95, 2.0]. Other statistical estimates in pre-surgery network: ILAE1 ( $n=34$ ) median 7.5 [95%CI 6, 8.5]; ILAE2 ( $n=8$ ) median 4.5 [95%CI 3, 8]; ILAE3+ median 9 [95%CI 8, 17]. In comparison, node abnormality computed from surgically-spared networks in **c)** discriminate ILAE1 from ILAE3+ with an effect size of  $d = 1.11$  [95%CI 0.42, 2.2] and ILAE2 from ILAE3+ with an effect size of  $d = 0.61$  [95%CI -0.92 2.04]. Other statistical estimates in surgically-spared network: ILAE1 ( $n=34$ ) median 6 [95%CI 5, 7.5]; ILAE2 ( $n=8$ ) median 3 [95%CI 2, 5.5]; ILAE3+ median 8 [95%CI 5, 10].

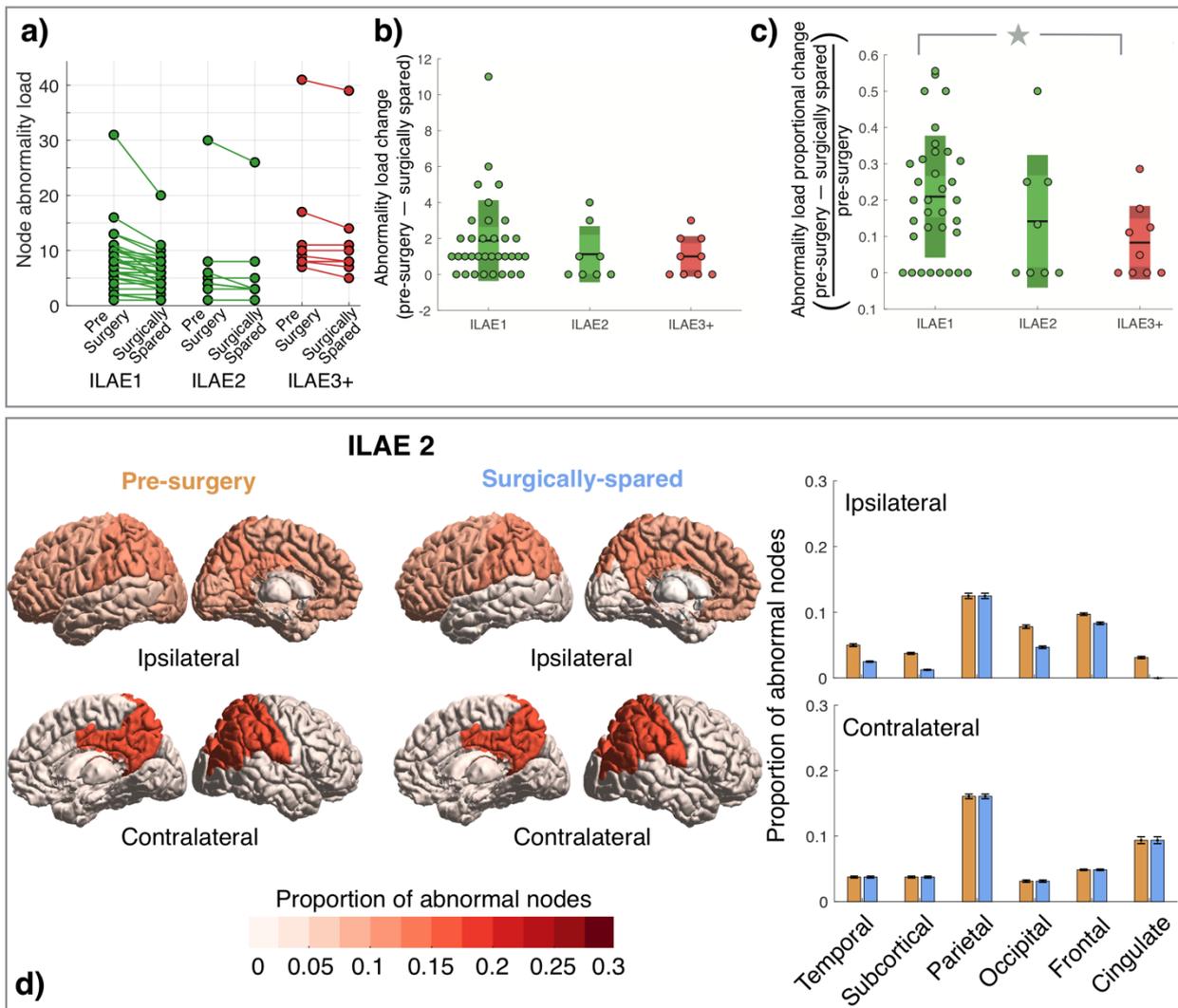


**Figure S3: The widespread effect of surgery in reducing node abnormality in seizure free group is consistent across the thresholds.** Furthering our results in Figure 4, we show the difference between the number of lobes effected in seizure free and non-seizure free group across thresholds in panel a). Positive differences in a) indicate that surgery effects the node abnormalities in more brain areas of ILAE 1 group compared to ILAE 3+ group. Along the two slices taken from the grid, the net effect of surgery is shown between ILAE 1 and ILAE 3+ groups in panel b) and c). The effect of surgery at each lobe is computed as:  $1 - \text{ratio of proportion of abnormal nodes in ILAE1 to ILAE3+}$ . The net effect of surgery is computed as the sum of the aforementioned ratio across all the ipsilateral and contralateral brain areas.



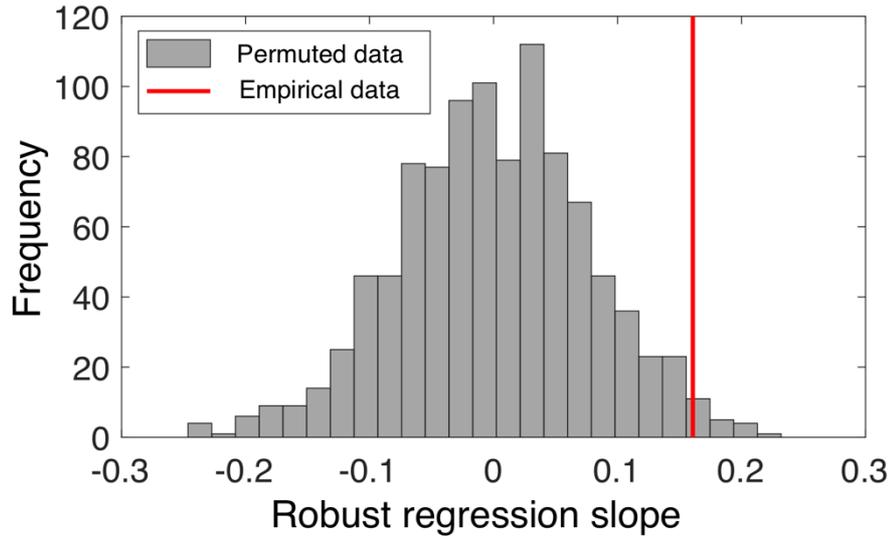
**Figure S4:** With networks inferred using Desikan-Killiany parcellation scheme the association between node abnormality and surgical outcomes in pre-surgery and surgically-spared networks are consistent across thresholds. **a)** Effect size (d-score) for node abnormality between ILAE 1 and ILAE 3+ groups in pre-surgery and surgically-spared networks. Positive effect size indicates higher node abnormality in ILAE 3+ patients compared to ILAE 1 patients. Medium to large positive effects size, colour coded in red, are evident across a large range of thresholds. **b)** AUC quantifying the discriminatory value of node abnormality is shown at every point on the threshold grid for pre-surgery and surgically-spared networks. **c)** 100 highest AUCs sampled from the threshold grid of pre-surgery and surgically-spared networks in panel **b)** are plotted against each other. Relatively higher AUCs are apparent in surgically-spared networks. **d)** Histogram of effect size sampled from pre-surgery and surgically-spared networks corresponding to the

threshold values of 100 highest AUCs of surgically-spared network. **e)** At an example point on the threshold grid corresponding to the highest AUC, the box plot shows significantly higher number of abnormal nodes in ILAE 3+ patient than in ILAE 1 patients.

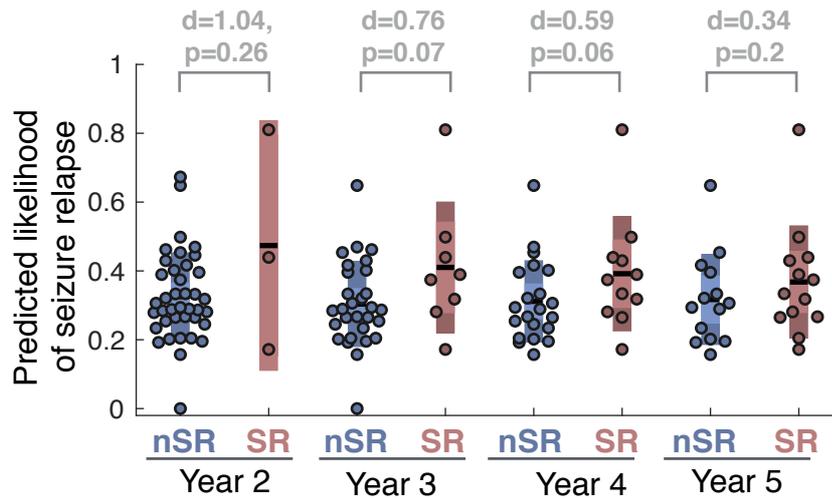


**Figure S5: Change in node abnormality load between pre-surgery and surgically-spared networks.** **a)** Reduction in the number of abnormal nodes between pre-surgery and surgically-spared networks are shown for ILAE 1, ILAE 2, and ILAE 3+ patients. ILAE 3+ patients have a greater number of abnormal nodes than ILAE 1 patients. Inspecting visually, the reduction of abnormality load, apparent by the slope of the lines, are more in ILAE 1 patients than in ILAE 3+. Reduction in ILAE 2 patients are intermediate to ILAE 1 and ILAE 3+. **b)** The absolute reduction in abnormal nodes between pre-surgery and surgically spared networks are higher on average in ILAE 1 patients but not statistically significant ( $p=0.14$ ;  $d=0.42$  [95%CI -0.25, 0.84]). **c)** The proportional drop in the number of abnormal nodes between pre-surgery and surgically-spared networks relative to the node abnormality load pre-surgery is significantly higher ( $p=0.01$ ;  $d=0.81$  [95%CI 0.2, 1.4]) in the ILAE 1 patients ( $n=34$ , median 0.2 [95%CI 0.14, 0.3]) compared to the ILAE 3+ patients ( $n=9$ , median 0.05 [95%CI 0, 0.15]). **d)** Equivalent of Figure 4 for ILAE 2 patients: Spatial reduction in node abnormality due to surgery in ILAE 2 patients are also intermediate to

ILAE 1 and ILAE 3+ patients. Bar plot shows the drop in surgically-spared network compared to pre-surgery network in five ipsilateral (temporal subcortical, occipital, frontal, and cingulate) only.



**Figure S6:** Steepness of regression slope obtained from robust regression tested for significance using permutation testing ( $p = 0.004$ , number of permutations = 1000).



**Figure S7: On combining network features with clinical attributes association with relapse is lost.** The predicted 12-month likelihood of seizure relapse was estimated from the SVM model at an example combination of features that yielded highest classification performance (corresponding to Figure 5b-c). Amongst the patients who were initially seizure-free (i.e., ILAE 1 or ILAE 2 at year 1), the likelihood of seizure relapse was not different to those who had a subsequent relapse.

## Supplementary Methods

### Imaging protocols

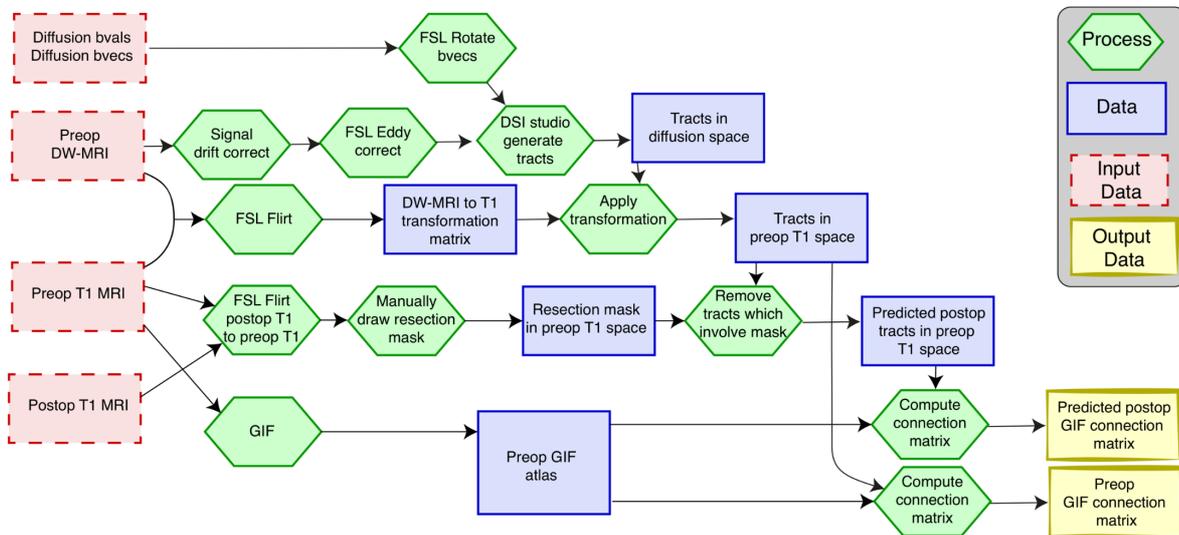
MRI data were acquired on a 3T GE Signa HDx scanner (General Electric, Waukesha, Milwaukee, WI). Standard imaging gradients with a maximum strength of  $40 \text{ mTm}^{-1}$  and slew rate  $150 \text{ Tm}^{-1}\text{s}^{-1}$  were used. All data were acquired using a body coil for transmission, and 8-channel phased array coil for reception. Standard clinical sequences were performed including a coronal 3D T1-weighted volumetric acquisition (matrix,  $256 \times 256 \times 170$ ; in-plane resolution,  $0.9375 \times 0.9375 \text{ mm}$  with slice thickness  $1.1 \text{ mm}$ ).

Diffusion MRI data were acquired using a cardiac-triggered single-shot spin-echo planar imaging sequence (Wheeler-Kingshott et al., 2002) with echo time =  $73 \text{ ms}$ . Sets of 60 contiguous  $2.4 \text{ mm}$ -thick axial slices were obtained covering the whole brain, with diffusion sensitizing gradients applied in each of 52 noncollinear directions (b-value of  $1200 \text{ mm}^2\text{s}^{-1}$  [ $\delta = 21 \text{ ms}$ ,  $\Delta = 29 \text{ ms}$  using full gradient strength of  $40 \text{ mTm}^{-1}$ ]) along with 6 non-diffusion weighted scans. The gradient directions were calculated and ordered as described elsewhere (Cook et al., 2007). The field of view was  $24 \times 24 \text{ cm}$ , and the acquisition matrix size was  $96 \times 96$ , zero filled to  $128 \times 128$  during reconstruction, giving a reconstructed voxel size of  $1.875 \times 1.875 \times 2.4 \text{ mm}$ . The DTI acquisition time for a total of 3480 image slices was approximately 25 min (depending on subject heart rate). These protocols are identical to our previous study (Taylor et al. 2018).

### Pre-processing pipeline

In summary, the pre-surgery and surgically-spared networks are generated as follows: the preoperative T1 image were parcellated into 114 cortical and subcortical regions of interest (ROIs) derived from the predefined Geodesic Information Flow atlas and separately in 82 ROIs using the Freesurfer Desikan-Killiany atlas in the native space of each participant. We registered the parcellated ROIs, resection mask, and tracts from deterministic tractography on dMRI data in native space. The pre-surgical streamline network is the connectivity matrix depicting the number of streamlines connecting two ROIs. The surgically-spared streamline network is inferred after removing the streamlines that intersected the resection mask. By definition, surgery can only cause an immediate reduction in the number of streamlines. Therefore, we specified that the surgically-spared network contains only those network edges which are not expected to

change in streamline count following surgery (i.e., edges where their streamlines do not pass through/into the resection cavity).



Reproduced from Taylor *et al.* 2018: Summary of processing pipeline to generate GIF network.

The pipeline is applied to each subject.

We applied the same image pre-processing steps that we previously established on this data in our recent study (Taylor *et al.* 2018). The image processing pipeline is summarised in the flowchart reproduced from Taylor *et al.* 2018.

Preoperative diffusion MRI data were first corrected for signal drift (Vos *et al.* 2017), then eddy current and movement artefacts were corrected using the FSL `eddy_correct` tool (Andersson and Sotiropoulos 2016). The b vectors were then rotated appropriately using the ‘`fdt-rotate-bvecs`’ tool as part of FSL (Jenkinson *et al.* 2012, Leemans & Jones, 2009). The diffusion data were reconstructed using generalised q-sampling imaging (Yeh, Wedeen, and Tseng 2010) with a diffusion sampling length ratio of 1.2. A deterministic fibre tracking algorithm (Yeh *et al.* 2013) was then used, allowing for crossing fibres within voxels, with seeds placed at the whole brain. Default tractography parameters from the 14 February 2017 build of DSI studio software were used as follows. The angular threshold used was 60 degrees and the step size was set to 0.9375mm. The anisotropy threshold was determined automatically by DSI Studio. Tracks with length less than 10 mm and more than 300mm were discarded. A total of 1,000,000 tracts were calculated per subject and saved in diffusion space.

To align the tracts with the ROIs we linearly registered the first non-diffusion-weighted image to the preoperative T1 image and saved the inverse of this transformation matrix using FSL FLIRT. We then multiplied every coordinate in every tract by this transformation matrix to get the tracts in T1 space. The quality of the registration between tracts, ROIs, and the resection mask was confirmed through visual inspection for all subjects. Since networks are constructed in native space, this removes any mismatching of track types due to potential nonlinear registration issues which is advantageous compared to previous studies of network change (Kuceyeski 2016). To generate preoperative connectivity matrices, we looped through all tracts and deemed two regions as connected if the two endpoints of the tract terminate in those regions. This generated a weighted connectivity matrix in which each entry in the matrix represents the number of streamlines connecting two regions. To generate surgically-spared connectivity matrices, we performed the same process as above with one exception. Any tract that had any point within the resection mask was excluded from building the matrix. The inferred surgically-spared network therefore always had fewer streamlines than the preoperative network.

#### Extended methodological details of predictive-model design and nested-cross validation

We predicted the patient-specific probability of seizure relapse using 13 preoperative clinical data, pre-surgery node abnormality, and the surgically-spared node abnormality. Preoperative clinical data (mentioned in Table 1, S1) included: sex, epilepsy onset age, age at surgery, epilepsy duration, number of anti-epileptic drugs taken preoperatively, history of status epilepticus, evidence of secondary generalised seizures, side of surgery, preoperative MRI pathology, evidence of hippocampal sclerosis, history of depression, history of psychosis, and history of other psychiatric disorder.

We performed this using support vector machine (SVM) implemented in MATLAB 'fitcsvm' classification library (Platt, 1999; Guyon et al., 2002). We applied a linear kernel because this enables the interpretation of weight vectors (i.e., the relative importance of each feature in the prediction), which were used to rank the importance of metrics in identifying patients who would have suboptimal seizure outcome. SVMs were initially trained with all 15 preoperative metrics: 13 clinical, 1 pre-surgery node abnormality, and 1 surgically-spared node abnormality. To identify the most informative metrics, after each round of SVM training, we removed the least important metric (in terms of its weight vector) and trained a new SVM with the remaining metrics. We repeated this process until only a single metric remained (Guyon et al., 2002; Fagerholm et al.,

2015). At each stepwise removal we recorded: a) the performance of classifier in classifying totally seizure free (ILAE 1) and non-seizure free (ILAE 3-6) patients, and b) the Spearman's rank correlation between the predicted probability of seizure relapse for each patient with the actual severity of seizure outcomes at one-year after surgery (ILAE class).

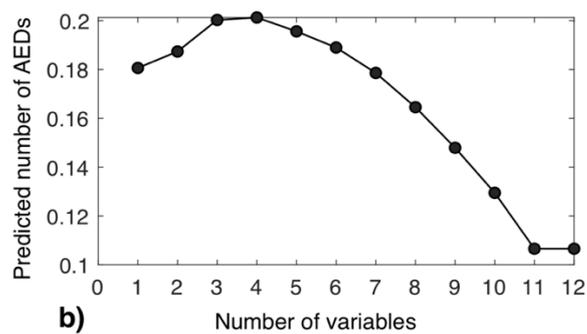
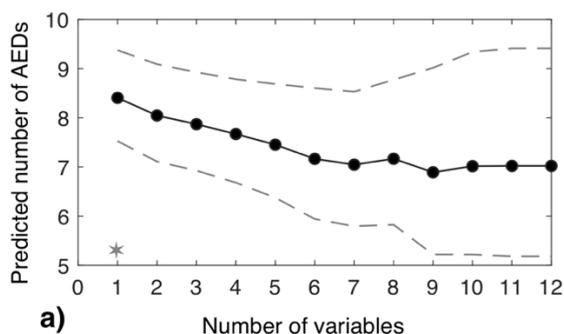
The performance of the classifier was estimated using binary classification. Given that ILAE 2 patients tend to relapse (Table 1, S1), and thus, are in the spectrum between the totally seizure free (ILAE 1) and non-seizure free (ILAE 3-6) patients, we first excluded the ILAE outcome group 2 patients (Fairclough et al., 2018). With these patients removed, our dataset consisted of 43 samples, 34 of which were labelled 1 corresponding to ILAE 1, and 9 were labelled -1, corresponding to ILAE 3-6. On this dataset, we performed nested-cross validation by combining a three-way split of the data (training-validation-testing) with leave-one-out cross-validation (CV) and grid search for SVM parameter (box-constraint) tuning. This was done to avoid upward bias in the metrics of performance estimates (Guyon and Elisseeff, 2003; Tsamardinos et al., 2018). Additionally, we avoided any bias in the selection of the most discriminatory threshold pair (i.e., z-score and percentage abnormality) to determine the node abnormality by computing it at every step of cross-validation after removing the test subject (Smialowski et al., 2009).

Specifically, in nested-cross validation, an external leave-one-out is implemented in which one patient is left out at every step for testing and the remaining patients used for training and validation. Training and validation were performed in the internal leave-one-out CV in which one patient is again left out for validation and the remaining used for model training combined with model parameter tuning. In our analysis, we tuned the model on 100 logarithmically spaced grid points between 1 and 10. At every point, the SVM is trained and its performance tested using the patient left out for validation by estimating AUC. We selected the model parameter that gave the highest cross-validated AUC. The classifier generalisation capability is then evaluated by computing the classification AUC, accuracy, sensitivity, and specificity using the patient originally left for testing in the external cross-validation. We also noted the probability with which each test patient was classified as non-seizure free. The intuition being that the predictive model, though blind to the non-seizure free outcome categories (i.e., all ILAE 3 to ILAE 6 are labelled as -1), would classify the patients with worse surgical outcome with a higher probability.

To determine where the ILAE outcome group 2 subjects fall on the spectrum, we treated all 8 ILAE outcome group 2 patients as test subjects. SVMs were trained and tuned, as described above, on all the remaining seizure free (ILAE 1) and non-seizure free (ILAE 3-6) patients (43 patients). On the classifier with highest discrimination between the seizure free and non-seizure free patients, we tested the features of ILAE 2 patients to note only the probability of classification to the non-seizure free group. We refer to these probabilities as the likelihood of seizure relapse because a high probability indicates a predicted propensity towards a non-seizure free outcome. Having obtained the likelihood of seizure relapse for all 51 patients, we compared this with the surgical outcome categories at year 1 and the actual seizure relapse in five years post-surgery. Note that the labels for all training data are binary and based on 12-month ILAE1 versus ILAE3-6 outcomes only. The model is therefore blind to severity of outcome (i.e. ILAE class 2, 3, 4, 5), and also blind to outcomes beyond 12 months.

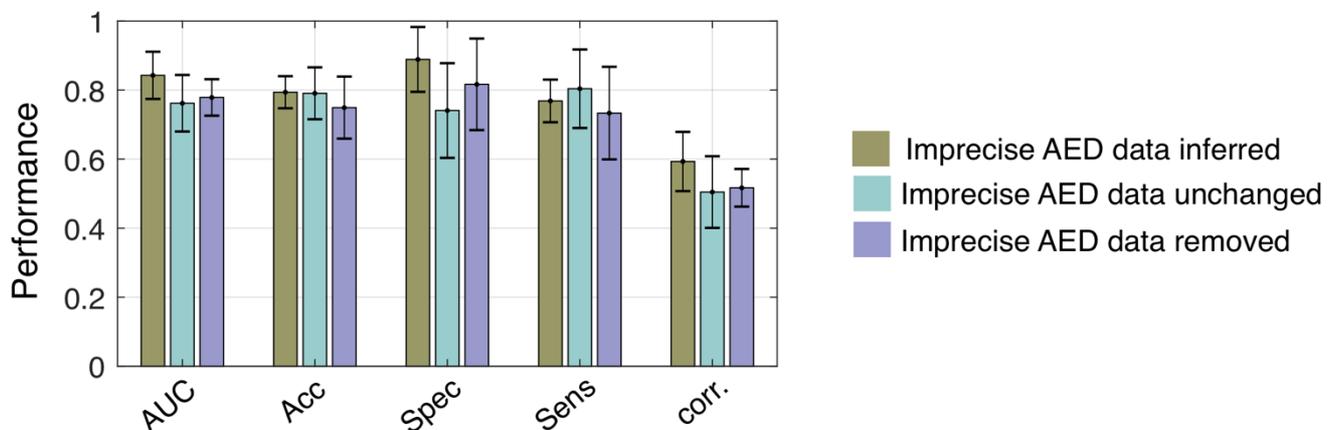
### Inferring imprecise AED information

For one patient, the clinical metadata was incomplete—the number of AEDs taken by patient 43 was imprecise. To infer a value for the number of AEDs for this patient, we investigated the predictive power of the remaining clinical data for AEDs prediction using multiple linear regression. The fitting of the linear regression model was achieved using stats package in R, which uses the Least Squares method to minimise the sum of the squares of residuals. Initially, we included all available clinical data in the linear regression model. However, to overcome multicollinearity and remove redundant predictors we discarded, at each round, the predictor with highest  $p$ -value, until the regression model and its coefficients were statistically significant. A significant model ( $p = 0.002$ ) with normally distributed residuals (Shapiro-Test:  $p = 0.7159$ ) was obtained using epilepsy duration as single significant predictor. The model was used to predict, for the patient with missing data, the number of AEDs and its 95% confidence interval.



**Figure 1.** Prediction of number of AEDs for the patient with imprecise AED data. The number of variables represents the number of predictors used by the model at each round. In panel **a)** the solid line represents the predictive AEDs value and the dash lines the upper and lower confidence interval. The adjusted  $r^2$  of the model at each round is shown in panel **b)**.

Since number of AEDs cannot be a fraction, we predicted the number of AEDs taken by patient 43 was 8. The following figure shows that the prediction performance remained similar regardless the imprecise data being imputed, left unchanged, or with patient 43 removed from the analysis. Thus, our results are robust to this missing data.



**Figure 2:** Consistent prediction performance was achieved regardless of how the imprecise AED data for patient 43 was treated. The bar plot shows the average performance metrics of SVMs in terms of AUC, accuracy, specificity, and sensitivity. The error bars represent the standard deviation across each step-wise feature removal.

### Contingency tables for risk reduction

The contingency table for absolute risk reduction and number needed to treat computation are shown in this section corresponding to: Figure S2b (association of abnormality load with year 1 seizure outcome in pre-surgery network), Figure 3e (association of abnormality load with year 1 seizure outcome in surgically-spared network), Figure 3g (association of abnormality load in surgically-spared network with seizure relapse in 5 years), Figure 6 (association of year 1 predicted likelihood of seizure relapse in year 1 ILAE 1-2 patients with actual seizure relapses at year 2, year 3, year 4, and year 5).

In year 1 the adverse event is poor (ILAE 3-6) seizure outcome. We compute the absolute risk reduction (ARR) of ILAE 3-6 seizure outcome by considering low abnormality load as a prognostic measure relative to the high abnormality load.

Confidence interval of ARR is computed as follows (Bussière and Wiebe 2005):

$$95\% \text{ CI for ARR} = \text{ARR} \pm 1.96 \sqrt{\frac{\text{CER} \times (1-\text{CER})}{\text{CS}} + \frac{\text{EER} \times (1-\text{EER})}{\text{ES}}}$$

Abbreviations: EN: Experiment group no adverse event; EE: Experiment group adverse event; CN: Control group no adverse event; CE: Control group adverse event; ES = total subjects in experiment group; CS: total subject in control group

*Table A: Contingency table for risk reduction in pre-surgery network corresponding to Fig S2b*

Year 1	Events (seizure outcome)		Total subjects	Event rate
	ILAE 1-2	ILAE 3-6		
Experiment: Low abnormality load < median of # abnormal nodes in ILAE3-6 group (9)	EN = 30	EE = 4	ES = 34	EER = EE/ES = 0.12
Control: High abnormality load ≥ median of # abnormal nodes in ILAE3-6 group (9)	CN = 12	CE = 5	CS = 17	CER = CE/CS = 0.29
Total	42	9		
Absolute risk reduction (ARR)	CER - EER = 0.17 [95%CI -0.07, 0.41]			
Number needed to treat (NNT)	1/ARR = 5.8 [95%CI -13.9, 2.42]			

*Table B: Contingency table for risk reduction in surgically spared network corresponding to Fig 3e*

Year 1	Events (seizure outcome)		Total subjects	Event rate
	ILAE 1-2	ILAE 3-6		
Experiment: Low abnormality load < median of # abnormal nodes in ILAE3-6 group (8)	EN = 30	EE = 2	ES = 32	EER = EE/ES = 0.0625
Control: High abnormality load ≥ median of # abnormal nodes in ILAE3-6 group (8)	CN = 12	CE = 7	CS = 19	CER = CE/CS = 0.368
Total	42	9		
Absolute risk reduction (ARR)	CER - EER = 0.31 [95%CI 0.08, 0.54]			
Number needed to treat (NNT)	1/ARR = 3.2 [95%CI 1.8, 12.9]			

For follow-up years the adverse event is occurrence of seizure relapse in year 1 ILAE1-2 patients. We compute the absolute risk reduction (ARR) of seizure relapse by considering low abnormality load as a prognostic measure relative to the high abnormality load.

*Table C: Contingency table for risk reduction corresponding to surgically spared network Fig 3g*

Year 5	Events (seizure relapse in 5 years)		Total subjects	Event rate
	No seizure relapse	Seizure relapse		
<b>Experiment: Low abnormality load &lt; median of # abnormal nodes in seizure relapse group (8)</b>	EN = 13	EE = 6	ES = 19	EER = EE/ES = 0.32
<b>Control: High abnormality load ≥ median of # abnormal nodes in seizure relapse group (8)</b>	CN = 1	CE = 7	CS = 8	CER = CE/CS = 0.88
<b>Total</b>	14	13		
<b>Absolute risk reduction (ARR)</b>	CER - EER = 0.56 [95%CI 0.25, 0.87]			
<b>Number needed to treat (NNT)</b>	1/ARR = 1.8 [95%CI 1.15, 3.96]			

Table D(i): Contingency table for risk reduction corresponding to Figure 6 at year 2:

Year 2	Events (seizure relapse at year 2)		Total subjects	Event rate
	No seizure relapse	Seizure relapse		
Experiment: Low predicted likelihood of seizure relapse < median of predicted likelihood in the seizure relapse group (0.18)	EN = 30	EE = 1	ES = 31	EER = EE/ES = 0.03
Control: High predicted likelihood of seizure relapse ≥ median of predicted likelihood in the seizure relapse group (0.18)	CN = 9	CE = 2	CS = 11	CER = CE/CS = 0.18
<b>Total</b>	39	3		
<b>Absolute risk reduction (ARR)</b>	CER - EER = 0.15 [95% CI -0.08, 0.38]			
<b>Number needed to treat (NNT)</b>	1/ARR = 6.7 [-11.78, 2.5]			

Table D(ii): Contingency table for risk reduction corresponding to Figure 6 at year 3:

Year 3	Events (seizure relapse at year 3)		Total subjects	Event rate
	No seizure relapse	Seizure relapse		
Experiment: Low predicted likelihood of seizure relapse < median of predicted likelihood in the seizure relapse group (0.18)	EN = 26	EE = 3	ES = 29	EER = EE/ES = 0.1
Control: High predicted likelihood of seizure relapse ≥ median of predicted likelihood in the seizure relapse group (0.18)	CN = 3	CE = 5	CS = 8	CER = CE/CS = 0.62
<b>Total</b>	29	8		
<b>Absolute risk reduction (ARR)</b>	CER - EER = 0.52 [95%CI 0.17, 0.87]			
<b>Number needed to treat (NNT)</b>	1/ARR = 1.92 [95%CI 1.14, 6]			

Table D(iii): Contingency table for risk reduction corresponding to Figure 6 at year 4:

Year 4	Events (seizure relapse at year 4)		Total subjects	Event rate
	No seizure relapse	Seizure relapse		
Experiment: Low predicted likelihood of seizure relapse < median of predicted likelihood in the seizure relapse group (0.18)	EN = 20	EE = 5	ES = 25	EER = EE/ES = 0.2
Control: High predicted likelihood of seizure relapse ≥ median of predicted likelihood in the seizure relapse group (0.18)	CN = 1	CE = 6	CS = 7	CER = CE/CS = 0.86
<b>Total</b>	21	11		
<b>Absolute risk reduction (ARR)</b>	CER - EER = 0.66 [95%CI 0.36, 0.96]			
<b>Number needed to treat (NNT)</b>	1/ARR = 1.52 [95%CI 1.04, 2.79]			

Table D(iv): Contingency table for risk reduction corresponding to Figure 6 at year 5:

Year 5	Events (seizure relapse at year 5)		Total subjects	Event rate
	No seizure relapse	Seizure relapse		
Experiment: Low predicted likelihood of seizure relapse < median of predicted likelihood in the seizure relapse group (0.18)	EN = 14	EE = 6	ES = 20	EER = EE/ES = 0.3
Control: High predicted likelihood of seizure relapse $\geq$ median of predicted likelihood in the seizure relapse group (0.18)	CN = 0	CE = 7	CS = 7	CER = CE/CS = 1
<b>Total</b>	14	13		
<b>Absolute risk reduction (ARR)</b>	CER - EER = 0.7 [95%CI 0.5, 0.9]			
<b>Number needed to treat (NNT)</b>	1/ARR = 1.4 [95%CI 1.11, 2]			

### Data availability

Data and code to reproduce figures are available at the following DOI:  
10.5281/zenodo.4486794

## Reference

Wheeler-Kingshott, Claudia A.M., Hickman, Simon J., Parker, Geoffrey J.M., Ciccarelli, Olga, Symms, Mark R., Miller, David H., Barker, Gareth J., 2002. Investigating cervical spinal cord structure using axial diffusion tensor imaging. *NeuroImage* 16 (1), 93–102.

Cook, Philip A., Symms, Mark, Boulby, Philip A., Alexander, Daniel C., 2007. Optimal acquisition orders of diffusion-weighted MRI measurements. *J. Magn. Reson. Imaging* 25 (5), 1051–1058.

Kuceyeski, Amy, Navi, Babak B., Kamel, Hooman, Raj, Ashish, Relkin, Norman, Togli, Joan, Iadecola, Costantino, O'dell, Michael, 2016. Structural connectome disruption at baseline predicts 6-months post-stroke outcome. *Hum. Brain Mapp.* 37 (7), 2587–2601.

S.B. Vos, C.M.W. Tax, P.R. Luijten, S. Ourselin, A. Leemans, M. Froeling, “The importance of correcting for signal drift in diffusion MRI”. *Magn Reson Med* 2017;77(1):285-299.

Platt JC. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv large margin Classif* 1999

Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. *J Mach Learn Res* 2003; 3: 1157–1182.

Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002

Fagerholm ED, Hellyer PJ, Scott G, Leech R, Sharp DJ. Disconnection of network hubs and cognitive impairment after traumatic brain injury. *Brain* 2015; 138: 1696–1709.

Tsamardinos I, Greasidou E, Borboudakis G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach Learn* 2018

Smialowski P, Frishman D, Kramer S. Pitfalls of supervised feature selection. *Bioinformatics* 2009

Bussièrè, M. and Wiebe, S., 2005. Progress in clinical neurosciences: measuring the benefit of therapies for neurological disorders. *Canadian journal of neurological sciences*, 32(4), pp.419-424.

## Supplementary Tables

<b>Term</b>	<b>Definition</b>
<b>Pre-surgery network</b>	Whole-brain connectivity using pre-operative diffusion MRI and pre-operative structural MRI only (Bonilha et al. 2015)
<b>Surgically-spared network</b>	Whole-brain connectivity using pre-operative diffusion MRI, pre-operative structural MRI, and a resection mask drawn on pre-operative structural MRI highlighting the surgery location (Taylor et al., 2017)
<b>Edge abnormality</b>	In the whole-brain connectivity the deviation of edges from normality. Normal distribution for the edges is inferred from a control population and deviation is measured in terms of z-score (Bonilha et al. 2015).
<b>Node abnormality</b>	Number of abnormal edges of a node normalised by the total number of edges in that node.
<b>Abnormality load</b>	The total number of abnormal nodes in the whole-brain connectivity of a patient.

**Table S1: Complete preoperative clinical information of all patients with postoperative ILAE seizure outcomes and relapse**

[Attached separately]

**Table S2: Summarising SVM performance estimates at each step-wise feature removal**

Feature eliminati on step	AUC	Accuracy	Sensitivity	Specificity	Correlation year-1 severity	
					Spearman's rho	p
1	0.76 [95%CI 0.51 0.91]	0.67	0.62	0.89	0.45 [95%CI -0.08 0.81]	0.0355
2	0.85 [95%CI 0.66 0.95]	0.79	0.76	0.89	0.49 [95%CI 0.02 0.84]	0.0241
3	0.80 [95%CI 0.48 0.94]	0.79	0.76	0.89	0.47 [95%CI -0.02 0.84]	0.0278
4	0.88 [95%CI 0.62 0.97]	0.86	0.85	0.89	0.57 [95%CI 0.09 0.88]	0.0084
5	0.91 [95%CI 0.75 0.98]	0.84	0.82	0.89	0.64 [95%CI 0.17 0.91]	0.0027
6	0.91 [95%CI 0.75 0.98]	0.84	0.82	0.89	0.64 [95%CI 0.21 0.91]	0.0027
7	0.92 [95%CI 0.78 0.98]	0.79	0.74	1.00	0.59 [95%CI 0.15 0.90]	0.0059
8	0.89 [95%CI 0.75 0.96]	0.79	0.74	1.00	0.66 [95%CI 0.25 0.88]	0.0018
9	0.89 [95%CI 0.74 0.96]	0.81	0.79	0.89	0.69 [95%CI 0.25 0.92]	0.0011
10	0.89 [95%CI 0.74 0.96]	0.79	0.74	1.00	0.71 [95%CI 0.32 0.94]	0.0008
11	0.91 [95%CI 0.77 0.97]	0.84	0.79	1.00	0.70 [95%CI 0.28 0.93]	0.0009
12	0.81 [95%CI 0.62 0.92]	0.79	0.79	0.78	0.64 [95%CI 0.20 0.89]	0.0026
13	0.74 [95%CI 0.42 0.90]	0.74	0.74	0.78	0.50 [95%CI -0.01 0.88]	0.0195
14	0.78 [95%CI 0.58 0.90]	0.74	0.71	0.89	0.59 [95%CI 0.08 0.91]	0.0062
15	0.72 [95%CI 0.46 0.91]	0.81	0.85	0.67	0.55 [95%CI -0.01 0.88]	0.0118

**Table S1: Complete preoperative clinical information of all patients with postoperative ILAE seizure outcomes and relapse data**

ID	Sex	Onset	Surg.	Dur.	AEDs	SE	SGS	Side	MRI	HS	Depr.	Psychos.	Other Psych.	ILAE yr1	ILAE yr2	Rel. yr2	ILAE yr3	Rel. yr3	ILAE yr4	Rel. yr4	ILAE yr5	Rel. yr5	Rel. in 5 yrs
1	M	31	45.9	14.9	4	N	Y	L	Abr	Y	N	N	N	1	2	nSR	3	SR	3	SR	1	SR	SR
2	F	3	49	46	15	N	Y	L	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
3	F	7	46	39	6	N	Y	L	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
4	F	9	42.5	33.5	10	N	Y	R	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
5	M	12	45.1	33.1	6	N	Y	L	Abr	Y	Y	N	Y	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
6	F	16	31	15	8	Y	Y	L	Abr	Y	Y	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
7	M	7	26.4	19.4	5	N	N	L	Abr	Y	N	N	N	1	1	nSR	1	nSR					
8	M	17	32.5	15.5	7	N	N	L	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
9	F	2	57.3	55.3	5	N	Y	R	Abr	Y	N	N	Y	1	1	nSR	1	nSR	3	SR	1	SR	SR
10	M	1	47.3	46.3	7	N	Y	L	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
11	M	20	35.2	15.2	7	N	Y	L	Nor	N	N	N	N	1	1	nSR	4	SR	3	SR	3	SR	SR
12	F	31	39.9	8.9	4	N	Y	L	Abr	N	N	N	N	1	2	nSR	1	nSR	1	nSR	1	nSR	nSR
13	M	13	21.1	8.1	6	N	Y	L	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	3	SR	SR
14	F	2	19.1	17.1	8	N	Y	L	Abr	Y	N	N	N	1	2	nSR	3	SR	1	SR	5	SR	SR
15	M	31	42.6	11.6	3	N	Y	R	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
16	F	15	20.1	5.1	5	N	Y	R	Nor	N	Y	N	Y	1	1	nSR	1	nSR	2	nSR	4	SR	SR
17	F	10	41.9	31.9	3	N	Y	R	Nor	N	Y	N	N	1	1	nSR	1	nSR	1	nSR			
18	F	7	48	41	8	N	Y	R	Abr	Y	N	N	N	1	1	nSR	1	nSR					
19	F	13	20.3	7.3	5	Y	Y	L	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR			
20	F	0.9	27.3	26.4	4	N	Y	L	Abr	N	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
21	M	32	40.6	8.6	4	N	N	L	Abr	N	N	N	Y	1	1	nSR	1	nSR	1	nSR			
22	F	1.5	31.7	30.2	3	N	N	R	Abr	Y	N	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
23	M	2	39.1	37.1	8	N	Y	L	Abr	Y	Y	N	N	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
24	F	17	29.2	12.2	7	N	Y	L	Abr	Y	Y	N	N	1	1	nSR							
25	F	12	48.2	36.2	5	N	Y	R	Abr	Y	N	N	N	1	1	nSR	1	nSR					
26	F	15	23.3	8.3	7	N	Y	R	Nor	N	N	N	Y	1	1	nSR	1	nSR	3	SR	3	SR	SR

27	F	38	43.2	5.2	5	N	N	L	Abr	N	N	N	Y	1	1	nSR	1	nSR	1	nSR	1	nSR	nSR
28	F	0.66	25	24.34	6	Y	Y	R	Nor	N	N	N	N	1	1	nSR	1	nSR					
29	M	0.01	52.9	52.89	7	N	Y	L	Abr	Y	Y	N	N	1	1	nSR							
30	M	11	20.6	9.6	5	Y	Y	R	Abr	Y	N	N	Y	1	1	nSR							
31	M	5	60.2	55.2	7	N	Y	L	Abr	Y	N	N	Y	1	1	nSR	1	nSR	1	nSR			
32	M	14	45.9	31.9	6	N	Y	R	Abr	N	N	N	N	1	1	nSR							
33	M	11	57.9	46.9	10	Y	Y	L	Abr	Y	N	Y	N	1	1	nSR							
34	M	8	35.4	27.4	8	N	N	L	Abr	Y	Y	Y	N	1	1	nSR	1	nSR					
35	F	9	28.5	19.5	10	N	N	R	Abr	N	Y	N	N	2	1	nSR	4	SR	4	SR	4	SR	SR
36	F	14	22.2	8.2	6	N	Y	R	Abr	Y	Y	N	N	2	3	SR	4	SR	4	SR	4	SR	SR
37	F	23	44.7	21.7	7	N	N	R	Nor	N	Y	N	Y	2	1	nSR	1	nSR	4	SR		SR	SR
38	F	34	38.6	4.6	2	N	Y	L	Abr	Y	N	N	Y	2	2	nSR	2	nSR	2	nSR	1	nSR	nSR
39	F	3	54.4	51.4	11	N	Y	R	Abr	Y	N	N	N	2	3	SR	3	SR		SR		SR	SR
40	M	8	40.2	32.2	10	N	Y	L	Abr	Y	Y	N	Y	2	1	nSR	1	nSR	1	nSR			
41	M	0.75	46.4	45.65	8	N	Y	R	Abr	Y	N	N	N	2	3	SR		SR		SR		SR	SR
42	F	22	33.4	11.4	7	N	Y	L	Abr	Y	Y	N	N	2	2	nSR	3	SR	4	SR	4	SR	SR
43	F	16	51.8	35.8	8*	N	Y	R	Abr	Y	N	N	Y	3	3	nSF	2	nSF	2	nSF	2	nSF	nSF
44	M	13	37	24	7	Y	Y	L	Abr	Y	N	N	N	3	3	nSF	4	nSF		nSF		nSF	nSF
45	F	13	31.6	18.6	7	Y	Y	L	Nor	N	N	N	Y	4	4	nSF	4	nSF	4	nSF	4	nSF	nSF
46	F	50	68.1	18.1	5	Y	Y	R	Abr	N	N	N	N	4	4	nSF	4	nSF	4	nSF	4	nSF	nSF
47	M	14	47.7	33.7	15	N	N	R	Nor	N	N	N	N	4	4	nSF	2	nSF	2	nSF	2	nSF	nSF
48	F	18	46.9	28.9	7	N	Y	R	Abr	N	N	N	N	4	4	nSF	4	nSF		nSF		nSF	nSF
49	F	21	42.2	21.2	11	N	N	L	Abr	Y	N	N	Y	4	4	nSF	4	nSF	4	nSF	1	nSF	nSF
50	F	15	45.3	30.3	10	N	N	L	Abr	Y	Y	N	N	4	4	nSF	4	nSF	4	nSF	4	nSF	nSF
51	F	11	47.7	36.7	13	N	Y	L	Abr	Y	Y	N	Y	5	5	nSF	5	nSF	5	nSF	5	nSF	nSF

**Abbreviations and colour coding:** Sex: male (M)/female (F); Onset: epilepsy onset age in years; Surg.: age at surgery in years; Dur.: epilepsy duration in years; AEDs: number of anti-epileptic drugs taken preoperatively; SE: history of status epilepticus in yes (Y) or no (N); SGS: evidence

of secondary generalised seizures in yes (Y) or no (N); side of surgery left (L) or right (R); *MRI*: preoperative MRI pathology (normal/abnormal); *HS*: evidence of hippocampal sclerosis; *Depr.*: history of depression in yes (Y) or no (N); *Psychos.*: history of psychosis in yes (Y) or no (N); *Other Psych.*: history of other psychiatric disorder in yes (Y) or no (N); *ILAE yr1*: post-surgery seizure outcome at year 1, *ILAE 1 (good outcome) and ILAE 2 (favourable outcome) are in green, ILAE 3-6 (poor outcome) are in red*; *ILAE yr2, yr3, yr4, yr5*: post-surgery seizure outcome at year 2, 3, 4, and 5; *Rel. yr2, yr3, yr4, yr5*: post-surgery seizure relapse by year 2, 3, 4, and 5; *Rel. in 5 yrs*: post-surgery seizure relapse in five years after surgery. Patients with *no seizure relapse (nSR) are shown in blue, seizure relapse (SR) is shown in brown, and never seizure free (nSF) patients are shown in red*. In labelling seizure relapse, we excluded the 9 patients who had seizure recurrence within one year after the surgery (i.e., *ILAE 3-6 at year 1*) to avoid any bias.

\*Number of AEDs taken before surgery by patient 43 was imprecise. It was known that the patient took more than 4 AEDs but only 4 AEDs were reported precisely. We therefore imputed this imprecise data from the remaining information as detailed in Supplementary Methods. We illustrate in Supplementary Methods that the prediction performance remains similar with the imprecise AED information left unchanged, corrected, or with patient 43 removed from analysis.

Patients and controls were age and gender matched: female patient (%) = 60.7%; Female controls (%) = 58.6%; patient age at dMRI =  $38.51 \pm 11.53$ ; controls age at dMRI =  $37.65 \pm 12.40$ .