New insights for predicting surgery outcome in patients with temporal lobe epilepsy. A systematic review

M. Baciu, L. O'Sullivan, L. Torlay, S. Banjac

Version of Record: https://www.sciencedirect.com/science/article/pii/S0035378723008846
Manuscript_9c817c3467dfecfcd1abdd676ce47b60

New insights for predicting surgery outcome in patients with temporal lobe epilepsy. A systematic review

Monica Baciu[1], Laura O'Sullivan[1], Laurent Torlay[1] & Sonja Banjac[1, CA]

[1]Univ. Grenoble Alpes, CNRS LPNC UMR 5105, 38000 Grenoble, France

CA, Corresponding Author

Sonja BANJAC PhD

LPNC, UMR CNRS 5105

BSHM, BP 47

38040 Grenoble Cedex 09, France

Phone: +33 (0)4 76 74 81 60

Email: Sonja.Banjac@univ-grenoble-alpes.fr

Abstract

Resective surgery is the treatment of choice for one-third of adult patients with focal, drug-resistant epilepsy. This procedure is associated with substantial clinical and cognitive risks. In clinical practice, there is no validated model for epilepsy surgery outcome prediction (ESOP). Meta-analyses on ESOP studies assessing prognostic factors report discrepancies in terms of study design. Our review aims to systematically investigate methodological and analytical aspects of studies predicting clinical and cognitive outcomes after temporal lobe epilepsy surgery. A systematic review of ESOP studies published between 2000 and 2022 from three databases (MEDLINE, Web of Science, and PsycINFO) was completed by following the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines. It yielded 4,867 articles. Among them, 21 corresponded to our inclusion criteria and were therefore retained in the final review. The risk of bias was assessed using A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies (PROBAST). Data extracted from the 21 studies were analyzed using narrative synthesis and descriptive statistics. Our findings show an increase in the use of multimodal datasets and machine learning analyses in recent ESOP studies, although regression remained the most frequently used approach. We also identified a more frequent use of network notions in recent ESOP studies. Nevertheless, several methodological issues were noted, such as small sample sizes, lack of information on the follow-up period, variability in seizure outcome, and the definition of neuropsychological post-operative change. Of 21 studies, only one provided a clinical tool to anticipate the cognitive outcome after epilepsy surgery. We conclude that methodological issues should be overcome before we move towards more complete models to better predict clinical and cognitive outcomes after epilepsy surgery. Recommendations for future studies to harness the possibilities of multimodal datasets and data fusion, are provided. A stronger bridge between fundamental and clinical research may result in developing accessible clinical tools.

Keywords: epilepsy, surgery, prediction, post-operative, systematic literature review

1. Introduction

Epilepsy is a disease characterized by seizures that ILAE (The International League Against Epilepsy) defines as "a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain" [1]. It is estimated that 60% of all epilepsies are focal epilepsies [2] in which the abnormal neural activity starts locally and spreads by engaging other pathological and healthy brain areas [3,4].

The most prevailing treatment of epilepsy is administering antiepileptic medication to suppress seizure generation and propagation (for a recent review of antiepileptic medication, see [5]). However, almost 30-40% of patients are pharmacoresistant [6,7]. Moreover, the risk of pharmacoresistance is higher in patients with focal seizures compared to those with generalized seizures [8]. For those patients presenting with focal seizures, curative surgery to remove or inactivate the epileptogenic focus remains the only solution to stop seizures occurring [9,10].

However, these procedures are associated with substantial clinical and cognitive risks. Post-surgical clinical outcome refers to success in managing seizure status after the surgery. For instance, Mohan et al. [11] reported that 47% of patients remained seizure-free at five and 38% at ten years. However, these curative surgeries have cognitive risks as they can lead to changes in neuropsychological functioning, impacting patients' quality of life [12,13]. A meta-analysis performed by Sherman et al. [14] showed that 44% of patients with left temporal lobe epilepsy faced verbal memory disorders, and 34% experienced a naming decline after surgery.

Given multiple risks of epilepsy surgery, each surgical candidate undergoes a comprehensive neuropsychological and neuroimaging assessment before surgery. The main goal of this evaluation is a cost-benefit analysis that reveals the risks of cognitive decline versus potential seizure freedom [12,15]. Accordingly, there has been a considerable effort to predict and prognosticate clinical and cognitive outcomes of epilepsy surgery using patients' preoperative characteristics [12,16–18]. Indeed, numerous studies aimed to predict the post-surgical outcome in epilepsy patients (herein ESOP – epilepsy surgery outcome prediction) based on patients' preoperative characteristics such as clinical [19–21], cognitive [19] and cerebral data provided by anatomical [20,22,23], functional neuroimaging [24] and EEG recordings [20,25].

However, these studies yielded different results. For instance, some studies reported the duration of epilepsy to be predictive of post-operative seizure freedom [26,27], while others did not [28–30]. Similarly, some studies in epilepsy patients found preoperative verbal

memory to be predictive of post-operative decline of this function [31,32], while others did not [30]. Furthermore, some researchers found cortical language mapping to predict post-operative naming decline in temporal epilepsy patients [24], but not others [33].

One solution to overcome these differences is performing meta-analyses and systematic literature reviews to systematically synthesize and merge the results of different studies with the same topic to reach robust and valid empirical evidence [34–36]. Indeed, there are various meta-analyses of studies on ESOP [10,36–39] that tried to reconcile diverse findings by combining and weighting the results of different ESOP studies based on the study quality to find the most robust and valid evidence [40]. These studies mostly focused on post-surgical clinical [10,38,41,42] or cognitive outcomes [14,39,43] and identifying valid prognostic features clinical or cognitive epilepsy surgery outcomes [30,37,38,40,44,45]. Importantly, one common conclusion of these meta-analyses and systematic reviews is that one of the main reasons for different findings across ESOP studies are differences in employed methodology [37,40,45]. For instance, a characteristic that differentiates between ESOP studies is the number of participants used to make the prediction model. This is a crucial point given that small sample sizes can lead to sampling bias [46]. Indeed, Armon et al. [29] warned that large sample sizes are needed (preferably from multiple institutions) for reliable identification of predictors for epilepsy surgery and external validation of these predictive models. Moreover, with the development of artificial intelligence, there is an increased interest in applying machine learning (ML) approaches for predictive studies of epilepsy [47]. This trend could be explained by ML's ability to identify complex relationships between large, multimodal data sets and parameters [47–49]. This is particularly true for epilepsy surgery since physicians (neurologists and neurosurgeons) base surgery decisions on multimodal data obtained with various neuroimaging methods (functional magnetic resonance imaging (fMRI), positron emission tomography (PET), electroencephalography (EEG)) and neuropsychological assessments for every preoperative evaluation. Consistent with this practice, studies show that leveraging the information obtained through different modalities may help to highlight new patterns difficult to observe or effects too weak to be detected, if only one isolated modality is used [50].

Nevertheless, there has been only one systematic literature review that dealt with methodology of ESOP studies. In 2001, McIntosh et al [36] analyzed methodological approaches of ESOP studies of that time. These authors concluded that ESOP studies face several issues such as small sample size and variability of seizure freedom definition. Recently, Yuan et al. [26] provided a valuable review of ML applications in diagnosis and

surgery prognosis for epilepsy patients. However, their review is not systematic, and it focuses only on ML and neuroimaging techniques, not addressing the methodological variability of these studies. Another point that has not been considered by the existing ESOP meta-analyses and systematic reviews is the clinical applicability of the findings. As pointed out recently by Busch et al. [31], although there are a number of models for predicting memory decline after temporal lobe resection, they are not adapted for widespread clinical use and are not easily accessible to physicians.

This narrative descriptive review aims to fill in that gap and provide a systematic and comprehensive synthesis of the methodology used in ESOP studies.

We specifically focused on the methodology used in those studies with the following questions:

1. What type of data is used in those studies as predictors of epilepsy surgery outcome, and is there a trend to shift from unimodal or multimodal datasets?
2. What were the characteristics of samples, outcome variables and analytical approaches in the ESOP studies?
3. How many ESOP studies provide tools for the clinic reinforcing thus a research-practice bridge?

We formulated the following hypotheses related to each goal:

1. With the development of neuroimaging techniques and their increased application in clinics [51] and neuroimaging studies [52], we expected that the multimodality of predictors in ESOP studies are increasingly reported.
2. Given the recommendations [36] and the development of power calculation tools (e.g., [53,54]), we expected the sample size of ESOP studies is growing with time. We anticipated that the ESOP studies focusing on clinical outcomes use standardized categories, following recommendations [36,55,56] and widely excepted taxonomies [1,57], while those focusing on cognitive outcomes use metrics that avoid measurement error [55]. Moreover, in line with our expectations regarding larger and richer datasets [49] and the potential of the ML approaches for data integration [47,48], we expected this approach to be more dominant in newer and multimodal ESOP studies.
3. We expected to identify a significant number of ESOP studies that provide clinical tools, as they are motivated by a practical question and goal (predicting the outcome of epilepsy surgery in day-to-day practice).

The present literature review is useful for several reasons: *i*) it can inform future predictive studies about the available methods and models; *ii*) it suggests that the change of

methodology over the years should illustrate a shift in the conceptualization of epilepsy (for instance the increase of network measures related to modern connectomic perspectives [23,58]); and *iii*) it informs physicians who have difficulties identifying potential epilepsy surgical candidates [59] on the studies that provide practical tools which *iv*) could hopefully encourage future studies towards limiting the gap between cognitive neuroscience evidence and its application in clinical settings. A bridge between these two fields of knowledge can contribute to more adapted individual care and better surgical outcomes. Following this notion, Box 1 presents reminders and precision of some methodological and technical terms used in the present paper that may only be familiar to some readers.

2. Methods

2.1. Study design

This systematic narrative literature review was performed following Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [34,60]. We were interested in studies aiming to predict clinical and cognitive outcomes of resective surgery in patients with drug-resistant temporal lobe epilepsy (TLE). We decided to focus specifically on TLE rather than focal epilepsies in general for several reasons. First, since surgical procedures in different focal epilepsies (e.g., frontal and temporal) focus on different regions, they have diverse outcomes and different rates of positive surgical outcomes [61]. Therefore, the predictions of surgical outcomes could rely on different pre-operative features depending on the type of focal epilepsy (e.g., normal MRI can be a negative prognostic factor in frontal but not TLE patients [62]). Furthermore, related to previous arguments, we chose to focus on temporal lobe epilepsy because it is the most prevalent focal epilepsy [3,63]. We defined *clinical outcome* as seizure freedom after surgery in terms of Engel's or ILAE classification [1,57,64]. The neuropsychological performance after surgery was taken as the *cognitive outcome*. The search was conducted on MEDLINE, Web of Science, and PsycINFO databases (the strings used for this search are presented in Appendix 1 : Table S1).

2.2. Inclusion criteria

We included in this systematic review only studies that fulfilled the following criteria:

A. Population

a. Papers including adult patients between 18-65 years of age. Studies with older adults > 65 years were not included because of comorbidities and antecedents, like neurodegenerative diseases, which are present more often in the elderly [65], can influence surgery outcome and hence, the choice of the modality of predictors in

ESOP studies. Furthermore, post-surgical recovery of cognitive functions in older epilepsy patients can be more limited compared to younger patients [66]. Children were excluded as the characteristics of pediatric epilepsy and recommendations for pediatric epilepsy surgery differ from those for adults, which could influence ESOP models and the modality of included predictors [67–69]. Moreover, if the study did not provide an age range, we verified that the mean ± standard deviation fell in the 18-65 range.

b. Patients with a diagnosis of drug-resistant TLE
c. Patients undergoing first-time surgery (curative surgery)

B. Surgery

a. Studies predicting resective surgery (partial and complete lobectomies) outcomes

C. Methodology

a. Studies predicting post-surgical clinical (seizures) and cognitive outcomes, quantitatively and 12 months after surgery at the earliest. Studies with follow-up periods shorter than 12 months were not included because of the stability of the observed changes [70–73] and the influence of variables related to recent surgical interventions, which could impact the selection of predictors. Furthermore, the long-term outcomes are the overall interest of the surgical intervention [74].
b. Studies reporting post-surgical clinical outcomes according to ILAE [75] or Engel taxonomy [76] (or other taxonomy with feasible correspondence to ILAE or Engel)

D. Language of publication

a. Studies published in English

E. Publication

a. Peer-reviewed articles
b. Papers published between January 2000 and May 2022.

2.2. Exclusion Criteria

We excluded from this systematic review the studies that fulfilled the following criteria:

A. Population

a. Patients with other focal epilepsies (e.g., frontal, occipital, etc.)
b. Patients with focal epilepsy as a complication/comorbidity (e.g., secondary to tumors; gliosis [77])
c. Patients who underwent multiple surgeries
d. Animal studies

e. Studies that did not provide sufficient sample details (e.g., did not specify sample' age range/mean, or the time from surgery to measurement of postoperative outcomes)

B. Surgery

a. Procedures other than surgical resection (e.g., callosotomy, neuromodulation, thermocoagulation, ablative therapies etc.)

C. Methodology

a. Patients with clinical and cognitive outcome reported before 12 months after surgery
b. Longitudinal studies (i.e., studies that followed the change in clinical and cognitive characteristics through time and not surgery-related)
c. Studies using post-operative predictors (information that is not available before the surgery e.g., existence of seizures post-operatively)[1]
d. Studies that did not use an analysis specifically predicting a post-surgical outcome (e.g., studies that compared seizure-free and non-seizure-free patients but did not use an analysis that aimed to predict the post-surgical outcome)
e. Studies that predicted any criterion variable other than the outcome of surgery (e.g., decision whether or not a patient will be operated)

D. Language of publication

a. Studies published in languages other than English

E. Publication

a. Patient samples published in more than one paper
b. Meta-analyses, conference proceedings, abstracts, pilot study, protocol, dissertations, letters, brief reports, statement papers, case reports, books, and pre-prints.

2.3. Search strategies and study selection process

The search was conducted using a combination of free-text and Medical Subject Heading (MeSH) terms (the complete search strings are presented in Appendix 1 : Table S1) on MEDLINE via PubMed, Web of Science, and PsycINFO via EBSCOhost. The database searches were conducted on the same day (June 3rd, 2022) to control for daily updates. The study selection process steps followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [34,60]. The screening was performed using Rayyan [78]. After duplicate removal, two researchers screened the title and abstract of all papers identified to exclude papers that did not meet the inclusion criteria. In the following

[1] We did not exclude studies that used the resection area (e.g., whether it included the activated region) as a predictor since that information can be known before the surgery while planning for it.

phase, full texts were screened for exclusion criteria, and the final selection of the papers that entered the review was made.

2.4. Data extraction process

Two researchers independently extracted data from the final review papers. Consensus among researchers was established for all studies. Based on the study's questions (see section 2.1.), the following information was extracted from each paper:

1. Modality of the investigated predictors and feature(s) (e.g., MRI, task fMRI, resting-state fMRI (rs-fMRI), diffusion-weighted imaging (DWI), EEG, PET, neuropsychological tests scores, biomarkers)
2. Outcome (clinical, cognitive with specific measures)
3. Population (characteristics, sample size)
4. Analysis used for outcome prediction (e.g., ML, logistic regression, discriminative analysis, etc.) and its details (e.g., was there a phase of feature/predictors selection, was model evaluation or validation performed)
5. Proposition of a tool developed for clinical practice (e.g., nomograms, website, etc.).

2.5. Strategy for data synthesis

Due to the variability of the methodology and analytical approaches employed in the studies, we performed a narrative review of the results and reported descriptive statistics of the selected studies (similar to [36,39]). In addition, we opted for descriptive statistics as the goal of this study was to describe the methodology of ESOP studies instead of concluding on the best predictors of post-surgical outcomes, which would require appropriate meta-analysis (for instance, see [10,36–39]).

2.6. Risk of bias

We assessed the risk of bias (RoB) and the applicability of the prediction models of the studies we identified using the PROBAST (Prediction model Risk Of Bias Assessment Tool) framework [79–82]. Following the PROBAST, RoB and applicability of four domains were evaluated for each study – participants, predictors, outcome, and analysis. Each study's overall RoB and applicability were evaluated based on these domains. One of the authors evaluated the selected studies using the PROBAST framework, and the other author verified the obtained evaluations.

3. Results

3.1. Overview of PRISMA flowchart

Figure 1 illustrates the steps of the selection process according to PRISMA guidelines [34,60]. Our literature review of the three databases using search strings (Supplementary

Material A) yielded 4867 articles. First, all duplicates (n = 2544) were removed. Two researchers screened the title and abstract of all papers identified, excluding papers that did not meet the inclusion criteria. After this phase, 576 papers were excluded.

In the second phase, the authors read the remaining 211 studies. Full texts of these studies were reviewed, and a further 190 studies were excluded. In the case where full texts were not available, the publishing authors were contacted. The reasons for exclusion were documented. Disagreements were resolved through a discussion with a third researcher. This phase led to a sample of 21 papers in the final review [22,83–102]. Selected papers were published between 2000 and 2021. Given the present study's interest in trends in ESOP studies over time, the papers were separated into those published between 2000-2010 (n = 6) and those during the 2011-2021 period (n = 15). These two periods were obtained by dividing the period of identified studies (i.e., 2000 – 2021) into two halves and do not reflect a specific turning point. A finer period separation was not possible, given the number of identified studies. Table 1 provides ESOP study characteristics and an overview of the extracted data.

3.2.Modality of the investigated feature(s)

We used the term “modality” for each “acquisition framework” [103] or the type of data that the study was using as a predictor. We categorized studies as unimodal if their predictors of post-surgical outcomes were obtained through the same type of acquisition (i.e., modality, for instance, DWI) and as multimodal if the predictors were acquired through at least two different acquisition frameworks (for instance DWI and neuropsychological scores)[2].

Within our sample of 21 selected ESOP studies, 12 used a unimodal dataset to predict post-surgical outcomes, and nine used a multimodal dataset. Panel A of Figure 2 shows the percentage of studies using unimodal or multimodal datasets in the two time periods (Table 1). As the figure illustrates, while studies mainly relied on one modality at the beginning of the 2000s (only one multimodal study), from 2011 onwards, there was a growing trend of combining predictors from different modalities (eight multimodal studies). The same trend was observed when we focused on studies that only predicted clinical outcomes (Panel A, Figure S1, Supplementary Material). Half of the studies predicting cognitive outcomes used unimodal datasets, and the other half multimodal datasets. We did not analyze dataset change through time for selected ESOP studies predicting cognitive outcomes as there was only one study during the 2000 – 2010 period.

[2] Studies were classified as unimodal regardless of the number of predictors, as long as they were all the same modality. Studies were classified as multimodal even if predictors were in the same analysis, as long as they were of different modalities.

Panel B of Figure 2 depicts the precise modalities (i.e., types of data) employed in the 21 identified ESOP studies. Predictors such as anatomical MRI, task-fMRI, DWI, microRNA, and neuropsychological test scores were identified only in the studies published after 2011. Results also indicate that after 2011, more diverse predictors were used. However, more studies were identified in 2011-2021 (n = 15) than in 2000-2010 (n = 6).

Multimodal studies included, on average, three different modalities (SD 0.94, range 2 – 4 modalities). Studies with at least two modalities usually combined a neuroimaging technique with clinical-demographic information [88] or neuropsychological scores [87,97]. The most multimodal datasets combined four modalities. Bonilha et al. [89] used clinical-demographic data, anatomical MRI, DWI, and surface EEG, Baxendale et al. [84] included clinical-demographic, neuropsychological scores, anatomic MRI, and surface EEG[3], while Larivière et al. [86] combined clinical-demographic, anatomic MRI, rs-fMRI, and DWI.

3.3. Outcome

Among 19 selected studies that predicted clinical outcomes[4], nine used Engel's taxonomy [85,86,88–90,92,94,95,98–100,102], six used ILAE taxonomy [22,83,84,91,93,96], while one study grouped seizure-free and non-seizure-free patients, not relying on any standardized taxonomy [87]. Despite this, naming and categorizing outcomes varied across studies (see Table 1). The majority of studies [86,88,90,92,94,99,100] using Engel taxonomy divided patients into those that were seizure-free (class 1) and non-seizure-free (classes 2-4), also termed good versus bad outcomes [88]. Three studies classified as seizure-free only patients with Engel class 1a [95,98,102], while one also included patients with class 1b [89]. All studies classified patients with class 2 and higher as non-seizure-free, except Ioriatti et al. [85], who included only patients with class 3 and higher, and Gaca et al. [95], who included class 1b in the non-seizure-free group as well. Similar diversity was present in studies that used ILAE taxonomy. While one part of the studies grouped classes 1 and 2 as seizure-free or good outcomes and other classes as non-seizure-free [22,83,84], others only considered patients with class 1 [91,93] or class 1a [96] as seizure-free.

There was also diversity between studies predicting post-operative cognitive outcomes regarding how this outcome was defined. Two studies used the change score of the chosen neuropsychological test as the difference between pre-operative and post-operative scores

[3] Specifically, these authors included the information on whether the surface EEG was non-localizing with the requirement for SEEG as a proxy of non-lateralizing EEG.

[4] We counted together those that focused exclusively on clinical outcomes (n = 17) and those that predicted both clinical and cognitive outcomes (n = 2).

[101,102], although [102] also expressed the post-operative change (i.e., improved or unchanged versus deteriorated) qualitatively. Similarly, Paff et al. [97] subtracted post-operative scores from pre-operative scores, but they used principal component scores instead of raw test scores, converting them to z-scores. On the other hand, Roger et al. [94] divided patients into those with good (improved or unchanged) and poor outcomes based on the significant post-operative change in neuropsychological performance by using the reliable change index.

We only included studies that reported post-surgical clinical and cognitive outcomes at the earliest 12 months after surgery. However, we did not define the duration of the post-surgical period. It was impossible to calculate the average post-operative period during which identified studies collected data on surgical outcomes because 11/21 reported mean and variation of the post-operative period, while 10/21 did not adequately provide this information. Those that did not provide adequate information used descriptive statements (e.g., "approximately 12 months after surgery" [101]) or did not state the maximal post-operative period at which the data was collected (e.g., "at least one year follow up" [92]).

3.4. Population

The mean sample size across all identified studies was 76.42 (SD 71.97, range 20 – 275 participants). The average sample size for the 2000 - 2010 studies was 94.17 (SD 87.22, range 26 – 252 participants) was relatively higher than that of the 2011 – 2021 studies (M 69.33, SD 67.04, range 20 – 275 participants). Although it should be noted that sample sizes variation was high in both periods, for the studies predicting clinical outcome the mean sample size was 81.58 (SD 73.86, range 26 – 275), and for those predicting cognitive outcome it was 60 (SD 53.16, range 20 – 138). With respect to dataset, the mean sample size of the unimodal studies was 70.83 participants (SD 66.30, range 26 – 252) and of multimodal studies 83.89 (SD 82.44, range 20 – 275). In Appendix 1, Table S2 presents the distribution of sample sizes in relation to the specific modalities. On average, the 2000-2010 studies that used clinical-demographic variables had the largest samples (M = 158, SD 132.94), for which the mean inclusion period was 9.5 years (see Appendix 1 : Table S2 for more details), while the smallest samples were identified in studies using task-fMRI (n = 1, M = 20) and rs-fMRI (n = 2, M 25;,SD 7.07).

The average time of patient inclusion was 8.54 years (SD 6.36, range 3 – 28) for all studies, 8.33 years for studies predicting clinical outcomes (SD 6.6, range 3 – 28), and seven years for those predicting cognitive outcomes (SD 3.46, range 5 – 11). The average inclusion time for multimodal studies was 11.83 years (SD 8.13, range 6 – 28) and for unimodal studies 5.71

(SD 2.36, range 3 – 9 years) We did not find any association between the inclusion time and sample size.

Most studies (n = 14) collected data from a single institution (e.g., university center or hospital), while two collected data from two centers [88,92], and five studies did not explicitly report this information. Sample of multi-center studies included on average 158 TLE patients (SD 132.94, range 64 – 252) with the mean inclusion period of 9.5 years (SD 0.71, range 9 – 10), while the single-center studies included on average 71 TLE patients (SD 69.29, range 20 – 275) for 9.33 years (SD 7.37, range 4 – 28).

Left lateralization of the epileptogenic zone or network was reported in 62.92% of patients on average (SD 18.99, range 36.67 – 100), specifically 60.29% (SD 17.26, range 36.67 – 100) in studies with clinical outcomes and 71.61% (SD 21.42, range 51.06 – 100) in studies predicting cognitive outcomes. Two studies included exclusively left TLE patients [22,101], while six studies did not provide this information.

3.5. Analysis used for outcome prediction

Most of the identified ESOP studies used linear [97,99,101] or logistic regression [84,88,92,95,96,100,102] or a standard statistical approach such as discriminant analysis [90,98] or receiver operator characteristic (ROC) curves [85,87,93][5] as principal analyses. Six studies used ML analyses [22,83,86,89,91,94], among which the most frequent was the support vector machine (SVM, 4/6 studies) (Table 1). While regression and discriminant analysis were present in the studies published between 2000 and 2010, ML studies were identified only in the 2011 – 2021 period.[6]

Ten ESOP studies tested the predictive power of the predictors chosen beforehand based on a theory or previous research. The rest of the studies used an analytical method to select predictors or features for the final model among a group of possible ones. To that end, most studies [93,98,100–102] performed a univariate analysis for each potential predictor before adding them to the multivariate analysis or used a step-wise regression [88,90,96]. Four ML studies used feature selection. Specifically, Feis et al. [22] used Fisher's criterion, Bonilha et

[5] Although ROC curves can be used as a part of the ML approach to characterize the sensitivity/specificity of a classifier, we separated these three studies into a specific group because they used ROC curves as a principal analysis. These three studies used ROC with specific values/scores rather than evaluating a classifier model's prediction. Rathore et al. [87] used ROC to calculate the predictive value of different Wada and neuropsychological scores. Keller et al. [93] used ROC to calculate the predictive value of "along-the-tract profile" values. Ioriatti et al. [85] used ROC to calculate the predictive value of miR-328-3p expression profiles.

[6] The distinction between ML and statistical modeling is widely discussed and not always evident. While regression can be considered a ML algorithm, we separated it in this paper similarly to [49]. Moreover, the ESOP studies whose analyses were classified as "regression" did not declare any use of a ML approach. However, if the authors used regression among other algorithms, and explicitly stated that they applied the ML approach (such as [94]), we classified the analysis of the study as "ML".

al. [89] used independent cross-validation, and Sinha et al. [91] ranked features based on importance for prediction through analysis repetition, removing the least important feature each time. In contrast, Roger et al. [94] used penalized linear model and held-out fold repeated 500 times.

Except for Dupont et al. [98], who developed their prediction model and validated it on the new group of patients, all identified studies developed models without external validation. All ML studies performed model evaluation (also called internal validation by some authors [104]) to avoid overfitting and to account for optimization, using different forms of cross-validation such as k-fold [83,89,94], stratified k-fold [86], leave-one-subject-out [22], and nested cross-validation [91], which can also evaluate the prediction of the unseen data.[7]

We also explored if unimodal and multimodal studies differed in the employed analytical approaches. Regression studies had, on average, 1.8 modalities of predictors, and ML studies had 2.3. Table 2 shows the distribution of the analyses in relation to the dataset used in all identified studies and separately for those focusing on clinical and cognitive outcomes. The descriptive results suggest a slightly higher application of ML analyses and regression in multimodal studies compared to unimodal studies in general (ML 33.3% versus 25%; regression 55.6% versus 41.7%, respectively), although regression remained the most used analysis regardless of the dataset used. However, the multimodal ESOP studies focused on clinical outcomes employed the ML approach more than the ones using unimodal predictors (42.9% versus 25%). Moreover, in multimodal ESOP studies focused on clinical outcomes, the ML approach was as frequent as regression (42.9% both). On the other hand, multimodal ESOP studies predicting cognitive outcomes used regression exclusively, while unimodal ESOP studies used both ML and regression analyses.

It should also be noted that one study published between 2000-2010 and seven between 2011-2021 echoed the idea of network and connectivity in their methodology, either by using DWI or rs-fMRI [83,86,89,91,93,101], calculating graph metrics [83,86,94,101] and connectome gradients [86] or comparing the predictive power of a network and individual regions [98].

[7] We categorized these studies as "development only" rather than "development and validation" because they did not include an evaluation of predictive performance in data external to the development sample. PROBAST framework does not recognize splitting a sample into train and validation subsamples (as is done in cross-validation) as a form of external validation.

3.6.Proposition of a tool developed for clinical practice

Among identified studies, only Roger et al. [94] provided a tool that can be applied in clinical practice. These authors provided a nomogram based on the logistic regression results using the cognitive features selected via ML analysis.

3.7. Quality assurance/bias

Table 3 reports the RoB and applicability assessment of the 21 ESOP studies based on the PROBAST tool [79,82]. Details on the criteria used to make these summary judgments can be found in Table S3 in the Supplementary Material. The overall RoB and applicability of all studies were found to be acceptable. As shown in Table 3, all studies had low bias concerns for participants, predictors, and outcomes, and the applicability of these domains was also acceptable. However, the RoB for analysis was mixed across the studies. Ten studies were rated "unclear RoB" for the analysis performed. According to PROBAST guidelines, only six studies [22,87,90,92,100,102] had a reasonable number of participants with the outcome for the number of predictors included in the model. Additionally, as mentioned (section 3.5), six studies [88,93,98,100–102] selected the predictors for the model using univariate analysis. Nevertheless, the PROBAST tool suggests that this method increases the possible bias in the prediction. Moreover, 14/21 studies did not include information regarding overfitting or optimism in their model performance. Finally, while studies did report candidate predictors [82], only Roger et al. [94] explicitly reported the predictors included in their final model. However, these authors did not provide details such as the weights of the predictors in the model.

4. Discussion

A thorough pre-operative assessment is performed for each epilepsy surgical candidate to weigh potential seizure freedom over risks of cognitive decline, as this procedure is associated with considerable clinical and cognitive risks. Nevertheless, clinical practice still lacks a widely validated model for predicting post-surgical outcomes in epilepsy patients [105]. Meta-analyses on predicting post-surgical outcomes in epilepsy patients often explained the discrepancy in findings with differences in employed methodology [36,37,40,45]. Therefore, we performed a systematic literature review using three databases (MEDLINE, Web of Science, and PsycINFO) focusing on methodological and analytical aspects of studies predicting clinical and cognitive outcomes after TLE surgery. The data was analyzed narratively and using descriptive statistics due to the number and high variability of the identified studies. To our knowledge, this is the first study since the study of McIntosh et al.

[36] in 2001 to systematically examine the methodology and analytical approach of ESOP studies. Our analysis is based on 21 identified studies with 1605 TLE patients. All studies' RoB and applicability were acceptable according to the PROBAST framework [82], so they were all included in the final review. In line with our goals, we will first discuss the modality of the predictors used by the ESOP studies examined here. We will then focus on sample characteristics, outcomes, and analytical methods. We will finish by addressing the clinical tools provided by the identified ESOP studies.

4.2. Modality

Since epilepsy is a complex disorder with various neurobiological and neuropsychological characteristics, no single modality can provide an exhaustive evaluation of a patient's condition [87]. Therefore, prediction of epilepsy surgery outcomes in the clinical setting is inherently multimodal [87,104]. Our expectations regarding the increase of multimodal ESOP studies with time were confirmed, as the majority of multimodal studies were published after 2011 (9/10). In the only multimodal study published in the early 2000s, Wong et al. [88] combined the clinical and FDG-PET (18-fluorodeoxyglucose PET) predictors using a multiple logistic regression model. A similar approach was identified in the studies published after 2011 that combined the predictors of different modalities by including them in the same regression analysis [84,100,101,106] or using them all as features for SVM [91]. Another approach was to test the predictive value for post-operative outcome for each modality separately without combining them. Rathore et al. [87] implemented such an approach by performing separate ROCs for Wada and neuropsychological scores, and Paff et al. [97] by executing separate regression analyses for MRI volumetry and neuropsychological scores. Finally, two studies fused modalities. Bonilha et al. [89] computed a single score for each patient by combining network architecture abnormalities and composite binary clinical variables. The predictive value of this model outperformed models based on unimodal data. Lariviere et al. [86] used "functional connectivity distance", which combined rs fMRI and geometric brain network properties to predict post-surgical seizure freedom. The model based on functional connectivity distance features outperformed the one combining standard clinical and structural neuroimaging parameters in predicting clinical post-surgical outcomes. Nevertheless, the authors did not test a combined model. Similar to these findings, in a recent meta-analysis, Alim-Marvasti et al. [40] noted that correspondence between different measures has the potential for predicting post-surgical outcomes. These findings corroborate that data fusion is more than the sum of parts and contains information that could be lost if the

relations between modalities are ignored [103]. Calhoun and Sui [50] proposed a spectrum of data fusion approaches (for more technical and mathematical details on multimodal fusion, see [52,103]). These authors showed how moving from a simple visual inspection of two modalities toward a symmetric fusion of multiple modalities increases information gain. Although Calhoun and Sui [50] used this continuum in the case of neuroimaging data, their idea can be expanded to clinical-demographic and neuropsychological data. For instance, including in a post-surgical predictive model separately, the volume of the hippocampus and delayed neuropsychological memory score would not be the same as including the correlation between them. The latter can be considered data fusion because both datasets are used equally. In some cases, asymmetric fusion is more appropriate given the research question, such as analyzing the functional connectivity of to-be-resected regions. Therefore our recommendation for future ESOP studies is to test the predictive potential of the interaction of modalities or to use "fusion metrics" (that combine symmetrically or asymmetrically two modalities in a single measure) as it has shown promising results [86,89].

4.3. Outcome

Another methodological aspect investigated in our study was the outcomes that ESOP studies tried to predict. Most of the identified studies (n=19) predicted post-surgical seizure freedom. All of those studies except [87] used standardized taxonomy, either Engel's [85,86,88–90,92,94,95,98–100,102] or ILAE [22,83,84,91,93,96]. However, we identified five different ways of dividing Engel's class into seizure-free and non-seizure-free outcomes and four for ILAE. These differences can limit the comparability of studies and the generalization of their findings [36]. Furthermore, these variabilities in outcome definition can also present an issue for future studies aiming to validate proposed ESOP models. Although we identified variability in outcome defining, it can still be seen as a positive trend compared to findings of previous systematic reviews [10,36] as all except one study [87] used one of two standardized taxonomies. Therefore, our recommendation for future ESOP studies is to use one of the proposed taxonomies and sustain from author-defined classifications.

Identified studies that predicted cognitive outcomes mostly used the difference between pre- and post-surgical test scores. Specifically, Uijl et al. [102] calculated the post-surgical change in verbal IQ and performance IQ of the Wechsler Adult Intelligence Scale, Audrain et al. [101] of the Boston Naming Test, while Paff et al. [97] calculated the change of principal component scores of verbal and visuospatial memory obtained using a battery of tests [97]. On the other hand, Roger et al. [94] calculated the Reliable Change Index of the DO80 test

scores. Previous studies pointed out that raw change scores can be vulnerable to practice effects and regression to the mean [107]. Two approaches can help investigate the effects of epilepsy surgery on cognitive functioning independent of test-retest artifacts – reliable change indices and standardized regression-based methodologies [55,56,107]. Therefore, to evaluate the "meaningful post-operative cognitive change" [55] that overcomes the probable range of measurement error [56], we recommend that future ESOP studies focusing on cognitive outcomes use one of the two suggested approaches to avoid potential biases.

Another issue we identified in the ESOP studies concerning the outcome was the follow-up period. Namely, almost half of the identified studies (10/21) did not adequately report the post-operative time, and there is no information if the outcome was measured only within the first few post-operative years or if it was a longer follow-up. Moreover, a frequent reason for excluding the studies from our review was the short follow-up period. Also, in some studies, the follow-up period was very variable among patients (for instance, in [106], it varied from three to 32 months, which was not controlled for in the ESOP model). The follow-up information is crucial for clinical outcomes since the chance of seizure remission is related to the duration of the follow-up period [37]. The lack of this information prevents investigating specific factors related to short-term and long-term post-surgical outcomes, exploring if mechanisms differ with respect to the length of the post-operative period, and informing rehabilitation planning [36]. Hence, the recommendation for future ESOP studies is to provide complete information on the post-surgical follow-up period of epilepsy patients.

4.4. Population

We hypothesized that ESOP studies' sample size would grow from the 2000s onwards. Nevertheless, we did not observe such a trend, and the average sample size even decreased comparing the 2000-2010 and the 2011-2021 studies. Based on the PROBAST standards [79,82], only six out of 21 ESOP studies had a reasonable number of participants relative to their model design. The performance of a prediction model can be overestimated with smaller sample sizes [29,36]. Small sample sizes can lead to biases, especially in studies that use univariate analysis for predictor selection for the final model (this will be discussed in the 4.4 section). PROBAST suggests that the number of patients per variable (here, the variable refers to the predictors considered during any stage of the prediction model process, not only the final model) should be at least 20 [82]. However, it was found that ML algorithms such as SVM require a higher number of participants per variable to avoid overfitting [108]. Nevertheless, the sample size needed to avoid overfitting depends on context, outcome

prevalence or risk, the number of candidate predictor parameters, and the anticipated overall model performance. Discussing results on small sample sizes in ESOP studies, McIntosh noted in 2001 that power calculations are uncommon in this area of research [36]. Since then, various recommendations and tools have been developed for sample size prediction (for instance, [53,109]). Recently Riley et al. [110] provided a package that implements the recommendations and allows calculating the minimum sample size specifically for prediction model development.

One strategy for increasing sample size that we believe some studies adopted was looser inclusion criteria, such as age (and follow-up period, see section 4.2). Namely, during the study selection process, in the full-text review process, the most frequent reason for the exclusion of ESOP studies was the age of the participants (Fig. 1). Specifically, studies tended to include participants younger than 18 years old. However, children and adult samples should be separated for several reasons. The neurobiological features of epilepsy in children and adults differ [67,111]. Furthermore, in children, the neuropsychological changes following surgery are influenced by ongoing maturation changes and higher plasticity at the behavioral and structural levels [69]. It was also shown that children and adults differ in post-operative seizure freedom rate [112] and factors related to surgery failure [113]. Therefore, our recommendation for future ESOP studies is to have as homogenous a sample as possible or to include in the model the characteristics by which they differ. Additionally, ESOP studies should test predictive models of epilepsy surgery separately in children and adults and use available tools to calculate the necessary sample size. Larger samples can be more easily achieved if multiple research centers are included [36]. Indeed, our results showed that studies that used samples from several medical centers had a larger mean sample size than the ones relying on only one center, while the inclusion time was similar. Nevertheless, this finding should be taken with caution, given that we identified only two multicenter studies. Furthermore, in addition to providing relatively large samples, multicenter studies can introduce methodological challenges related to imaging acquisition across sites, under-recruitment at single sites, and lack of common measurements [114].

4.5. Analyses

Our results show that regression and ML were the two primary analyses in the ESOP studies. Our descriptive statistics showed that the ML approach was used more often in multimodal ESOP studies. Indeed, ML was claimed to be better adapted than classical statistical modeling for larger datasets [48–50]. Furthermore, ML includes nonlinear complex associations and

interactions between predictors (expected in a large multimodal dataset) more automatically than regression [49,115]. However, a recent systematic review did not find ML better than regression for performance in clinical prediction models [49]. It was also found that ML needs more data than regression to achieve a stable area under the ROC-curve and small optimism [108], although other studies suggest that ML methods are more appropriate for datasets with more variables than the number of subjects [115]. Therefore, analysis choice in ESOP studies should be based on the underlying hypotheses of the model and the dataset. Authors should opt for regression when their model is based on theory and human knowledge (e.g., experts) and when they are focused more on inference by creating and fitting a specific predictive model. On the other hand, ML should be used in data-driven studies which focus on finding the best learning algorithm based on patterns of rich data to predict the outcomes of unobserved cases. Since integrating multimodal data seems to be a growing trend in ESOP studies and since multimodal fusion is still being explored, ML approaches seem to be a good option.

An essential aspect of the model development is the selection of predictors. Our results showed that almost half of the identified ESOP studies tested the predictors chosen before the analyses, while the others selected predictors based on their predictive value. Nevertheless, the PROBAST framework warns that the univariate analyses of predictors before their inclusion in the final multivariate analysis should be avoided as it can lead to incorrect selection. Moons et al. [82] point out that the bias can arise because, in this case, the predictive power is evaluated as a single predictor and not in the context of other predictors. Moreover, there is also a possibility that the univariate associations with the outcome and individual predictors are accidental. ML approaches such as penalized linear mode with held-out fold used in [91] and removing the least important feature through analysis repetition used in [88] could prevent this, as all predictors are analyzed together. The other possibility is to use strategies that omit predictors, such as backward selection, as used in [85,87,93]. On the other hand, this strategy can lead to potential overfitting of the model to the dataset, so it should be followed by validation and optimism adjustment strategies [82].

Studies developing predictive models should address the optimism of their predictive performance. In other words, if their models are not too much adapted to the development dataset. For that, methods such as cross-validation are usually used, as was identified in all ML studies in our sample. However, this model optimization should not be confused with its validation [116], as a predictive model's final goal is to predict unseen data accurately [22]. Some ML studies in our sample addressed this issue through their choice of the cross-

validation method, such as leave-one-subject-out used by Feis et al. [22], nested cross-validation used by Sinha et al. [91], and stratified k-fold used by Larivière et al. [86] in which models can be evaluated in previously unseen data. However, some authors [82,104,117] consider splitting datasets as internal rather than external validation. A prediction model can show excellent performances in the sample used for its development, but before it is introduced into clinical practice, it needs to be validated in another sample that was not used for its development [104]. In our sample, we identified only one such study. Dupont et al. [98] first identified the prediction pattern of network hypometabolism in the group of patients with good and worst outcomes after anterior temporal lobectomy. Then, these authors validated their equation in the group of TLE patients with the intermediate post-surgical prognosis (equivalent to Engel's class 1b, c, and d). It is also suggested that external validation should be performed in a separate study by different researchers [117]. Nevertheless, no study validated an existing model among the identified ESOP studies. Therefore, our recommendation for future ESOP studies is to either validate the existing ESOP models and, if necessary, propose their modifications based on the results, or to include external validation if they develop a new model (in addition to internal validation). Once again, the solution for acquiring different samples of patients could be multicenter collaboration [36].

Our literature review findings also suggest that from 2011, the notion of networks was more present in the methodology of ESOP studies. Among studies in the early 2000s, only Dupont et al. [98] addressed the network notion through their study design. These authors showed that the pattern of hypometabolism of a network of connected regions outperformed the hypometabolism of individual temporal regions in predicting seizure freedom. This result substantiates that epilepsy is a network disease, as rather than a single region, a more extensive network is engaged in seizure genesis and propagation. This trend is even more evident in later studies (after 2011) that analyzed the abnormality of networks [89] and nodes [91], local connectivity changes [83], tract pathology [93], multimodal functional anomalies and connectome gradients [86]. Roger et al. [94] also used graph theory approach to analyze the properties of neuropsychological networks of epilepsy patients who became seizure-free after surgery and those that did not.

4.6. Research-clinic bridge

This systematic literature review did not corroborate our hypothesis that ESOP studies would provide clinical tools. Only one study among 21 presented a clinical tool [94]. Based on their prognostic model, Roger et al. [94] proposed a clinical nomogram, a visual statistical

instrument that can predict an individualized patient-centered outcome of resective epilepsy surgery based on several predictors [118,119].

It could be criticized that we did not identify studies connecting ESOP research and clinic because of our inclusion/exclusion criteria (e.g., minimum of 12-month follow-up, only post-op). Nevertheless, even in the 211 ESOP studies that were full-text reviewed for eligibility (Fig. 1), there were only six studies that proposed a tool for clinics such as nomograms for clinical [118] and cognitive outcome prediction [31,119], online risk calculator for clinical outcome [105] and clinical and cognitive outcome [120][8], and seizure freedom score [122] for predicting epilepsy surgery outcome. These studies were recently published, so we can see a trend towards bridging research and clinic. The recommendation for future ESOP studies is to try to translate their prediction model into a tool applicable and accessible to clinical practice to help ameliorate epilepsy surgery decision-making, rehabilitation, and prehabilitation. In doing so, future ESOP models should rely on techniques, measures, and tests available in most surgical centers to increase their tools' applicability and the possibility for their validation. Alternatively, the code should be made openly available if a new composite or fusion metric is calculated.

ESOP tools should support surgical practice since human decision-making can be subject to cognitive biases [84,123]. However, ESOP models and resulting clinical tools should not replace the practitioner's decision [122]. Furthermore, the surgical decision in the case of epilepsy is complex and manifold, and the patient's perspective and satisfaction should not be neglected [119]. We reviewed the literature on two crucial epilepsy surgery outcomes – clinical and cognitive, that are the goal of the surgical team. However, from the patient's perspective, the goal is psychosocial well-being, and studies showed that seizure freedom is insufficient to achieve this (for a review, see [124]). Hence, the practitioner's decision should also incorporate the patient's perspective.

Our study has several limitations. We defined rigorous inclusion and exclusion criteria, resulting in fewer included studies. For instance, some ESOP studies included participants aged 16 or 17 [19,125,126]. Considering crucial factors such as maturation and brain plasticity (see section 4.3), we decided to keep the 18-65 age range [111,112]. However, we acknowledge that there is no explicitly defined age limit for pediatric epilepsy surgery [67]. Furthermore, we included studies whose post-operative follow-up period was at least 12 months without defining the duration of this period. Although previous studies show that the

[8] Although in their later paper Morita-Sherman et al. [121] nevertheless state that further improvements are needed for the online risk calculator for seizure-freedom prediction to be ready for general clinical usage.

follow-up duration is related to the clinical outcome of epilepsy surgery [37], as pointed out, information on follow-up duration was not always presented in the identified studies (see sections 3.3 and 4.2). Hence this criterion would lead to an even smaller sample of ESOP studies. We could have captured more fine methodological time trends in ESOP studies if our criteria had been more loosely defined. Nevertheless, the most prevalent reasons for study exclusion (Fig. 1) suggest that the characteristics of epilepsy patients in ESOP studies are very variable. Unless these differences are not accounted for in the ESOP model, they could bias the results or mask the effects of certain factors. In addition, in order to obtain the most coherent data, we focused our literature search only on TLE as the most prevalent focal epilepsy. Including other focal epilepsies might have brought a more thorough overview of the ESOP research. However, it would also lead to higher variability of the already variable results despite our strict criteria. Moreover, we acknowledge that our review and resulting conclusions could be culturally biased as we only included papers written in English. Finally, we provide a narrative synthesis and only descriptive statistics of the methodology of the selected ESOP studies. Despite our numerous inclusion and exclusion criteria, the variability of design and methodology of the identified studies did not allow us to perform a meta-analysis. This finding suggests, on one hand, the need for future meta-analysis which would systematically evaluate ESOP models (such as the meta-analysis of Alim-Marvasti et al. [40] on prognostic features for epilepsy surgery outcome), and on the other, more straightforward recommendations for future ESOP models and tools.

4.7. Conclusions

In this study, we explored the methodological characteristics of studies aiming to predict clinical or cognitive outcomes of epilepsy surgery. The narrative synthesis of the results showed us how the transformation of epilepsy conceptualization into a complex multifactorial disease is reflected in the methodological and analytical features of ESOP studies. It is becoming more prevalent that different neuroimaging techniques, clinical and neuropsychological testing, and analyses that support the modeling of complex relationships between their respective measurements are needed to evaluate all functional and structural aspects of epilepsy and predict resective surgery outcomes. This was evidenced in our results by an increase in different modalities, multimodal datasets, and ML applications in ESOP studies. Nevertheless, it should be pointed out that some methodological issues noted by McIntosh in 2001 [36] and early 2000s reviews [37,41], such as small sample sizes, lack of information (e.g., on the duration of the follow-up period), variable seizure outcome

definition across studies, heterogeneity of epilepsy patients, remained present in the recent ESOP studies. Therefore, Figure 3 presents the synthesis of the hypothesis and results of the present study, as well as the recommendation for future studies. Moreover, Box 2 includes suggestions for clinicians engaged in the pre-operative evaluation of epilepsy surgery candidates. We believe that future ESOP studies should aim to collect larger sample sizes with rich multimodal data to verify proposed ESOP models and translate them into clinical tools, thus bridging research and clinic.

Acknowledge

This work was supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) and the CDT *Brain and Cognition* Program (CerCoG@UGA) funded by IDEX UGA.

5. Bibliography

[1] Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, et al. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017;58:522–30. https://doi.org/10.1111/epi.13670.

[2] Téllez-Zenteno JF, Hernández-Ronquillo L. A Review of the Epidemiology of Temporal Lobe Epilepsy. Epilepsy Res Treat 2012;2012:1–5. https://doi.org/10.1155/2012/630853.

[3] Barr WB. Neuropsychological assessment of patients with epilepsy. In: Barr WB, Morrison C, editors. Handb. Neuropsychol. Epilepsy, Springer; 2015, p. 1–36.

[4] Kramer MA, Cash SS. Epilepsy as a Disorder of Cortical Network Organization. The Neuroscientist 2012;18:360–72. https://doi.org/10.1177/1073858411422754.

[5] Khateb M, Bosak N, Herskovitz M. The Effect of Anti-seizure Medications on the Propagation of Epileptic Activity: A Review. Front Neurol 2021;12:674182. https://doi.org/10.3389/fneur.2021.674182.

[6] Borger V, Hamed M, Taube J, Aydin G, Ilic I, Schneider M, et al. Resective temporal lobe surgery in refractory temporal lobe epilepsy: prognostic factors of postoperative seizure outcome. J Neurosurg 2021;135:760–9. https://doi.org/10.3171/2020.7.JNS20284.

[7] Fattorusso A, Matricardi S, Mencaroni E, Dell'Isola GB, Di Cara G, Striano P, et al. The Pharmacoresistant Epilepsy: An Overview on Existant and New Emerging Therapies. Front Neurol 2021;12:674483. https://doi.org/10.3389/fneur.2021.674483.

[8] Chen Z, Brodie MJ, Liew D, Kwan P. Treatment Outcomes in Patients With Newly Diagnosed Epilepsy Treated With Established and New Antiepileptic Drugs: A 30-Year Longitudinal Cohort Study. JAMA Neurol 2018;75:279. https://doi.org/10.1001/jamaneurol.2017.3949.

[9] Schoenberg MR, Werz MA, Drane DL. Epilepsy and seizures. Little Black Book Neuropsychol. Syndr.-Based Approach, Springer; 2011, p. 423–520.

[10] Téllez-Zenteno JF, Dhar R, Wiebe S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. Brain 2005;128:1188–98. https://doi.org/10.1093/brain/awh449.

[11] Mohan M, Keller S, Nicolson A, Biswas S, Smith D, Osman Farah J, et al. The long-term outcomes of epilepsy surgery. PLOS ONE 2018;13:e0196274. https://doi.org/10.1371/journal.pone.0196274.

[12] Baxendale S, Thompson P. Red flags in epilepsy surgery: Identifying the patients who pay a high cognitive price for an unsuccessful surgical outcome. Epilepsy Behav 2018;78:269–72. https://doi.org/10.1016/j.yebeh.2017.08.003.

[13] Helmstaedter C. Cognitive outcomes of different surgical approaches in temporal lobe epilepsy. Epileptic Disord 2013;15:221–39. https://doi.org/10.1684/epd.2013.0587.

[14] 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: Cognitive Change after Epilepsy Surgery. Epilepsia 2011;52:857–69. https://doi.org/10.1111/j.1528-1167.2011.03022.x.

[15] Massot-Tarrús A, White K, Mirsattari SM. Comparing the Wada Test and Functional MRI for the Presurgical Evaluation of Memory in Temporal Lobe Epilepsy. Curr Neurol Neurosci Rep 2019;19:1–14. https://doi.org/10.1007/s11910-019-0945-8.

[16] 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;21:14. https://doi.org/10.1684/epd.2019.1065.

[17] Aull-Watschinger S, Pataraia E, Czech T, Baumgartner C. Outcome predictors for surgical treatment of temporal lobe epilepsy with hippocampal sclerosis. Epilepsia 2008;49:1308–16. https://doi.org/10.1111/j.1528-1167.2008.01732.x.

[18] Thom M, Mathern GW, Cross JH, Bertram EH. Mesial temporal lobe epilepsy: How do we improve surgical outcome? Ann Neurol 2010;68:424–34. https://doi.org/10.1002/ana.22142.

[19] Armañanzas R, Alonso-Nanclares L, DeFelipe-Oroquieta J, Kastanauskaite A, de Sola RG, DeFelipe J, et al. Machine Learning Approach for the Outcome Prediction of Temporal Lobe Epilepsy Surgery. PLoS ONE 2013;8:e62819. https://doi.org/10.1371/journal.pone.0062819.

[20] Memarian N, Kim S, Dewar S, Engel J, Staba RJ. Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. Comput Biol Med 2015;64:67–78. https://doi.org/10.1016/j.compbiomed.2015.06.008.

[21] Wiebe S. A Randomized, Controlled Trial of Surgery for Temporal-Lobe Epilepsy. N Engl J Med 2001;345:311–8. https://doi.org/10.1056/NEJM200108023450501.

[22] Feis D-L, Schoene-Bake J-C, Elger C, Wagner J, Tittgemeyer M, Weber B. Prediction of post-surgical seizure outcome in left mesial temporal lobe epilepsy. NeuroImage Clin 2013;2:903–11. https://doi.org/10.1016/j.nicl.2013.06.010.

[23] Munsell BC, Wee C-Y, Keller SS, Weber B, Elger C, da Silva LAT, et al. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. NeuroImage 2015;118:219–30. https://doi.org/10.1016/j.neuroimage.2015.06.008.

[24] You X, Zachery AN, Fanto EJ, Norato G, Germeyan SC, Emery EJ, et al. fMRI prediction of naming change after adult temporal lobe epilepsy surgery: Activation matters. Epilepsia 2019;60:527–38. https://doi.org/10.1111/epi.14656.

[25] Antony AR, Alexopoulos AV, González-Martínez JA, Mosher JC, Jehi L, Burgess RC, et al. Functional Connectivity Estimated from Intracranial EEG Predicts Surgical Outcome in Intractable Temporal Lobe Epilepsy. PLoS ONE 2013;8:e77916. https://doi.org/10.1371/journal.pone.0077916.

[26] Jeong S-W, Lee SK, Hong K-S, Kim K-K, Chung C-K, Kim H. Prognostic Factors for the Surgery for Mesial Temporal Lobe Epilepsy: Longitudinal Analysis. Epilepsia 2005;46:1273–9. https://doi.org/10.1111/j.1528-1167.2005.33504.x.

[27] DeSalvo MN, Tanaka N, Douw L, Cole AJ, Stufflebeam SM. Contralateral Preoperative Resting-State Functional MRI Network Integration Is Associated with Surgical Outcome in Temporal Lobe Epilepsy. Radiology 2020;294:622–7. https://doi.org/10.1148/radiol.2020191008.

[28] Elsharkawy AE, Alabbasi AH, Pannek H, Oppel F, Schulz R, Hoppe M, et al. Long-term outcome after temporal lobe epilepsy surgery in 434 consecutive adult patients: Clinical article. J Neurosurg 2009;110:1135–46. https://doi.org/10.3171/2008.6.JNS17613.

[29] Armon C, Radtke RA, Friedman AH, Dawson DV. Predictors of Outcome of Epilepsy Surgery: Multivariate Analysis with Validation. Epilepsia 1996;37:814–21. https://doi.org/10.1111/j.1528-1157.1996.tb00033.x.

[30] Mathon B, Bielle F, Samson S, Plaisant O, Dupont S, Bertrand A, et al. Predictive factors of long-term outcomes of surgery for mesial temporal lobe epilepsy associated with hippocampal sclerosis. Epilepsia 2017;58:1473–85. https://doi.org/10.1111/epi.13831.

[31] 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.

[32] Ljunggren S, Andersson-Roswall L, Imberg H, Samuelsson H, Malmgren K. Predicting verbal memory decline following temporal lobe resection for epilepsy. Acta Neurol Scand 2019;140:312–9. https://doi.org/10.1111/ane.13146.

[33] 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 2005;7:273–8. https://doi.org/10.1016/j.yebeh.2005.05.016.

[34] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021:n71. https://doi.org/10.1136/bmj.n71.

[35] Wiebe S. Effectiveness and Safety of Epilepsy Surgery: What is the Evidence? CNS Spectr 2004;9:11.

[36] McIntosh AM, Wilson SJ, Berkovic SF. Seizure Outcome after Temporal Lobectomy: Current Research Practice and Findings. Epilepsia 2001;42:1288–307. https://doi.org/10.1046/j.1528-1157.2001.02001.x.

[37] 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.

[38] Bjellvi J, Olsson I, Malmgren K, Wilbe Ramsay K. Epilepsy duration and seizure outcome in epilepsy surgery: A systematic review and meta-analysis. Neurology 2019;93:e159–66. https://doi.org/10.1212/WNL.0000000000007753.

[39] Parra-Díaz P, García-Casares N. Memory assessment in patients with temporal lobe epilepsy to predict memory impairment after surgery: a systematic review. Neurol Engl Ed 2019:S2173580819300173. https://doi.org/10.1016/j.nrleng.2017.02.011.

[40] Alim-Marvasti A, Vakharia VN, Duncan JS. Multimodal prognostic features of seizure freedom in epilepsy surgery. J Neurol Neurosurg Psychiatry 2022;93:499–508. https://doi.org/10.1136/jnnp-2021-327119.

[41] Téllez-Zenteno JF, Ronquillo LH, Moien-Afshari F, Wiebe S. Surgical outcomes in lesional and non-lesional epilepsy: A systematic review and meta-analysis. Epilepsy Res 2010;89:310–8. https://doi.org/10.1016/j.eplepsyres.2010.02.007.

[42] Widjaja E, Jain P, Demoe L, Guttmann A, Tomlinson G, Sander B. Seizure outcome of pediatric epilepsy surgery: Systematic review and meta-analyses. Neurology 2020;94:311–21. https://doi.org/10.1212/WNL.0000000000008966.

[43] Ives-Deliperi VL, Butler JT. Naming outcomes of anterior temporal lobectomy in epilepsy patients: A systematic review of the literature. Epilepsy Behav 2012;24:194–8. https://doi.org/10.1016/j.yebeh.2012.04.115.

[44] Krucoff MO, Chan AY, Harward SC, Rahimpour S, Rolston JD, Muh C, et al. Rates and predictors of success and failure in repeat epilepsy surgery: A meta-analysis and systematic review. Epilepsia 2017;58:2133–42. https://doi.org/10.1111/epi.13920.

[45] Englot DJ, Chang EF. Rates and predictors of seizure freedom in resective epilepsy surgery: an update. Neurosurg Rev 2014;37:389–405. https://doi.org/10.1007/s10143-014-0527-9.

[46] Beghi E, Tonini C. Surgery for epilepsy: Assessing evidence from observational studies. Epilepsy Res 2006;70:97–102. https://doi.org/10.1016/j.eplepsyres.2006.03.013.

[47] Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman MLD, et al. Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. World Neurosurg 2018;109:476–86. https://doi.org/10.1016/j.wneu.2017.09.149.

[48] Yuan J, Ran X, Liu K, Yao C, Yao Y, Wu H, et al. Machine Learning Applications on Neuroimaging for Diagnosis and Prognosis of Epilepsy: A Review. J Neurosci Methods 2022;368:109441. https://doi.org/10.1016/j.jneumeth.2021.109441.

[49] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004.

[50] Calhoun VD, Sui J. Multimodal Fusion of Brain Imaging Data: A Key to Finding the Missing Link(s) in Complex Mental Illness. Biol Psychiatry Cogn Neurosci Neuroimaging 2016;1:230–44. https://doi.org/10.1016/j.bpsc.2015.12.005.

[51] Duncan JS, Trimmel K. Advanced neuroimaging techniques in epilepsy. Curr Opin Neurol 2022;35:189–95. https://doi.org/10.1097/WCO.0000000000001007.

[52] Zhang Y-D, Dong Z, Wang S-H, Yu X, Yao X, Zhou Q, et al. Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. Inf Fusion 2020;64:149–87. https://doi.org/10.1016/j.inffus.2020.07.006.

[53] Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav Res Methods 2009;41:1149–60. https://doi.org/10.3758/BRM.41.4.1149.

[54] Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. Behav Res Methods Instrum Comput 1996;28:1–11. https://doi.org/10.3758/BF03203630.

[55] Baxendale S, Thompson P. Defining meaningful postoperative change in epilepsy surgery patients: Measuring the unmeasurable? Epilepsy Behav 2005;6:207–11. https://doi.org/10.1016/j.yebeh.2004.12.009.

[56] Iverson GL. Reliable Change Index. In: Kreutzer J, DeLuca J, Caplan B, editors. Encycl. Clin. Neuropsychol., Cham: Springer International Publishing; 2019, p. 1–4. https://doi.org/10.1007/978-3-319-56782-2_1242-3.

[57] Engel JJ. Epilepsy Board Quick Review: Selected Tables and Figures from Seizures and Epilepsy. USA: Oxford University Press; 2014.

[58] Larivière S, Bernasconi A, Bernasconi N, Bernhardt BC. Connectome biomarkers of drug-resistant epilepsy. Epilepsia 2021;62:6–24. https://doi.org/10.1111/epi.16753.

[59] Roberts JI, Hrazdil C, Wiebe S, Sauro K, Vautour M, Wiebe N, et al. Neurologists' knowledge of and attitudes toward epilepsy surgery: A national survey. Neurology 2015;84:159–66. https://doi.org/10.1212/WNL.0000000000001127.

[60] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160. https://doi.org/10.1136/bmj.n160.

[61] Noe K, Sulc V, Wong-Kisiel L, Wirrell E, Gompel JJV, Wetjen N, et al. Long-term Outcomes After Nonlesional Extratemporal Lobe Epilepsy Surgery. JAMA Neurol 2013;70:1003–8. https://doi.org/10.1001/jamaneurol.2013.209.

[62] Kumar A, Valentín A, Humayon D, Longbottom AL, Jimenez-Jimenez D, Mullatti N, et al. Preoperative estimation of seizure control after resective surgery for the treatment of epilepsy. Seizure 2013;22:818–26. https://doi.org/10.1016/j.seizure.2013.06.010.

[63] Thom M, Bertram EH. Temporal lobe epilepsy. Handb. Clin. Neurol., vol. 107, Elsevier; 2012, p. 225–40. https://doi.org/10.1016/B978-0-444-52898-8.00014-8.

[64] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017;58:512–21. https://doi.org/10.1111/epi.13709.

[65] Thomas B, Aupy J, Penchet G, De Montaudouin M, Bartolomei F, Biraben A, et al. Predictive factors of postoperative outcome in the elderly after resective epilepsy surgery. Rev Neurol (Paris) 2022;178:609–15. https://doi.org/10.1016/j.neurol.2021.08.011.

[66] Grivas A, Schramm J, Kral T, von Lehe M, Helmstaedter C, Elger CE, et al. Surgical Treatment for Refractory Temporal Lobe Epilepsy in the Elderly: Seizure Outcome and Neuropsychological Sequels Compared with a Younger Cohort. Epilepsia 2006;47:1364–72. https://doi.org/10.1111/j.1528-1167.2006.00608.x.

[67] Cross JH, Jayakar P, Nordli D, Delalande O, Duchowny M, Wieser HG, et al. Proposed Criteria for Referral and Evaluation of Children for Epilepsy Surgery: Recommendations of the Subcommission for Pediatric Epilepsy Surgery. Epilepsia 2006;47:952–9. https://doi.org/10.1111/j.1528-1167.2006.00569.x.

[68] Ryvlin P, Cross JH, Rheims S. Epilepsy surgery in children and adults. Lancet Neurol 2014;13:1114–26. https://doi.org/10.1016/S1474-4422(14)70156-5.

[69] Smith ML. Neuropsychology in epilepsy: Children are not small adults. Epilepsia 2010;51:68–9. https://doi.org/10.1111/j.1528-1167.2009.02451.x.

[70] Giovagnoli A, Parente A, Didato G, Manfredi V, Deleo F, Tringali G, et al. The course of language functions after temporal lobe epilepsy surgery: a prospective study. Eur J Neurol 2016;23:1713–21. https://doi.org/10.1111/ene.13113.

[71] Alpherts W, Vermeulen J, van Rijen P, da Silva F, van Veelen C, Dutch Collaborative Epilepsy Surg. Verbal memory decline after temporal epilepsy surgery? A 6-year multiple assessments follow-up study. NEUROLOGY 2006;67:626–31. https://doi.org/10.1212/01.wnl.0000230139.45304.eb.

[72] Andersson-Roswall L, Malmgren K, Engman E, Samuelsson H. Verbal memory decline is less frequent at 10years than at 2years after temporal lobe surgery for epilepsy. Epilepsy Behav 2012;24:462–7. https://doi.org/10.1016/j.yebeh.2012.05.015.

[73] Helmstaedter C, Elger CE, Vogt VL. Cognitive outcomes more than 5 years after temporal lobe epilepsy surgery: Remarkable functional recovery when seizures are controlled. Seizure 2018;62:116–23. https://doi.org/10.1016/j.seizure.2018.09.023.

[74] Baxendale S. The impact of epilepsy surgery on cognition and behavior. Epilepsy Behav 2008;12:592–9. https://doi.org/10.1016/j.yebeh.2007.12.015.

[75] Wieser HG, Blume WT, Fish D, Goldensohn E, Hufnagel A, King D, et al. Proposal for a New Classification of Outcome with Respect to Epileptic Seizures Following Epilepsy Surgery. Epilepsia 2001;42:282–286. https://doi.org/10.1046/j.1528-1157.2001.4220282.x.

[76] 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.

[77] Taube J, Witt J-A, Grote A, Delev D, Enkirch J, Hattingen E, et al. Preoperative and postoperative memory in epilepsy patients with 'gliosis only' versus hippocampal sclerosis: a matched case–control study. J Neurol Neurosurg Psychiatry 2022:jnnp-2022-329224. https://doi.org/10.1136/jnnp-2022-329224.

[78] Ouzzani M. Rayyan - A Web and Mobile App for Systematic Reviews. Syst Rev 2016;5:1–10. https://doi.org/10.1186/s13643-016-0384-4.

[79] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019;170:51. https://doi.org/10.7326/M18-1376.

[80] Navarro CLA, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ 2021;375:2281. http://dx.doi.org/10.1136/bmj.n2281.

[81] Jong Y de, Ramspek CL, Zoccali C, Jager KJ, Dekker FW, Diepen M van. Appraising prediction research: a guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). Nephrology 2021;26:939–47. https://doi.org/10.1111/nep.13913.

[82] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med 2019;170:W1. https://doi.org/10.7326/M18-1377.

[83] Chen X, Wang Y, Kopetzky SJ, Butz-Ostendorf M, Kaiser M. Connectivity within regions characterizes epilepsy duration and treatment outcome. Hum Brain Mapp 2021;42:3777–91. https://doi.org/10.1002/hbm.25464.

[84] Baxendale S, Thompson P, McEvoy A, Duncan J. Epilepsy surgery: How accurate are multidisciplinary teams in predicting outcome? Seizure 2012;21:546–9. https://doi.org/10.1016/j.seizure.2012.05.008.

[85] Ioriatti ES, Cirino MLA, Lizarte Neto FS, Velasco TR, Sakamoto AC, Freitas-Lima P, et al. Expression of circulating microRNAs as predictors of diagnosis and surgical outcome in patients with mesial temporal lobe epilepsy with hippocampal sclerosis. Epilepsy Res 2020;166:106373. https://doi.org/10.1016/j.eplepsyres.2020.106373.

[86] Larivière S, Weng Y, Vos de Wael R, Royer J, Frauscher B, Wang Z, et al. Functional connectome contractions in temporal lobe epilepsy: Microstructural underpinnings and predictors of surgical outcome. Epilepsia 2020;61:1221–33. https://doi.org/10.1111/epi.16540.

[87] Rathore C, Kesavadas C, Sarma SP, Radhakrishnan K. Usefulness of Wada test in predicting seizure outcome following anterior temporal lobectomy. Epilepsy Res 2013;107:279–85. https://doi.org/10.1016/j.eplepsyres.2013.09.003.

[88] Wong CH, Bleasel A, Wen L, Eberl S, Byth K, Fulham M, et al. The topography and significance of extratemporal hypometabolism in refractory mesial temporal lobe epilepsy examined by FDG-PET: Topography and Significance of Extratemporal Hypometabolism. Epilepsia 2010;51:1365–73. https://doi.org/10.1111/j.1528-1167.2010.02552.x.

[89] Bonilha L, Jensen JH, Baker N, Breedlove J, Nesland T, Lin JJ, et al. The brain connectome as a personalized biomarker of seizure outcomes after temporal lobectomy. Neurology 2015;84:1846–53. https://doi.org/10.1212/WNL.0000000000001548.

[90] Eberhardt KE, Stefan H, Buchfelder M, Pauli E, Hopp P, Huk W, et al. The Significance of Bilateral CSI Changes for the Postoperative Outcome in Temporal Lobe Epilepsy: J Comput Assist Tomogr 2000;24:919–26. https://doi.org/10.1097/00004728-200011000-00019.

[91] Sinha N, Wang Y, Moreira da Silva N, Miserocchi A, McEvoy AW, de Tisi J, et al. Structural brain network abnormalities and the probability of seizure recurrence after epilepsy surgery. Neurology 2021;96:e758–71. https://doi.org/10.1212/WNL.0000000000011315.

[92] Burneo JG, Black L, Martin R, Devinsky O, Pacia S, Faught E, et al. Race/Ethnicity, Sex, and Socioeconomic Status as Predictors of Outcome After Surgery for Temporal Lobe Epilepsy. Arch Neurol 2006;63:1106. https://doi.org/10.1001/archneur.63.8.1106.

[93] Keller SS, Glenn GR, Weber B, Kreilkamp BAK, Jensen JH, Helpern JA, et al. Preoperative automated fibre quantification predicts postoperative seizure outcome in temporal lobe epilepsy. Brain 2017;140:68–82. https://doi.org/10.1093/brain/aww280.

[94] Roger E, Torlay L, Banjac S, Mosca C, Minotti L, Kahane P, et al. Prediction of the clinical and naming status after anterior temporal lobe resection in patients with epilepsy. Epilepsy Behav 2021;124:108357. https://doi.org/10.1016/j.yebeh.2021.108357.

[95] Gaça LB, Garcia MTFC, Sandim GB, Assumption Leme IB, Noffs MHS, Carrete H, et al. Morphometric MRI features and surgical outcome in patients with epilepsy related to hippocampal sclerosis and low intellectual quotient. Epilepsy Behav 2018;82:144–9. https://doi.org/10.1016/j.yebeh.2018.03.011.

[96] Krendl R, Lurger S, Baumgartner C. Absolute spike frequency predicts surgical outcome in TLE with unilateral hippocampal atrophy. Neurology 2008;71:413–8. https://doi.org/10.1212/01.wnl.0000310775.87331.90.

[97] Paff M, Boutet A, Elias GJB, Germann J, Mansouri A, Loh A, et al. Lateralizing magnetic resonance imaging findings in mesial temporal sclerosis and correlation with seizure and neurocognitive outcome after temporal lobectomy. Epilepsy Res 2021;171:106562. https://doi.org/10.1016/j.eplepsyres.2021.106562.

[98] Dupont S, Semah F, Clémenceau S, Adam C, Baulac M, Samson Y. Accurate Prediction of Postoperative Outcome in Mesial Temporal Lobe Epilepsy: A Study Using Positron Emission Tomography With ^{18} Fluorodeoxyglucose. Arch Neurol 2000;57:1331–6. https://doi.org/10.1001/archneur.57.9.1331.

[99] Voets NL, Bernhardt BC, Kim H, Yoon U, Bernasconi N. Increased temporolimbic cortical folding complexity in temporal lobe epilepsy. Neurology 2011;76:138–44. https://doi.org/10.1212/WNL.0b013e318205d521.

[100] Asadi-Pooya AA, Nei M, Sharan A, Sperling MR. Historical Risk Factors Associated with Seizure Outcome After Surgery for Drug-Resistant Mesial Temporal Lobe Epilepsy. World Neurosurg 2016;89:78–83. https://doi.org/10.1016/j.wneu.2016.02.023.

[101] Audrain S, Barnett AJ, McAndrews MP. Language network measures at rest indicate individual differences in naming decline after anterior temporal lobe resection. Hum Brain Mapp 2018;39:4404–19. https://doi.org/10.1002/hbm.24281.

[102] Uijl SG, Leijten FSS, Arends JBAM, Parra J, van Huffelen AC, van Rijen PC, et al. The intracarotid amobarbital or Wada test: unilateral or bilateral? Acta Neurol Scand 2009;119:199–206. https://doi.org/10.1111/j.1600-0404.2008.01079.x.

[103] Lahat D, Adali T, Jutten C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. Proc IEEE 2015;103:1449–77. https://doi.org/10.1109/JPROC.2015.2460697.

[104] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med 2015;162:55–63. https://doi.org/10.7326/M14-0697.

[105] Benjumeda M, Chandramohan D, Chang EF, Hall JA, Bielza C, Larrañaga P, et al. Patient specific prediction of temporal lobe epilepsy surgical outcomes. Epilepsia 2021;62:2113–22. https://doi.org/10.1111/epi.17002.

[106] Osipowicz K, Sperling MR, Sharan AD, Tracy JI. Functional MRI, resting state fMRI, and DTI for predicting verbal fluency outcome following resective surgery for temporal lobe epilepsy. J Neurosurg 2016;124:929–37. https://doi.org/10.3171/2014.9.JNS131422.

[107] 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.

[108] van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14:137. https://doi.org/10.1186/1471-2288-14-137.

[109] Campbell MK, Thomson S, Ramsay CR, MacLennan GS, Grimshaw JM. Sample size calculator for cluster randomized trials. Comput Biol Med 2004;34:113–25. https://doi.org/10.1016/S0010-4825(03)00039-8.

[110] Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441. https://doi.org/10.1136/bmj.m441.

[111] Harvey AS, Cross JH, Shinnar S, Mathern GW, the Pediatric Epilepsy Surgery Survey Taskforce. Defining the spectrum of international practice in pediatric epilepsy surgery patients. Epilepsia 2008;49:146–55. https://doi.org/10.1111/j.1528-1167.2007.01421.x.

[112] Cloppenborg T, May TW, Blümcke I, Fauser S, Grewe P, Hopf JL, et al. Differences in pediatric and adult epilepsy surgery: A comparison at one center from 1990 to 2014. Epilepsia 2019;60:233–45. https://doi.org/10.1111/epi.14627.

[113] Barba C, Cossu M, Guerrini R, Gennaro GD, Villani F, Palma LD, et al. Temporal lobe epilepsy surgery in children and adults: A multicenter study. Epilepsia n.d.;62:128–42. https://doi.org/10.1111/epi.16772.

[114] Tognin S, van Hell HH, Merritt K, Winter-van Rossum I, Bossong MG, Kempton MJ, et al. Towards Precision Medicine in Psychosis: Benefits and Challenges of Multimodal Multicenter Studies—PSYSCAN: Translating Neuroimaging Findings From Research into Clinical Practice. Schizophr Bull 2020;46:432–41. https://doi.org/10.1093/schbul/sbz067.

[115] Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods 2018;15:233–4. https://doi.org/10.1038/nmeth.4642.

[116] Brereton RG. Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. TrAC Trends Anal Chem 2006;25:1103–11. https://doi.org/10.1016/j.trac.2006.10.005.

[117] Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol 2016;69:245–7. https://doi.org/10.1016/j.jclinepi.2015.04.005.

[118] Jehi L, Yardi R, Chagin K, Tassi L, Russo GL, Worrell G, et al. Development and validation of nomograms to provide individualised predictions of seizure outcomes after epilepsy surgery: a retrospective analysis. Lancet Neurol 2015;14:283–90. https://doi.org/10.1016/S1474-4422(14)70325-4.

[119] Busch RM, Hogue O, Kattan MW, Hamberger M, Drane DL, Hermann BP, 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.

[120] Morita-Sherman M, Louis S, Vegh D, Busch RM, Ferguson L, Bingaman J, et al. Outcomes of resections that spare vs remove an MRI-normal hippocampus. Epilepsia 2020;61:2545–57. https://doi.org/10.1111/epi.16694.

[121] Morita-Sherman M, Li M, Joseph B, Yasuda C, Vegh D, De Campos B, et al. Incorporation of quantitative MRI in a mode to predict temporal lobe epilepsy surgery outcome. Brain Commun 2021;3:fcab164. https://doi.org/10.1093/braincomms/fcab164.

[122] Garcia Gracia C, Yardi R, Kattan MW, Nair D, Gupta A, Najm I, et al. Seizure freedom score: A new simple method to predict success of epilepsy surgery. Epilepsia 2015;56:359–65. https://doi.org/10.1111/epi.12892.

[123] Jehi L. Algorithms in clinical epilepsy practice: Can they really help us predict epilepsy outcomes? Epilepsia 2021;62:S71–7. https://doi.org/10.1111/epi.16649.

[124] Wilson SJ, Saling MM, Lawrence J, Bladin PF. Outcome of temporal lobectomy: expectations and the prediction of perceived success. Epilepsy Res 1999;36:1–14. https://doi.org/10.1016/S0920-1211(99)00016-9.

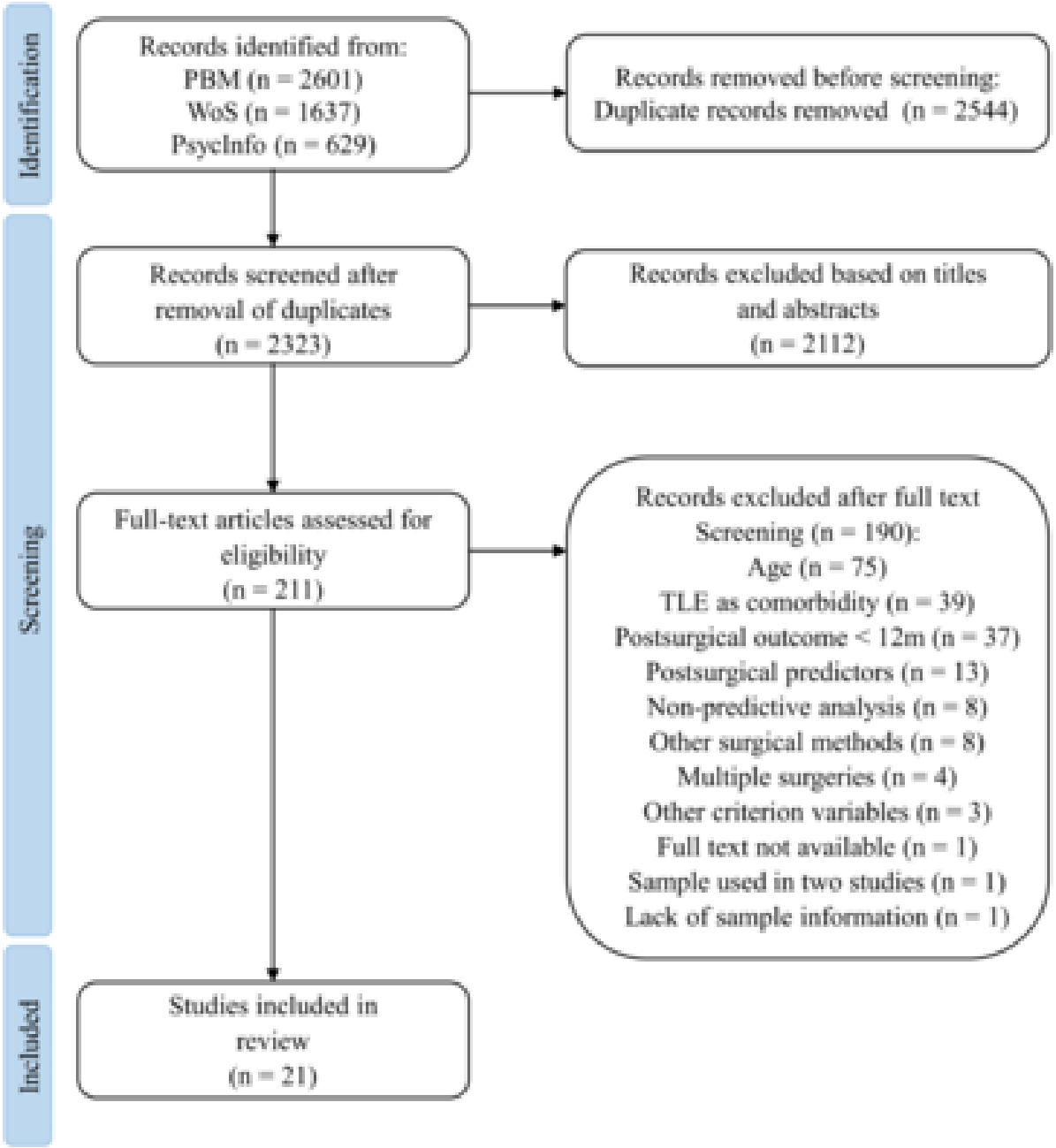
[125] Busch RM, Dulay MF, Kim KH, Chapin JS, Jehi L, Kalman CC, et al. Pre-Surgical Mood Predicts Memory Decline after Anterior Temporal Lobe Resection for Epilepsy. Arch Clin Neuropsychol 2011;26:739–45. https://doi.org/10.1093/arclin/acr067.

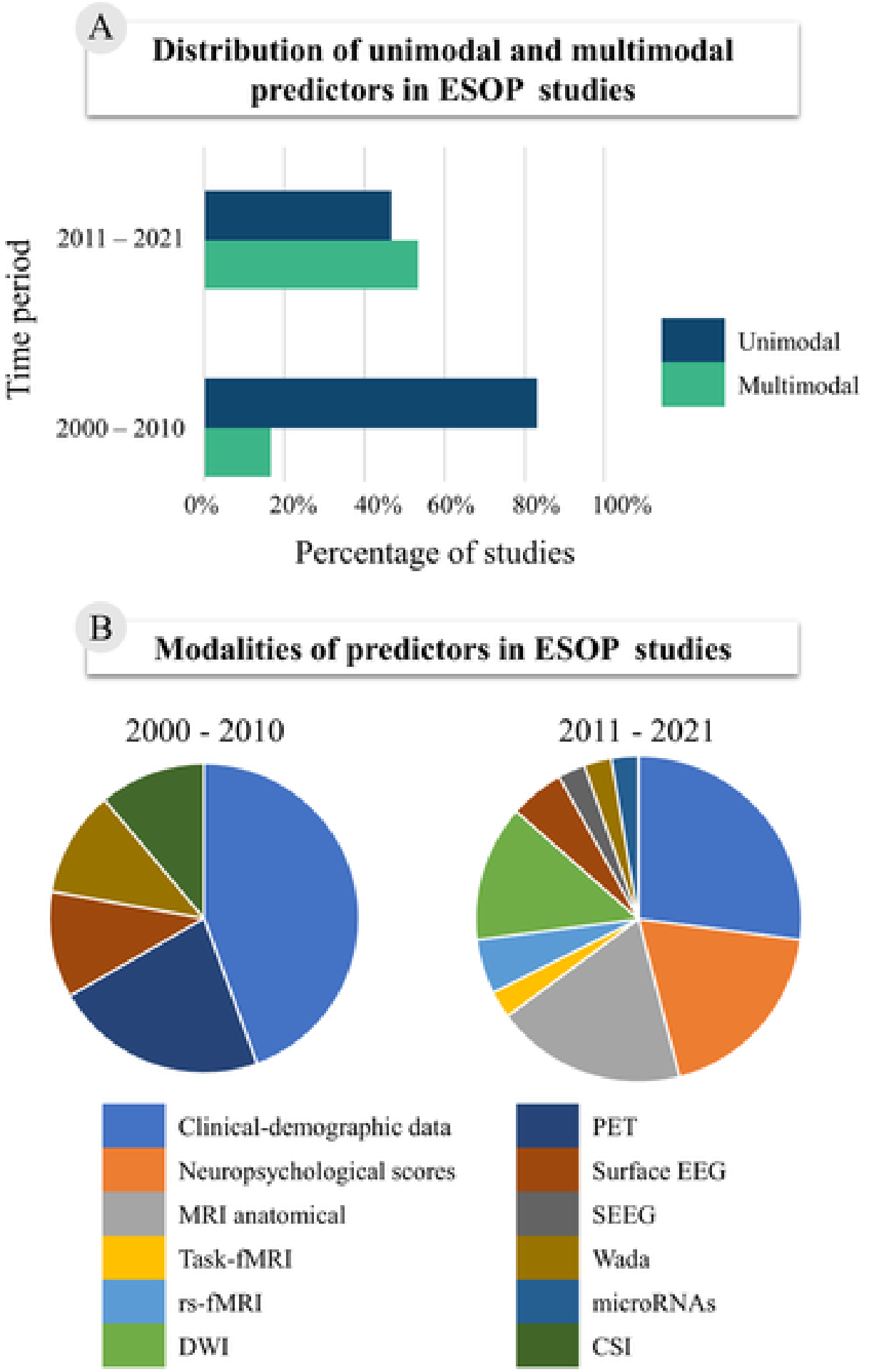
[126] Bernhardt BC, Bernasconi N, Hong S-J, Dery S, Bernasconi A. Subregional Mesiotemporal Network Topology Is Altered in Temporal Lobe Epilepsy. Cereb Cortex 2016;26:3237–48. https://doi.org/10.1093/cercor/bhv166.

Figure 1. PRISMA flow diagram of the systematic literature search and study selection, adapted from [34]. The systematic literature review of the three databases yielded 4867 articles (the exact strings are presented in Table S1). Following duplicate removal, two researchers screened the title and abstract of all papers identified, excluding papers that did not meet the inclusion criteria. The full texts of the remaining papers were screened, after which the final selection of 22 ESOP studies was made. The screening was performed using Rayyan application [78]. **Abbreviations**: PMB = PubMed; WoS = Web of Science.

Figure 2. Distribution of the identified ESOP studies with respect to the modality of their datasets and period of publication. Panel A shows the percentage of unimodal and multimodal ESOP studies from 2000 – 2010 and 2011 – 2021. Panel B shows two pie charts illustrating the proportion of studies using a specific modality among all ESOP studies in the two periods.

Figure 3. Domains of ESOP studies that were of focus in this systematic literature review. The hypothesis and principal results are presented for each domain and recommendations for future ESOP studies.


Identification
Screening
Included
Records identified from:
PBM (n = 2601)
WoS (n = 1637)
PsycInfo (n = 629)
Records removed before screening:
Duplicate records removed (n = 2544)
Records screened after removal of duplicates (n = 2323)
Records excluded based on titles and abstracts (n = 2112)
Full-text articles assessed for eligibility (n = 211)
Records excluded after full text Screening (n = 190):
Age (n = 75)
TLE as comorbidity (n = 39)
Postsurgical outcome < 12m (n = 37)
Postsurgical predictors (n = 13)
Non-predictive analysis (n = 8)
Other surgical methods (n = 8)
Multiple surgeries (n = 4)
Other criterion variables (n = 3)
Full text not available (n = 1)
Sample used in two studies (n = 1)
Lack of sample information (n = 1)
Studies included in review (n = 21)


A
Distribution of unimodal and multimodal predictors in ESOP studies
Time period
2011 – 2021
2000 – 2010
Unimodal
Multimodal
0%
20%
40%
60%
80%
100%
Percentage of studies
B
Modalities of predictors in ESOP studies
2000 - 2010
2011 - 2021
Clinical-demographic data
Neuropsychological scores
MRI anatomical
Task-fMRI
rs-fMRI
DWI
PET
Surface EEG
SEEG
Wada
microRNAs
CSI

Domain	Hypothesis	Principal findings	Recommendation
Modality of predictors	Increase of multimodal predictors in ESOP studies with time	After 2011 increase in the number of multimodal studies	Including interaction of modalities in ESOP models by including symmetric or asymmetric data fusion
Methods - outcome	Clinical ESOP studies should be using standardized taxonomies Cognitive ESOP studies should be using RCI or SRB	Despite using Engel's or ILAE taxonomy, outcome categories differed Variability in post-operative score change calculation	Using the proposed taxonomies and not author-defined classifications Calculating RCI or SRB as meaningful post-operative cognitive change to avoid test-retest artifacts
Methods – sample	The sample size of ESOP studies would grow with time	Sample size decreased after 2011 Majority of studies did not have sufficient sample size	Using and reporting clear inclusion and exclusion criteria Collecting a homogenous sample or including in ESOP model the characteristics by which patients differ Using available tools to calculate necessary sample size Multicenter studies
Analytical approach	ML would be dominated in later and multimodal studies	Regression and ML were the two primary analyses in the ESOP studies Multimodal clinical ESOP studies employed more the ML than the unimodal studies From 2011 the notion of networks more present in the methodology of ESOP studies	Regression for models based on theory and expert knowledge focused on inference ML for data-driven studies with rich multimodal datasets Future studies should aim to validate the existing ESOP models
Clinical tools	The number of tool-providing ESOP studies should be high	Only one study provided a tool for clinical practice	Future ESOP studies to translate their prediction model into a clinical tool Models and tools should rely on techniques, measures, and tests available in most surgical centers

Table 1. The identified ESOP studies and their characteristics.

Outcome	Period	Studies	Center	Inclusion period	Modality of predictors	Clinical outcome	Cognitive outcome	Sample size	% of LTLE	Analysis	Predictor selection	Model evaluation or validation	Clinical tool	Study Conclusions
clinical outcomes	2000 – 2010	Burneo et al. [92]	M	1994 - 2003	C-D	Engels: SF (E1) vs. NSF (E2 - E4)		252	44	R	No	No	No	Sex plays role in outcome of TLE surgery, but SES does not.
		Dupont et al. [98]	U	NR	PET	Engels: good (E1a) vs. worst (E2 - E3) outcome		30	NR	DA	Yes	Yes	No	Pattern of hypometabolism in a network of connected regions predicts SF better than single region hypometabolism
		Eberhardt et al. [90]	NR	NR	CSI	Engels: SF (E1) vs. NSF (E2 - E4)		26	53.8	DA	No	No	No	Bilateral metabolic CSI changes have predictive value for postop outcome in TLE.
		Krendl et al. [96]	NR	1997-2000	Surf. EEG	ILAE: SF (1a) vs. NSF (all other classes)		55	NR	R	Yes	No	No	Preop spike freq. strong predictor for surgical outcome.
		Wong et al. [88]	M	1994-2004	C-D, PET	Engel: good (E1) vs. bad outcome (E2 - E4)		64	NR	R	Yes	No	No	RH in MTLE assoc. with poorer surgical outcome, esp. if in contralateral hemisphere.
	2011 – 2021	Asadi-Pooya et al. [100]	U	1986-2014	C-D, NPS	Engels: SF (E1) vs. NSF (E2 - E4)		275	NR	R	Yes	No	No	Gender, race, FH, FS, history SE, duration of disease, IQ and SF not predictors of outcome.
		Baxendale et al. [84]	U	2000-2007	C-D, NPS, MRI,	ILAE: SF (ILAE 1-2) vs. NSF (all other classes)		94	53.2	R	No	No	No	Probabilities based on LR improve clinical estimates

			srf. EEG, SEEG								acc. of postop outcome in pts with very good/very poor chance of SF.
Bonilha et al. [89]	U	NR	C-D, MRI, DWI, srf. EEG	Engels: SF (E1a, E1b) vs. NSF (E2-E4)	35	NR	ML	Yes	Yes	No	Individual variations in C-T + preop clinical data biomarkers for estimating postop outcomes for TLE pts.
Chen et al. [83]	NR	NR	DWI	ILAE: Good (1-2) vs. bad outcome (3-5)	33	57.6	ML	No	Yes	No	Connectivity within regions more efficient biomarker for postop SF prediction.
Feis et al. [22]	U	2007-2011	MRI	ILAE: favorable (1-2) vs. unfavorable outcome (3-6)	49	100	ML	Yes	Yes	No	Results support using single T1-weighted MRI in routine preop of TLE.
Gaça et al. [95]	U	2004-2013	NPS	Engels: SF (E1a) vs. NSF (E1b- E4)	106	53.8	R	No	No	No	IQ not predictor of surgical outcome in low IQ MTLE-HS pts.
Ioriatti et al. [85]	U	NR	microRNAs	Engels: Favorable (E1) vs. unfavorable outcome (E3-E4)	28	NR	ROC	No	No	No	miR-654-3p only microRNA with statistical power to differentiate Engel I from Engel III-IV pts.
Keller et al. [93]	U	2006-2011	DWI	ILAE: SF (1) vs. NSF (2-6)	43	62.8	ROC	Yes	No	No	Pathology of ipsilateral dorsal fornix and contralateral PHG WM classified SF and NONSF pts.
Larivière et al. [86]	U	2008-2017	C-D, MRI, rs-fMRI, DWI	Engels: SF (E1) vs. NSF free (E2- E4)	30	36.7	ML	No	Yes	No	Multimodal functional anomalies predicted SF better than clinical and structural imaging
Rathore et al. [87]	U	1996-2002	NPS, Wada	None: SF vs. NSF	151	85.4	ROC	No	No	No	Wada test limited use for predicting

													SF following ATL.
		Sinha et al. [91]	U	NR	C-D, MRI, DWI	ILAE: SF (1) vs. NSF (3-5)		51	58.8	ML	Yes	Yes	No	Node abnormality individual, noninvasive marker for better estimation of SF 1yr postop
		Voets et al. [99]	NR	NR	MRI	Engels: SF (E1) vs. NSF (E2- E4)		43	51.2	R	No	No	No	Folding complexity of bilateral temporal cortices in LTLE and of contralateral PHG in RTLE predictive of postoperative outcome.
cognitive outcomes	2011 – 2021	Audrain et al. [101]	U	NR	NPS, task f-MRI, rs-fMRI		Pre-post change score of naming test	20	100	R	Yes	No	No	FC matrix similarity better predictor of postop outcome than degree of to-be-resected regions, NC, modularity and language task fMRI laterality.
		Paff et al. [97]	U	2007-2018	NPS, MRI		Pre-post change of PC verbal memory and visuospatial memory scores	35	60.0	R	No	No	No	Left PHG atrophy predicted cognitive outcome (naming) in LTLE
clinical and cognitive outcomes	2000 – 2010	Uijl et al. [102]	NR	1997-2002	Wada	Engels: SF (E1a) vs. NSF (E2- E4)	Pre-post change scores of Verbal and performance IQ	138	75.4	R	Yes	No	No	Bilateral IAP has added value predicting postop verbal memory and IQ.
	2011 – 2021	Roger et al. [94]	U	2014-2019	NPS	Engels: SF (E1) vs. NSF (E2- E4)	RCI of naming test	47	51.1	ML	Yes	Yes	Yes	Specific NPS for language, memory and EF can globally predict SF.

Abbreviations: M = multicenter; U = uncenter; NR = not reported; C-D = clinical-demographic data; NPS = neuropsychological scores; srf. EEG = surface EEG; SF = seizure-free; NSF = non-seizure-free; PC = principal component; RCI = reliable change index; R = regression; DA = discriminant analysis; ROC = Receiver operating characteristic; ML = machine learning; pts = patients; CSI = Chemical shift spectroscopy imaging; SES = Socioeconomic status; RH = remote hypometabolism; FH = family history of epilepsy; FS = febrile seizures; SE = status epilepticus; IQ = intelligence quotient; LR = logistic regression; C-T = Connectome topography; FC = functional connectivity PHG = parahippocampal gyrus; EF = Executive functions; MRI = magnetic resonance imaging; RS = resting state.

Table 2. Distribution of analyses across ESOP studies in relation to the datasets.

	All ESOP studies		Clinical outcome		Cognitive outcome	
Analysis	Multimodal	Unimodal	Multimodal	Unimodal	Multimodal	Unimodal
Discriminant analysis	0 (0)	16.7% (2)	0 (0)	16.7% (2)	0 (0)	0 (0)
ROC	11.1% (1)	16.7% (2)	14.3% (1)	16.7% (2)	0 (0)	0 (0)
Regression	55.6% (5)	41. 7% (5)	42.9% (3)	41.7% (5)	100% (2)	50% (1)
ML	33.3% (3)	25% (3)	42.9% (3)	25% (3)	0 (0)	50% (1)

Abbreviations: ROC = Receiver-Operator Characteristic curve; ML = Machine Learning.

Table 3. PROBAST results for 21 selected ESOP studies. The notes for ROB and APP of each domain are based on the PROBAST questions presented in Table S3. The overall ROB and APP are based on the notes of each domain respectively. Table is adapted according to PROBAST guidelines [79, 81, 82].

Outcome	Period	Studies	ROB Participants	ROB Predictors	ROB Outcome	ROB Analysis	APP Participants	APP Predictors	APP Outcome	Overall ROB	Overall App
clinical outcomes	2000 – 2010	Burneo et al. [92]	+	+	+	+	+	+	+	+	+
		Dupont et al. [98]	+	+	+	?	+	+	+	+	+
		Eberhardt et al. [90]	+	+	+	?	+	+	+	+	+
		Krendl et al. [96]	+	+	+	?	+	+	+	+	+
		Wong et al. [88]	+	+	+	+	+	+	+	+	+
	2011 – 2021	Asadi-Pooya et al. [100]	+	+	+	?	+	+	+	+	+
		Baxendale et al. [84]	+	+	+	?	+	+	+	+	+
		Bonilha et al. [89]	+	+	+	+	+	+	+	+	+

		Chen et al. [83]	+	+	+	+	+	+	+	+	+
		Feis et al. [22]	+	+	+	+	+	+	+	+	+
		Gaça et al. [95]	+	+	+	?	+	+	+	+	+
		Ioriatti et al. [85]	+	+	+	+	+	+	+	+	+
		Keller et al. [93]	+	+	+	?	+	+	+	+	+
		Larivière et al. [86]	+	+	+	+	+	+	+	+	+
		Rathore et al. [87]	+	+	+	?	+	+	+	+	+
		Sinha et al. [91]	+	+	+	+	+	+	+	+	+
		Voets et al. [99]	+	+	+	+	+	+	+	+	+
cognitive outcomes	2011 – 2021	Audrain et al. [101]	+	+	+	+	+	+	+	+	+
		Paff et al. [97]	+	+	+	?	+	+	+	+	+

clinical and cognitive outcomes	2000 – 2010	Uijl et al. [102]	+	+	+	?	+	+	+	+	+
	2011 – 2021	Roger et al. [94]	+	+	+	+	+	+	+	+	+

Abbreviations: ROB = Risk of Bias; APP = Applicability; + = low concern; ? = unclear.