CHAPTER 2

CHAPTER 2

What is the current evidence on decision-making after referral for temporal lobe epilepsy surgery? A review of the literature.

SG Uijl

FSS Leijten

J Parra

JBAM Arends

AC van Huffelen

KGM Moons

Seizure 2005; 14(8):534-540.

Abstract

Objectives. Many patients thought to have temporal lobe epilepsy, are evaluated for surgical treatment. Decision-making in epilepsy surgery is a multidisciplinary, phased process involving complex diagnostic tests. This study reviews the literature on the value of different tests to decide on whether to operate.

Methods. Articles were selected when based on the consensus decision whether to perform temporal lobe surgery, or on the consensus localization or lateralization of the epileptic focus. The articles were scrutinized for sources of bias as formulated in methodological guidelines for diagnostic studies (STARD).

Results. Most studies did not fulfill the criteria, largely because they addressed prognostic factors in operated patients only. Ten articles met our inclusion criteria. In most articles a single test was studied; SPECT accounted for five papers. Unbiased comparison of the results was not possible.

Conclusion. Surprisingly little research in epilepsy surgery has focused on the decision-making process as a whole. Future studies of the added value of consecutive tests are needed to avoid redundant testing, enable future cost-efficiency analyses, and provide guidelines for diagnostic strategies after referral for temporal lobe epilepsy surgery.

Introduction

Epilepsy surgery is an established treatment for patients with seizures refractory to medical therapy. Temporal lobe epilepsy surgery in particular has a good outcome, with 70% of patients becoming seizure-free and 95% reaching a worthwhile reduction of seizure frequency of at least 90%.[13;14]

The diagnostic work-up to decide whether or not patients should undergo surgery is a consecutive, stepwise process focusing on the lateralization and localization of the epileptic focus, and risk factors that may compromise the surgical outcome.[23-26] A recent survey among epilepsy surgery centers worldwide showed that centers use the same phased diagnostic approach with more or less comparable techniques.[16]

During the last two decades, the number of tests in the diagnostic work-up has increased. It is recognized that different tests may provide overlapping information and that the risk of false-positive results increases with the number of tests used.[27-29] Although many diagnostic tests have been thoroughly evaluated, they were often studied in isolation. Given the consecutive diagnostic protocol, it is more important to know the relative or independent contribution of each consecutive diagnostic test to the decision-making process.[3;4] Other fields of medicine have shown that whereas a diagnostic test may be accurate, it may not have any added value to other tests and may thus be redundant for the diagnostic or therapeutic decision-making process.[30;31]

We searched for current evidence on the accuracy of different tests to the decision whether to perform temporal lobe epilepsy surgery or not. Thus, we searched for articles that studied the diagnostic, rather than prognostic, accuracy of one or more tests.

Methods

Selection of the current literature

A literature search was conducted using Medline, ScienceDirect and BioMedNet (January 1990 - March 2003) to identify publications on the diagnostic work-up regarding surgery in suspected temporal lobe epilepsy. We used combinations of the following keywords: epilepsy; temporal; temporal lobe / diagnosis; diagnos*; diagnostic techniques and procedures; prediction; presurg*; process assessment health care. The reference lists of retrieved papers and personal files were scrutinized for additional sources.

We looked for the most important methodological biases in diagnostic studies,[20;21;32-34] as recently stipulated in the STARD guidelines.[2;35] Table 2.1 shows the criteria we used that address outcome, description of assessed tests, and patient recruitment.

Table 2.1. Criteria used to select studies in hierarchical order

Criteria
1 Proper description of outcome: Decision to perform surgery yes / no using a consensus diagnosis *or* Localization of a temporal epileptic focus using a consensus diagnosis *or* Lateralization of a temporal epileptic focus using a consensus diagnosis.
2 Proper description of the diagnostic tests under study
3 Proper description of patient recruitment: Patients should be selected if temporal lobe epilepsy may be present (making them potential candidates for epilepsy surgery). Patients should not be selected on having undergone surgery (to avoid verification bias).

1. Outcome. Articles were included if they studied the diagnosis of an operable unilateral temporal focus or the decision to operate as outcome variable. In epilepsy surgery practice, unlike in other diagnostic areas, there is not a unique 'gold standard' or reference test to assess the final diagnostic outcome. In the absence of a single established reference standard, judgment of an expert panel

is ideally used as reference.[9] Fortunately, in epilepsy surgery the final diagnosis depends on such a consensus among a multidisciplinary team that takes into account all information from diagnostic test results and known prognostic and diagnostic factors. Accordingly, the consensus decision or the consensus diagnosis of the localization or lateralization of a unilateral temporal focus was used as outcome measure. Studies using a single test as outcome (e.g. MRI or invasive EEG monitoring) were excluded from this review.

2. Description of assessed tests. Studies were included only if they provided original data on the test results and described the tests under study. Overviews and review articles were excluded, but their references were checked.
3. Patient recruitment. We selected studies that included all patients suspected of having temporal lobe epilepsy who were analyzed for epilepsy surgery, and excluded studies focusing on only those patients who actually underwent surgery. This was to avoid verification or work-up bias leading to overestimation of the predictive values, sensitivity, and specificity of the diagnostic tests under study.[20;33;34] Essentially, the population should reflect the population of all referred candidates for epilepsy surgery as encountered in practice.

Statistical analysis

The sensitivity, specificity, positive likelihood ratio, and positive and negative predictive value of the diagnostic test studied in relation to the outcome (consensus decision for surgery or consensus localization or lateralization of the epileptic focus) were retrieved or calculated.

Results

Using the mentioned keywords, 654 articles were identified (table 2.2), of which only 102 reported on the diagnostic work-up of epilepsy surgery, with the final consensus decision or diagnosis as outcome. Most (86%) of the other 552 articles were excluded because of a different study outcome. Another 76 articles (14%)

used a single test, such as intracranial or video EEG monitoring, as reference test instead of a consensus diagnosis, and were therefore excluded. Of the 102 selected articles, 77 studied one or more diagnostic tests and provided original test results; 25 were either reviews or overviews. However, only 10 of the 77 articles fulfilled the stringent STARD criterion of adequate patient recruitment to avoid verification bias.[20;33;34]

Table 2.2. Inclusion of studies

Criteria		*Excluded*	*Included*
	Start search		654
1	Outcome	552	102
2	Assessed diagnostic test	25	77
3	Patient recruitment	67	10

These selected 10 papers are presented in table 2.3. Two studies used the decision whether or not to operate as study outcome,[36;37] seven dealt with the localization of the epileptic focus,[38-44] and one dealt with the lateralization of the epileptic focus.[45] Only one study included more than 100 patients.[38] All studies were retrospective, except for the study by Oliviera et al.[41]

Table 2.4 shows the sensitivity, specificity, positive likelihood ratio, and positive and negative predictive values of the assessed diagnostic tests. These estimates were either provided directly or calculated from the data provided. They could not be calculated from the article by Kilpatrick et al.[37] These authors did describe the diagnostic work-up until the decision for surgery, but presented the results for a selected group of operated patients only. Only one of the articles presented parameters of uncertainty (e.g. 95% confidence intervals). [41]

Table 2.3. Selected papers

Author (year)	*Study population*		*Diagnostic tests*										
	N	*Positive outcome*	*Clin. Exam.*	*MRI*	*InterI EEG*	*Video EEG*	*SPECT*	*PET*	*NPT*	*IntraC EEG*	*TPAS*	*Wada test*	*Surgical diagnosis*
Outcome: Decision for surgery													
Dellabadia[36] (2002)	69	33		✦✓		✦✓	✓	✦✓	✓				
Kilpatrick[37] (1997)	75	65		✦✓		✦✓	✦✓	✦✓	✦✓				
Outcome: Localization epileptic focus													
Henkel[38] (2002)	336	223		✓		✦✓	✓	✓					
Brekelmans[39] (1998)	82	60		✓	✓	✓		✓		✦			
O'Brien[40] (1999)	34	24		✓		✓	✦✓						
Oliveira[41] (1999)	48	43	✓	✓		✓	✦✓		✓			✓	
Tatum[42] (1995)	20	17		✓	✓	✓	✦		✓		✓	✓	✓
Velasco[43] (2002)	93	84		✓		✓	✦✓			✓			
Lee[44] (2002)	24	3		✓	✓	✓	✦✓	✓					✓
Outcome: Lateralization epileptic focus													
Ogden-Epker[45] (2001)	56	56							✦[a]				

N= total number of patients; ClinExam. = clinical examination; InterI EEG = interictal EEG; NPT = neuropsychological testing; IntraC EEG = Intracranial EEG; TPAS = thiopental activation study; Positive outcome: number of patients undergoing surgery or number of patients with localized or lateralized (operable) epileptic temporal focus; ✦ Test under study, of which the diagnostic accuracy was estimated; ✓ Tests used to set the outcome and form the consensus diagnosis, i.e. the tests that were included in reference standard; [a] The outcome was the consensus of the diagnostic work-up; which tests were included in this consensus was not specified.

Table 2.4. Accuracy parameters.

Authors	*Assessed diagnostic test*	*N*	*Sensitivity*	*Specificity*	*LR*	*Positive predictive value*	*Negative predictive value*
Standard tests, performed in all patients							
Dellabadia[36]	MRI	69	0.66	0.68	*2.06*	0.68	0.66
	Sleep-deprived EEG		0.66	0.68	*2.06*	0.68	0.66
Henkel[38]	Seizure semiology during video-EEG (Abdominal aura)	336	*0.52*	*0.88*	*4.33*	*0.90*	*0.49*
Ogden-Epker [45]	*Neuropsychological testing*	56	0.66	-	-	-	-
Ancillary tests, performed in a specific selection of patients							
Dellabadia[36]	PET	69	0.86	0.59	*2.10*	0.68	0.80
Brekelmans[39]	*Subdural EEG monitoring*	82	*0.60*	*0.82*	*3.33*	*0.90*	*0.43*
	Depth EEG monitoring		*0.87*	*0.55*	*1.93*	*0.84*	*0.60*
O'Brien[40]	Postictal SPECT	34	*0.83*	*0.10*	*0.92*	*0.69*	*0.20*
Oliveira[41]	Ictal SPECT	48	0.93 (0.79-1.06) [a]	0.93 (0.79-1.06) [a]	*13.08*	0.93 (0.79-1.06) [a]	0.93 (0.79-1.06) [a]
	Interictal SPECT		0.77 (0.64-1.09) [a]	0.65 (0.51-0.79) [a]	*2.20*	0.69 (0.56-0.82) [a]	0.74 (0.60-0.88) [a]
Tatum[42]	Interictal SPECT	20	0.67	0.25	*0.89*	-	-
Velasco[43]	*Interictal SPECT*	93	*0.81*	-	-	-	-
Lee[44]	1st + 2nd ictal SPECT	24	0.54	-	-	-	-

N = Total number of patients; LR = Likelihood ratio of positive test; Italic values: diagnostic parameters that were not given in the article, but retrieved by the authors from the results sections; - = Values could not be retrieved or estimated from the described results; [a] 95% confidence interval

Of the standard tests, only seizure semiology obtained from video EEG appeared to have good diagnostic accuracy, i.e. identified patients suitable for surgery with a relatively high specificity, likelihood ratio, and positive predictive value. Sleep-deprived EEG and neuropsychological testing had a rather poor diagnostic accuracy. Surprisingly, MRI also showed modest positive and negative predictive values. Of the ancillary tests, usually performed when standard tests provide conflicting results, SPECT was investigated in five articles. From the papers that met our methodological stringent criteria, only ictal SPECT showed a relatively high diagnostic accuracy (sensitivity, specificity, positive, and negative predictive values all 0.93, likelihood ratio 13.08). By contrast, interictal and postictal SPECT had a rather poor diagnostic accuracy. Intracranial monitoring with subdural electrodes showed a relatively high specificity, likelihood ratio, and positive predictive value and PET appeared to be useful for excluding candidates from surgery, having a high sensitivity and negative predictive value.

DellaBadia et al. were the only investigators who evaluated combinations of tests.[36] They assessed eligibility for surgery after one or more positive interictal tests, after two or more positive interictal tests, and after three positive interictal tests. This, however, was regardless of which interictal test was included. They showed that the sensitivity decreased from 0.97 to 0.40 when more interictal tests were positive, while the specificity increased from 0.35 to 0.91.

Discussion

This methodological study searched the available literature on the value of diagnostic tests for the decision whether or not to perform temporal epilepsy surgery. Applying stringent STARD criteria, we conclude that there are surprisingly few unbiased studies in the literature that deal with decision making in epilepsy surgery. Notwithstanding the importance of seminal papers on epilepsy surgery that did not meet our criteria, our review shows that the information currently available in the literature is not sufficient to quantify the relative or independent contribution of each consecutive diagnostic test in the decision-making for epilepsy surgery. Below, we will discuss some methodological issues and clinical implications of our findings.

Methodological issues

The focus of this review was on the whole decision-making process for temporal lobe epilepsy. We used a limited number of stringent criteria, dealing with the most important sources of bias in diagnostic research in general. We then had to exclude the majority of articles, including some of the seminal articles on epilepsy surgery. Most of these articles were written with a different perspective, focusing only on the outcome after surgery. Other studies compared diagnostic tests for interchangeability.

In total, 84% (552 of 654 articles) of the articles were excluded because the outcome criterion of the study was inappropriate for our purposes. Most of the published articles have limited their focus on the prognostic accuracy of tests to predict the outcome of surgery, such as seizure freedom one or two years after surgery. Although these articles are useful and have influenced diagnostic practice,[46-51] they do not primarily deal with the diagnostic decision-making. Although the prognostic value of a test may be an important factor in the decision-making process, taking into account operated patients only introduces a verification or work-up bias. The decision not to operate on potentially good candidates for surgery may be based on diagnostic factors that are not detected when studying operated patients only. Unfortunately, many of the prognostic studies did not include data on non-operated patients, which would have enabled us to include these papers in our review.

A number of articles assessed interchangeability of tests, and therefore used one test, e.g. chronic intracranial monitoring, as a reference, instead of a consensus from all tests. Such studies contain verification bias as only a sample of patients will undergo the (invasive) reference procedure.[20;33;34]

Clinical implications

Unfortunately, the 10 diagnostic articles that meet our criteria do not reflect current practice of work-up for temporal epilepsy surgery. Most studies deal with SPECT, which is not a routine investigation for candidates for temporal lobe epilepsy surgery, whereas only one article concerned MRI which is performed in every patient. It was not possible to compare the predictive power of the tests described in these studies (table 4) because of the considerable variation in the results, as shown for example by the variability in the sensitivity of ictal SPECT.[41;44] Moreover, ictal SPECT could be performed in only 29% of patients in one study (Oliveira et al.), which highlights the difficulty of performing ictal SPECT in practice.[41]

There is little information on the predictive value of the most basic tests used in surgical decision-making, such as medical history, standard EEG, and MRI. Video-EEG monitoring and dedicated MRI techniques are often used to guide the surgical decision-making process without much being known about the independent contribution of these tests to decision-making. The study by Henkel et al. addressed only one aspect of video-EEG monitoring.[38] MRI has been extensively studied in relation to prognosis of the outcome of epilepsy surgery only. We know that MRI evidence of unilateral hippocampal atrophy is a potent predictor of a good postoperative outcome,[46;49-51] but this is not necessarily a good diagnostic indicator to set a decision for surgery as evidenced by its modest values for sensitivity and specificity reported by Dellabadia et al. (see table 4).[36]

A recent article by the Multicentre Study Group of Epilepsy Surgery described the presurgical decision-making process for epilepsy surgery in general, and the factors influencing the decision to have surgery in a qualitative manner.[52] None of the 10 articles we reviewed addressed the issue of the added diagnostic value of commonly used tests for surgical decision-making. However, such studies do exist regarding the prognosis of epilepsy surgery. For example, Armon et al. performed a multivariate analysis of the predictors of outcome of surgery, assessing the added value of different predictors.[47] Study designs such as these should as well be applied to the decision-making process to tailor the diagnostic approach in a more cost-effective manner.[3;4]

To answer the question which diagnostic tests truly contribute to decision making in epilepsy surgery and in which order these tests should be performed, the following study design would be desirable.[2-4;20;21;32-35] All epilepsy patients who are potential candidates for temporal lobe epilepsy surgery should be included during a specific period. All these patients should undergo the diagnostic tests in the chronological order commonly applied in clinical practice. The results of each test should be documented for each patient. For each patient, the final decision 'surgery or not' should be made by a multidisciplinary team using the consensus diagnosis method, again in accordance with clinical practice. This decision will be based on all patient information (including known diagnostic and prognostic factors, but probably also, as yet insufficiently studied factors) and can be considered as the reference test result. Hence, for all patients the results of the tests under study as well as the reference outcome is known. This allows a multivariate analysis and modeling of the decision making process, and show which test parameters independently contribute to the final decision 'surgery or not'. Furthermore, such a design makes it possible to characterize specific subgroups of patients requiring a minimum number of tests.

We conclude that few articles have tried to quantify the relevancy of tests for surgical decision-making in temporal lobe epilepsy patients. Knowledge of which test parameters really contribute to determine eligibility as well as ineligibility for surgery is necessary before burdensome, costly, and risky tests can be replaced by more convenient ones. Such knowledge will be essential to future cost-efficiency analyses of epilepsy surgery.[53] It will also allow us to provide tailored clinical guidelines of diagnostic strategies for patients referred for temporal lobe epilepsy surgery.[3;4] And finally, for the future of epilepsy surgery in developing countries where many facilities are lacking,[53;54] such information could be the basis for good risk-benefit assessment without extensive testing.