Received: 30 May 2018 | Accepted: 27 September 2018

DOI: 10.1111/epi.14588

Epilepsia®

FULL-LENGTH ORIGINAL RESEARCH

Diagnostic accuracy of functional magnetic resonance imaging, Wada test, magnetoencephalography, and functional transcranial Doppler sonography for memory and language outcome after epilepsy surgery: A systematic review

Elisabeth Schmid[1,2,3] | Aljoscha Thomschewski[1,2,3,4] | Alexandra Taylor[1,2] | Georg Zimmermann[1,3,5] | Margarita Kirschner[1] | Teia Kobulashvili[1] | Francesco Brigo[6,7] | Matea Rados[8] | Christoph Helmstaedter[9] | Kees Braun[8] | Eugen Trinka[1,3,4] | on behalf of the E-PILEPSY consortium

[1]Department of Neurology, Christian Doppler Medical Center, Paracelsus Medical University, Salzburg, Austria

[2]Department of Psychology, Paris Lodron University of Salzburg, Salzburg, Austria

[3]Spinal Cord Injury and Tissue Regeneration Center Salzburg, Salzburg, Austria

[4]Center for Cognitive Neuroscience, Salzburg, Austria

[5]Department of Mathematics, Paris Lodron University of Salzburg, Salzburg, Austria

[6]Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

[7]Department of Neurology, Franz Tappeiner Hospital, Merano, Italy

[8]Department of Child Neurology, University Medical Center, Utrecht, The Netherlands

[9]Department of Epileptology, Medical Center, University of Bonn, Bonn, Germany

Correspondence
Eugen Trinka, Department of Neurology, Christian Doppler Medical Center, Paracelsus Medical University, Salzburg, Austria.
Email: e.trinka@salk.at

Summary

Objective: The European Union–funded E-PILEPSY project was launched to develop guidelines and recommendations for epilepsy surgery. In this systematic review, we aimed to assess the diagnostic accuracy of functional magnetic resonance imaging (fMRI), Wada test, magnetoencephalography (MEG), and functional transcranial Doppler sonography (fTCD) for memory and language decline after surgery.

Methods: The literature search was conducted using PubMed, Embase, and CENTRAL. The diagnostic accuracy was expressed in terms of sensitivity and specificity for postoperative language or memory decline, as determined by pre- and postoperative neuropsychological assessments. If two or more estimates of sensitivity or specificity were extracted from a study, two meta-analyses were conducted, using the maximum ("best case") and the minimum ("worst case") of the extracted estimates, respectively.

Results: Twenty-eight papers were eligible for data extraction and further analysis. All tests for heterogeneity were highly significant, indicating large between-study variability ($P < 0.001$). For memory outcomes, meta-analyses were conducted for Wada tests (n = 17) using both memory and language laterality quotients. In the best case, meta-analyses yielded a sensitivity estimate of 0.79 (95% confidence interval [CI] = 0.67-0.92) and a specificity estimate of 0.65 (95% CI = 0.47-0.83). For the worst case, meta-analyses yielded a sensitivity estimate of 0.65 (95% CI = 0.48-0.82) and a specificity estimate of 0.46 (95% CI = 0.28-0.65). The overall quality of evidence, which was assessed using Grading of Recommendations Assessment, Development, and Evaluation methodology, was rated as very low. Meta-analyses concerning diagnostic accuracy of fMRI, fTCD, and MEG were not feasible due to small numbers of studies (fMRI, n = 4; fTCD,

n = 1; MEG, n = 0). This also applied to studies concerning language outcomes (Wada test, n = 6; fMRI, n = 2; fTCD, n = 1; MEG, n = 0).

Significance: Meta-analyses could only be conducted in a few subgroups for the Wada test with low-quality evidence. Thus, more evidence from high-quality studies and improved data reporting are required. Moreover, the large between-study heterogeneity underlines the necessity for more homogeneous and thus comparable studies in future research.

KEYWORDS

diagnostic accuracy, epilepsy surgery, language, memory, systematic review

1 | INTRODUCTION

In 2014, the European Union funded E-PILEPSY, a pilot network of 28 reference centers for refractory epilepsy and epilepsy surgery (http://www.ucl.ac.uk/www.e-pilepsy.eu). Its overall objectives are to enhance access to epilepsy surgery in Europe and to increase the number of patients cured of drug-resistant epilepsy. In a first step, the current practices in brain imaging and electromagnetic source localization procedures,[1] long-term video-electroencephalographic monitoring,[2] and neuropsychological assessments[3] were evaluated. In a second step, the network aimed to create recommendations and guidelines for surgical evaluation and epilepsy surgery based on the best available evidence.

Epilepsy surgery is an elective procedure considered to be an effective treatment for patients with drug-resistant epilepsy.[4] However, patients may experience postoperative cognitive impairments.[5,6] After temporal lobe resection, which is the most common type of epilepsy surgery,[4] memory and language impairments have been reported.[5,7] The observed memory impairments tend to be material-specific (verbal/visual) depending on language lateralization.[6] After temporal lobe resection involving the speech-dominant hemisphere, verbal memory decline is more consistent and well documented[8] as compared to visual memory loss in the nondominant hemisphere.[8,9] In a systematic review by Sherman et al,[5] an estimated risk of 44% for verbal memory decline after left-sided temporal lobe surgery was reported (vs 20% after right-sided surgery). For visual memory, no difference with regard to side of surgery was found (21% after left-sided surgery vs 23% after right-sided surgery). Furthermore, language impairments have been reported in 34% of patients with left-sided temporal lobe surgery.[5]

To estimate the risk of postoperative memory and language impairments, various methods have been applied to examine the lateralization and localization of language and/or memory functions preoperatively. The intracarotid amobarbital test, or so-called selective Wada test,[10] is still considered the gold standard for assessing language lateralization.[11] However, memory lateralization and its predictive value for postoperative decline are less valid,[12–16] as memory testing during selective Wada test assesses more than mesial temporal lobe functions.[16] Furthermore, aphasia may have a major impact on verbal memory testing during cortical anesthesia of the speech-dominant hemisphere.[17] Thus, the superselective Wada test was developed, in which barbiturate is injected into the posterior cerebral artery[18] or anterior choroidal artery.[19] This enables memory testing while preserving language functions. Noninvasive alternatives conducted in epilepsy centers for presurgical evaluation of language and memory lateralization include functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and functional transcranial Doppler sonography (fTCD).[3] The diagnostic accuracy of these methods for postoperative language and memory decline has been the focus of numerous studies. However, most studies only report mean differences in group data or correlations as outcome parameters, thus making it difficult to estimate the individual risk for possible postoperative decline in clinical practice.

Key Points

- Diagnostic accuracy of fMRI, Wada test, MEG, and fTCD was expressed in terms of sensitivity and specificity of each method
- Meta-analyses could be conducted for the Wada test only; overall quality of evidence was rated as very low
- High variability exists regarding protocols, stimuli, neuropsychological tests, and assessment of language and memory functions
- Substantial between-study heterogeneity indicates the need for more comparable studies
- The majority of papers could not be included in the analysis due to insufficient data reporting, thus emphasizing the need for guidelines

Therefore, the objective of this systematic review was to assess the diagnostic accuracy of selective and superselective Wada test, fMRI, MEG, and fTCD for memory and language decline after epilepsy surgery in terms of sensitivity and specificity.

2 | MATERIALS AND METHODS

The study protocol was published on PROSPERO, an international prospective register of systematic reviews (CRD42016043927)[20] in July 2016. The results of this systematic review are reported in agreement with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)[21] statement.

2.1 | Literature search strategy and data extraction

An extensive literature search was conducted on October 30, 2016, within the PubMed, Embase, and CENTRAL databases (for the search syntaxes, see Appendix S1). All abstracts that met initial inclusion criteria (see Table 1) were identified, and articles of possible relevance were then further reviewed when inclusion criteria for data extraction (reported estimates of sensitivity and specificity or individual subject data allowing for retrospective calculation of these estimates were available) were met.

In the case of multiple papers reporting results of the same data or subgroup datasets, the study with the largest sample size was included. When there were several neuropsychological datasets concerning different cognitive domains, or when more than one test per cognitive domain or more than one score per test was reported within one study, all datasets were included. In the case of studies reporting multiple postoperative datasets, the dataset with the longest follow-up interval was selected.

TABLE 1 Inclusion criteria for systematic review of diagnostic accuracy of fMRI, Wada test, MEG, and fTCD for memory and language outcomes after epilepsy surgery

- Original research article or (systematic) reviews containing original research data.
- Patients with epilepsy who underwent resective epilepsy surgery.
- Pre- and postsurgical neuropsychological examinations for memory and/or language.
- Data from neuropsychological methods (fMRI and/or selective and/or superselective Wada test and/or MEG and/or fTCD) conducted preoperatively.
- The minimum sample size is $n \geq 5$.
- The minimum age of patients included is 12 years.
- English-language publication.

fMRI, functional magnetic resonance imaging; fTCD, functional transcranial Doppler sonography; MEG, magnetoencephalography.

The screening of abstracts and full text as well as the data extraction was independently performed by a combination of two of three reviewers (E.S., A.Th., and A.Ta.). Discrepancies were resolved by discussion to achieve a consensus decision or by referral to the third reviewer. Extracted data comprised study characteristics (eg, author, year, study design) as well as patient characteristics (eg, sex, age at onset, age at surgery). Furthermore, detailed characteristics of the index tests (fMRI, selective/superselective Wada test, fTCD, MEG) as well as the reference tests (neuropsychological tests) were extracted. Outcome measures were extracted in terms of sensitivity and specificity of the index tests for cognitive decline in a specific domain (memory and/or language) assessed by the reference tests (pre- and postoperative neuropsychological test scores). A full list of extracted variables is provided in Table S1. The quality appraisal of each study was done using the qualitative assessment of diagnostic accuracy tool (QUADAS-2).[22] The cumulative quality of evidence was rated using an adapted version of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE)[23] methodology with due regard to International League Against Epilepsy (ILAE) recommendations.[24]

2.2 | Statistical methods

The diagnostic accuracy was expressed in terms of sensitivity and specificity, which were based on the reported laterality quotients (LQs) of the respective index tests. Sensitivity (true-positive rate) was defined as *true-positive / (true-positive + false-negative)*; specificity (true-negative rate) was defined as *true-negative / (false-positive + true-negative)*. For classification of postoperative outcome in relation to index test results and its predictive value, see Table S2.

Cognitive decline was assessed by calculating the difference in standardized pre- and postoperative neuropsychological test scores with a decrease in >1 SD defined as significant. If estimates of sensitivity and specificity for significant postoperative decline were not reported by the study authors, estimates were calculated independently by both reviewers.

Univariate meta-analyses for both sensitivity and specificity were carried out for various subgroups of our database, depending on the number of studies eligible for the respective analyses. A random-effect model was used, and estimation was carried out with the DerSimonian-Laird algorithm. In addition, the Knapp-Hartung adjustment was applied. For sensitivity and specificity estimates that were either 0 or 1, SDs were calculated by applying the add two successes and two failures rule.[25] Because only a small number of papers could be included, neither a further differentiation between material-specific memory functions

(verbal/visual memory) nor other subgroup analyses were possible. However, if two or more estimates of sensitivity or specificity were extracted from a certain study for one specific domain, two meta-analyses were conducted, one using the minimum of all estimates extracted from that study (referred to as "worst case") and another taking the maximum of all estimates from that study (referred to as "best case"). For studies reporting only a single estimate of sensitivity or specificity per domain, that value was included in both analyses. Conversely, if the minimum or the maximum was not unique, the value with the smallest associated variance was used. Statistical analyses were carried out using the software environment R (v3.1.1, R Core Team 2014).[26]

3 | RESULTS

The literature search resulted in 3538 articles (654 duplicates). The interrater reliability and percentage of agreement were moderate to good (Cohen kappa 0.49 for abstract and 0.71 for full-text screening; for more detailed results and a detailed list of agreement rates between the reviewers, see Tables S3 and S4).

Ninety-eight papers met the predefined initial inclusion criteria (see Table 1). However, only 29 papers were eligible for data extraction and further analyses, primarily due to lack of reported data in the majority of studies (for PRISMA flow diagram, see Figure S1). An overview of the excluded papers is given in Table S5. In most cases, only group data or correlations were presented (n = 56). Thus, estimates of sensitivity and specificity could not be calculated retrospectively. Other reasons for exclusion were multiple reports of the same sample in different articles (n = 7) and no standardized or adequate neuropsychological assessments (n = 6). One paper met all inclusion criteria for data extraction; however, it was excluded after a consensus decision was made by the reviewers: Mikuni et al[27] described five patients with mesial temporal lobe epilepsy and left cerebral dominance for language and memory functions as assessed by selective Wada test. After the selective Wada test procedure, the basal temporal language area was further evaluated before epilepsy surgery by implantation of long-term subdural electrodes.[27] Due to this procedure, the surgical outcome was considered to be biased by the reviewers, making a comparison with other studies unfeasible. Therefore, 28 papers reporting data from 941 patients were included in the final analyses (for an overview, see Table S6) and from these papers, 60 datasets could be extracted. Several papers reported results from either more than one method (index test), more than one cognitive domain (language, visual/verbal memory), or more than one neuropsychological test or subtest. The majority of papers reported data from selective Wada tests (n = 22) with 46 extracted datasets. Six papers reported results from fMRI examinations (nine extracted datasets), one from fTCD examinations (four extracted datasets), and one from superselective Wada test with barbiturate injected into the anterior choroidal artery (one extracted dataset). Regarding MEG, no eligible papers could be found.

An overview of the extracted datasets, grouped according to methods and paradigms/tests used for calculating LQs and their diagnostic accuracy, is shown in Figure S2. Because datasets for more than one method (fMRI and Wada test) could only be extracted in two studies, no statistical comparison between various methods applied on the same patient sample could be conducted. For further analyses, the extracted datasets were divided into the domains in which postoperative outcomes were predicted: memory outcomes (n = 49) and language outcomes (n = 11; see Figure S3). In the majority of cases, language LQs were used to predict memory outcomes calculated by selective Wada test results (n = 21 vs n = 18 using memory LQs).

All authors reported cohort studies, and the majority of the data collection was retrospective (n = 25). Only two prospective studies and one mixed design were included. All patients in the included studies had temporal lobe epilepsy (TLE). The mean age at surgery was 34 years (in 14/60 datasets it was not reported), and the mean duration of epilepsy was 21 years (in 24/60 datasets it was not reported). A summary of design and patient characteristics from all papers included in the final analyses is provided in Table 2. Overall, all tests for heterogeneity were highly significant ($P < 0.001$), indicating a high between-study variability. Moreover, high diversity was found among the neuropsychological tests and protocols used for Wada tests and fMRI.

Regarding pre- and postoperative neuropsychological testing, a variety of tests were used for language (n = 4), verbal memory (n = 9), and visual memory (n = 6) evaluation (see also Table 2). Moreover, within different memory tests, various domains (eg, learning, free recall, recognition) were examined. Differences also exist with regard to the timeframe when postoperative follow-up examinations were conducted (mean = 17 months, SD = 23, minimum = 2, maximum = 120; in 9/60 datasets it was not reported).

When a specific design was reported, a high diversity in Wada test protocols was also observed. Nine studies (of 23) did not report details about the protocols used. The remaining 14 studies (including the superselective Wada test study) all reported different protocols and different stimuli items (see Table S7). Furthermore, there was great variability with respect to the methods used for the LQ calculation. Fifteen studies did not report a specific formula, and the remaining eight studies reported a variety of calculation formulas (see Table S8). The only consensus was found regarding the drug used, as all but two studies (in

TABLE 2 Overview of studies included in final analysis (n = 28)

Study	Selection criteria	N	Mean age at surgery, y (SD)	Surgery in speech-dominant hemisphere, %	Percentage with best seizure outcome	Methods	Neuropsychological reference tests
Andersson-Roswall et al (2012)	N/A	51	33.69	45.1	53 (seizure-free)	Language Wada	The Claeson-Dahl Learning and Retention Test The Cronholm-Molander Memory Test
Bartha et al (2004)	Left TLE	10	43.10 (5.3)	100	100 (Engel I)	Language Wada	Aachen Aphasia Test
Baxendale et al (2007)	Typical language representation	91	N/A	N/A	N/A	Memory Wada	Adult Memory and Information Processing Battery
Binder et al (2008)	IQ > 70	11	38.55 (11.74)	N/A	N/A	Language fMRI	Selective Reminding Test
Busch et al (2008)	Left TLE, IQ > 70, right-handed	10	27.90 (13.88)	90	60 (Engel I)	Memory Wada	WMS (verbal and visual memory)
Dodrill & Ojemann (1997)	N/A	25	30.40	N/A	36 (seizure-free)	Memory Wada	WMS (verbal memory)
Dulay et al (2009)	IQ > 75, left speech dominance	75	30.62	44	80 (Engel I)	Language Wada	Selective Reminding Test Nonverbal Selective Reminding Test
Dupont et al (2010)	Atypical neuropsychological profile	23	40.61 (8.49)	30.43	91.30 (seizure-free)	Memory Wada	Marilyn Jones-Gotman Verbal Learning Test and/or Rey Auditory Verbal Learning Test; Rey-Taylor Complex Figure Test and/or Aggie Figure Learning Test
		24	40.67 (8.79)	29.17	91.67 (seizure-free)	Language Wada Memory fMRI	Marilyn Jones-Gotman Verbal Learning Test and/or Rey Auditory Verbal Learning Test; Rey-Taylor Complex Figure Test and/or Aggie Figure Learning Test
Elshorst et al (2009)	Left TLE	59	32.50 (9.20)	N/A	79.66 (Engel I)	Memory Wada	CVLT or Rey Auditory Verbal Learning Test
Hamberger et al (2005)	Left speech dominance, left TLE	19	31.50 (13.13)	100	73.68 (Engel I)	Language Wada	BNT
Hori et al (2007)	N/A	18	32.50 (8.51)	61.11	50 (Engel I)	Language Wada	WAIS (verbal IQ)
Janecek et al (2013)	Bilateral Wada	22	34.96	100	78.27 (Engel I)	Language Wada	BNT
Janecek et al (2013)	Discordant Wada and fMRI results	10	37.30 (10.78)	90	80 (seizure-free)	Language fMRI	BNT

(Continues)

TABLE 2 (Continued)

Study	Selection criteria	N	Mean age at surgery, y (SD)	Surgery in speech-dominant hemisphere, %	Percentage with best seizure outcome	Methods	Neuropsychological reference tests
Kubu et al (2000)	N/A	10	40.50	80	40 (seizure-free)	Language Wada	WMS (verbal memory)
							WMS (visual memory)
						Memory Wada	WMS (verbal memory)
							WMS (visual memory)
		8	40.50	87.50	50 (seizure-free)	Language Wada	Rey-Osterrieth Complex Figure
						Memory Wada	Rey-Osterrieth Complex Figure
Langfitt & Rausch (1996)	English native speaker	59	29.33	38.98	64.41 (Engel I)	Language Wada	BNT
LoGalbo et al (2005)	IQ > 70, right-handed	10	30.10 (8.24)	100	N/A	Language Wada	CVLT
Mani et al (2008)	Left language dominance, left TLE	31	N/A	100	N/A	Memory Wada	WMS (verbal and visual memory)
						Language Wada	WMS (verbal and visual memory)
Martens et al (2014)	N/A	40	N/A	47.50	N/A	Language fTCD	Rey Auditory Verbal Learning Test
		41	N/A	46.34	N/A	Language fTCD	Word Generation Task
		21	N/A	52.38	N/A	Language fTCD	Aachen Aphasia Test
		43	N/A	44.19	N/A	Language fTCD	DCS
Mayanagi et al (2001)	N/A	18	25.28 (5.72)	61.11	N/A	Language Wada	WMS (verbal memory)
Pravata et al (2014)	N/A	5	26.20 (14.86)	N/A	100 (seizure-free)	Language fMRI	WAIS (verbal IQ)
Richardson et al (2004)	Right-handed; English native speaker; left TLE	10	32.80 (8.69)	N/A	100 (Engel I)	Memory fMRI	Adult Memory and Information Processing Battery
Rössler et al (2015)	Left speech and verbal memory dominance	7	37.71 (15.13)	100	85.71 (Engel I)	Language Wada	Berlin Amnesia Test
		12	41 (16.46)	100	91.67 (Engel I)	Language fMRI	Berlin Amnesia Test
Sabsevitz et al (2001)	Nonlesional left TLE, left speech dominance, IQ > 69	21	34.33	100	76.19 (Engel I)	Memory Wada	Selective Reminding Test
Setoain et al (2004)	N/A	38	30	60	85 (Engel I)	Memory Wada	WMS (verbal and visual memory)
Stroup et al (2003)	Unilateral speech dominance	132	37.80 (9.80)	44	78 (seizure-free)	Language Wada	CVLT, WMS (verbal memory)
Vulliemoz et al (2008)	N/A	9	31.11 (14.04)	N/A	N/A	Memory Wada	Rey Auditory Verbal Learning Test

(Continues)

TABLE 2 (Continued)

Study	Selection criteria	N	Mean age at surgery, y (SD)	Surgery in speech-dominant hemisphere, %	Percentage with best seizure outcome	Methods	Neuropsychological reference tests
Wyllie et al (1991)	Left speech dominance	20	29	100	N/A	Language Wada	WMS (verbal memory)
		17	29	0	N/A	Language Wada	WMS (visual memory)
Yu et al (2010)	Left speech dominance	68	30 (7.90)	45.59	72.06 (seizure-free)	Memory Wada Language Wada	WAIS (verbal IQ) WAIS (verbal IQ)

Only studies included in final analyses are shown.

BNT, Boston Naming Test; CVLT, California Verbal Learning Test; DCS, Diagnosticum für Cerebralschädigung; fMRI, functional magnetic resonance imaging; fTCD, functional transcranial Doppler sonography; IQ, intelligence quotient; N, number of patients; N/A, not available; TLE, temporal lobe epilepsy; WAIS, Wechsler Adult Intelligence Scale; WMS, Wechsler Memory Scale.

1/23 studies it was not reported) reported the use of amobarbital (with the other two reporting a combination of amobarbital and methohexital [Brevital]). Moreover, selective Wada tests were conducted bilaterally (in 4/23 studies it was not reported), whereas the superselective Wada test was carried out unilaterally.

Regarding fMRI, in four (of six) studies, a description of the paradigms used was provided, with the other two referring to paradigms used in previous publications. Altogether, five different paradigms were described (see Table S9). For the calculation of the LQs, in five studies, two different formulas as well as study-specific estimates were reported (one study did not give any information; see Table S10).

With regard to fTCD, the only paper included did not state specific details concerning paradigms, used stimuli items, or calculation of hemispheric dominance.

The risk of bias within individual studies was assessed using the QUADAS-2 tool.[22] None of the included studies was free from bias. A high risk of bias was observed regarding patient selection (82%), index tests (75%) and flow and timing (50%). Lower risk of bias was found regarding the reference standard (14%). Further details on risk of bias and applicability assessments are shown in the supplementary material (Figure S4).

3.1 | Memory outcomes

Due to the small sample of included papers (n = 20), no further differentiation between material-specific (verbal/visual) memory functions was possible. The results from various protocols were used to calculate the LQs with multiple formulas, which were then used to calculate sensitivity and specificity.

Meta-analyses could only be conducted for Wada tests (n = 17) using both memory and language LQs. Because superselective Wada test results were reported by only one study, no further differentiation between the different types of Wada tests was made and all datasets were included in the meta-analyses. In the best case, meta-analyses yielded a sensitivity estimate of 0.79 (95% confidence interval [CI] = 0.67-0.92) and a specificity estimate of 0.65 (95% CI = 0.47-0.83). In the worst case, meta-analyses yielded a sensitivity estimate of 0.65 (95% CI = 0.48-0.82) and a specificity estimate of 0.46 (95% CI = 0.28-0.65). Furthermore, subgroup analyses were conducted for studies using either language (n = 10) or memory LQs (n = 10). Sensitivity was higher for language LQs: 0.89 (95% CI = 0.81-0.98) for best case and 0.74 (95% CI = 0.52-0.97) for worst case versus 0.68 (95% CI = 0.49-0.86) and 0.56 (95% CI = 0.33-0.79) for memory LQs. However, specificity was generally higher for memory LQs: 0.70 (95% CI = 0.47-0.94) for best case and 0.66 (95% CI = 0.43-0.88) for worst case

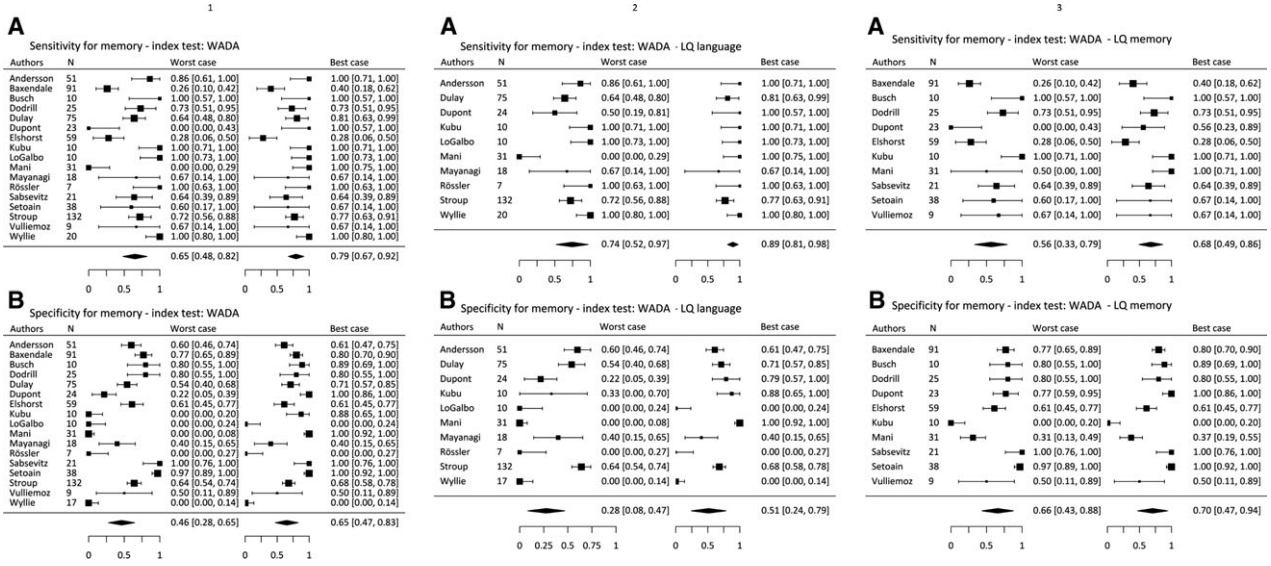

FIGURE 1 Memory: pooled estimates of sensitivity and specificity and 95% confidence intervals in "best case" as well as "worst case" for prediction of postoperative memory decline by Wada test for (1) all studies combined, (2) studies using language laterality quotients (LQs) for calculating estimates of sensitivity and specificity, and (3) for studies using memory LQs for calculating estimates of sensitivity and specificity

versus 0.51 (95% CI = 0.24-0.79) and 0.28 (95% CI = 0.08-0.47) for language LQs. Nevertheless, substantial overlaps of confidence intervals indicate a high variability of these estimates. For detailed results, see Figure 1.

The quality of evidence was assessed using the GRADE[23] approach. The baseline quality was defined as moderate, as only retrospective studies could be included. The overall quality of evidence was rated as very low across all conducted meta-analyses. The GRADE evidence profiles for each conducted meta-analysis are presented in the supplementary material (Tables S11-S16).

Results of meta-analyses without the superselective Wada test (thus only selective Wada test results included) can be found in the supplementary material (Table S17). In summary, no substantial differences regarding pooled estimates were found.

For fMRI and fTCD, a meta-analysis was not feasible due to the small number of eligible studies (for results, see Table 3).

3.2 | Language outcomes

For the prediction of language outcome, only nine studies met inclusion criteria. Thus, meta-analyses for estimates of sensitivity (eight datasets) and specificity (11 datasets) were not feasible. Moreover, studies reported data from three different methods: Wada test (n = 6), fMRI (n = 2), and fTCD (n = 1). We therefore only report study-specific sensitivities and specificities together with 95% CIs (Table 3).

Three studies[28–30] reported verbal intelligence quotient (IQ) changes as outcome parameters predicted by language[28,30] as well as language and memory Wada test LQs.[29] Because no subtest results were available, the overall verbal IQ score was included in the analyses.

A complete overview of the different neuropsychological tests that were used in each study to calculate pre- and postoperative changes in relation to the methods used to calculate sensitivity and specificity can be found in the supplementary material (Tables S18-S25).

4 | DISCUSSION

The aim of this systematic review was to evaluate the diagnostic accuracy of fMRI, selective and superselective Wada test, MEG, and fTCD for language and memory decline after epilepsy surgery and to subsequently provide a comprehensive summary of evidence and develop recommendations regarding our research questions. However, only a small number of studies were eligible for respective analyses and pooled estimates could only be analyzed for Wada test and memory decline. Although the added value of the Wada test as a routine examination in comparison with fMRI has been discussed numerous times (for review, see Bauer et al,[11] Binder,[31] Dym et al,[32] Massot-Tarrús et al[33]), only a few studies evaluated and compared its diagnostic accuracy (for an overview, see Szaflarski et al[34]). Our meta-analyses indicate a higher sensitivity of language LQs compared with memory LQs in predicting postsurgical memory decline, as has previously been reported.[17,35] However, the substantial overlaps in the CIs between studies must be considered, and caution is needed when drawing conclusions. Furthermore,

TABLE 3 Study-specific estimates of sensitivities and specificities with 95% CIs for each study not included in meta-analyses

Method	Study	N	LQ	Sensitivity estimate and 95% CI	Specificity estimate and 95% CI
Memory outcomes					
fMRI	Binder et al (2008)	11	Language	1.00 (0.78-1.00)	1.00 (0.67-1.00)
	Dupont et al (2010)	24	Memory	0.30 (0.03-0.57)	0.43 (0.18-0.68)
				0.00 (0.00-0.43)	0.52 (0.32-0.72)
	Richardson et al (2004)	10	Memory	1.00 (N/A)	0.40 (N/A)
				1.00 (N/A)	1.00 (N/A)
				1.00 (N/A)	0.40 (N/A)
	Rössler et al (2015)	12	Language	1.00 (0.63-1.00)	0.00 (0.00-0.18)
fTCD	Martens et al (2014)	40	Language	0.67 (0.30-1.00)	0.56 (0.38-0.74)
		43	Language	0.56 (0.38-0.74)	0.44 (0.11-0.77)
Language outcomes					
Wada test	Bartha et al (2004)	10	Language	1.00 (0.71-1.00)	0.00 (0.00-0.25)
	Hamberger et al (2005)	19	Language	1.00 (0.75-1.00)	0.00 (0.00-0.16)
	Hori et al (2007)	18	Language	N/A	0.33 (0.11-0.55)
	Janecek et al (2013)	22	Language	1.00 (0.76-1.00)	0.00 (0.00-0.14)
	Langfitt & Rausch (1996)	59	Language	0.88 (0.64-1.00)	0.69 (0.57-0.81)
	Yu et al (2010)	68	Language	0.47 (0.22-0.72)	0.55 (0.41-0.69)
			Memory	0.00 (0.00-0.24)	1.00 (0.92-1.00)
fMRI	Janecek et al (2013)	10	Language	1.00 (0.73-1.00)	0.00 (0.00-0.27)
	Pravata et al (2014)	5	Language	N/A	0.00 (0.00-0.27)
fTCD	Martens et al (2014)	41	Language	0.38 (0.05-0.71)	0.52 (0.34-0.70)
		21	Language	N/A	0.48 (0.26-0.70)

CI, confidence interval; fMRI, functional magnetic resonance imaging; fTCD, functional transcranial Doppler sonography; LQ, laterality quotient; N, number of patients; N/A, not available.

estimates of sensitivity and specificity were calculated from reported LQs, which have been derived from a variety of different calculation formulas. Moreover, a high heterogeneity between studies was observed, making a comparison difficult both between and within methods.

Regarding possible postoperative language decline, no meta-analyses could be conducted. Furthermore, those studies we could include in our review only report surgeries in non–language-relevant areas. Series reporting declines in language functions after extratemporal surgeries are rare, although such series must be assumed to involve language-eloquent cortex more likely than temporal lobe surgery. In general, negative sequels like aphasia or global amnesia must be suggested as a complication rather than an expected cognitive risk. However, reliable numbers of such catastrophic outcomes are not available. Moreover, the best information possible to obtain from the Wada test or other index tests would be evidence that memory/language is not represented in the pathological hemisphere and evidence as to whether a patient is at risk of decline, but not to which degree such decline can be expected.[36,37] Apparently, even the Wada test—which is considered the gold standard—does not provide a satisfying predictive accuracy, although it is a highly demanding and invasive procedure. Thus, it should only be considered in patients in whom suggested eloquent cortices might be affected by surgery, as is already common practice in Europe.[38] If surgery is, for example, done close to or within potentially eloquent language cortex, the resection borders need to be determined by use of intracranial electrocorticoencephalography or stimulation during awake surgery.

Concerning pre- and postoperative neuropsychological examinations, a variety of tests have been used to objectively monitor postsurgical outcomes. A high diversity among neuropsychological diagnostic tests for presurgical evaluation has already been demonstrated in various surveys[39,40]; the most recent one has been conducted within the framework of the E-PILEPSY project.[3] Currently, there are no recommendations regarding specific tests to be used, although the ILAE has recently updated the concepts and principles of neuropsychological assessments in epilepsy patients.[41] Another aspect that should be considered is the relatively short follow-up

period for postoperative neuropsychological examinations reported in reviewed studies. After an initial decline in the early postoperative stages,[42] memory functions stabilize 2 years after surgery,[43] with no further decline reported after 6 years.[43,44] Therefore, a longer follow-up period should be considered in future study proposals.

With regard to the index tests, no standardizations have been established yet either. Therefore, as demonstrated in our review, a high heterogeneity regarding protocols, stimuli, calculation formulas of LQs, analyses, et cetera exist. Even for the Wada test, first described in 1948,[10] no universally accepted standard protocol for either language or memory lateralization exists (for a review, see Haag et al[38]). Therefore, a variety of heterogeneous protocols and stimuli are used across centers. This may in particular affect the results of memory assessments[12,45] and make comparisons difficult. As for fMRI, again, no standardized protocols exist and various stimuli for language and memory fMRI are applied across centers using different paradigms. Various studies comparing multiple fMRI paradigms also showed great diversity regarding the validity of fMRI in presurgical evaluation.[35,46,47] A recently published survey focusing on clinicians' use of language fMRI by Benjamin and colleagues[48] also emphasized the importance of further studies comparing commonly used fMRI paradigms to predict postsurgical decline.

In a first attempt to establish practice guidelines, the American Academy of Neurology[34] recently published evidence-based recommendations on diagnostic accuracy and prognostic value of fMRI in the presurgical evaluation of patients with epilepsy. However, in terms of predictive value of fMRI for memory or language outcome after surgery, these recommendations are based on a few studies only (language, n = 2; verbal memory, n = 12; visual memory, n = 1), with small sample size and heterogeneous characteristics. Thus, the strength of recommendations is mostly rated as level C (possibly effective). In addition, and contrary to our systematic review, studies reporting only correlations as outcome parameters have been included. Overall, in agreement with our conclusion, the authors emphasize the need for further research and evidence regarding, among other things, comparisons between different methods (fMRI vs Wada test) and fMRI language and memory tasks with regard to lateralization of functions and prediction of postsurgical outcomes, or various fMRI analyses.

Limitations of our systematic review include the lack of differentiation between material-specific (verbal/visual) memory functions, which should be addressed in future reviews. Due to the limited number of studies included, different surgical approaches (eg, standard anterior temporal lobe resections vs selective amygdalohippocampectomy), which may have an influence on surgical outcome,[7,49] could not be considered. Furthermore, other factors such as antiepileptic medication were also not controlled for across studies. Moreover, in many studies, a biased patient population with regard to speech dominance was included, thus not allowing for a general interpretation of the results. Because only patients with TLE could be included, conclusions regarding extra–temporal lobe epilepsy cannot be drawn either. This also greatly restricted our attempt to examine the diagnostic accuracy for language decline. In addition, conclusions regarding children cannot be drawn, as we only included patients aged 12 years and older. Furthermore, only two prospective studies were included; hence, conclusions are primarily derived from retrospective analyses. There is also a potential bias due to our inclusion criteria, as full-text articles that were not available in English were not considered for inclusion.

Moreover, due to the small number of articles eligible for analyses, our conclusions are drawn from only a small part of existing evidence. Therefore, more meta-analyses also including group data, correlations, and effect sizes regarding hemispheric lateralization and/or prediction of postoperative memory and language outcomes are needed. Finally, although our systematic review focused on commonly used procedures, other methods such as transcranial magnetic stimulation (TMS)[50] were not included and should be systematically evaluated in the future. However, a literature search conducted in July 2018 by our group regarding our research question but also including TMS as a method of interest identified no additional studies that would have been eligible for inclusion and further analyses in this systematic review.

In general, substantial heterogeneity in terms of varying protocols, methods, included populations, et cetera may severely challenge the goal of developing valid recommendations for the prediction of cognitive decline after epilepsy surgery. Furthermore, the majority of studies that met the inclusion criteria could not be included in further analyses due to an extensive lack of reported data. We therefore want to emphasize the need for methodological recommendations on proper data reporting, which researchers should adhere to in future studies. We also urge authors to publish data at the individual patient level, to allow retrospective analyses of sensitivity and specificity. In a first attempt, we propose a checklist of requirements specifying which data should be reported in future studies in this research area (see Table 4). Since we conducted our literature search in October 2016, no further studies that would have been eligible for analyses have been added to the PubMed database (search conducted in July 2018); this further underlines the apparent need for recommendations regarding data reporting in publications.

Compliance with minimal standards as well as the implementation of standardized protocols across multiple centers will be crucial to objectively assess the values of prognostic methods such as Wada test, fMRI, fTCD, and MEG in predicting cognitive outcomes after epilepsy surgery and further allow valid comparisons among various methods.

TABLE 4 Checklist: Requirements that should be adhered to when reporting data in studies

General subject characteristics (eg, age, gender) at individual patient level
Epilepsy-related characteristics including type of epilepsy, age at onset, seizure types, description of structural lesions, drugs, and dosage at individual patient level
Surgery-related characteristics: Type of surgery, outcome, and complications at individual patient level
Neuropsychological data: Exact test scores at individual patient level, names and descriptions of used tests, examined domains (subtests), versions, language, norms, time between repeated testing.
Index test–related:
Exact LQs for every subject, exact description of procedure, exact description of paradigms and stimuli used, LQ calculations, and cutoff scores.
fMRI-related parameters: MRI scanner, teslas, correction procedures, software for analyses used, et cetera.
Wada test–related parameters: artery, drug, dosage, control (electroencephalogram), pretests, order of investigated hemispheres (if bilateral), time between injections (if bilateral), et cetera.
fTCD-related parameters: machine, pulse description, et cetera.
MEG-related parameters: number of channels, preprocessing, investigated frequencies, artifact correction, et cetera.

fMRI, functional magnetic resonance imaging; fTCD, functional transcranial Doppler sonography; LQ, laterality quotient; MEG, magnetoencephalography.

5 | CONCLUSION

Meta-analyses were only possible in certain subgroups for Wada tests. Those few analyses indicate that language LQs derived from Wada tests seem to be more sensitive in predicting memory decline in patients with TLE than memory LQs. However, the overall quality of evidence was rated as very low. We further identified a general lack of data, insufficient standardization regarding paradigms/tests/protocols, and missing reports of individual subject data as the main challenges, preventing us from drawing general conclusions and developing recommendations for cognitive outcome prediction after epilepsy surgery by fMRI, Wada test, fTCD, and MEG. Moreover, the high between-study heterogeneity indicates the need for more homogeneous and thus more comparable studies across centers. Having sufficient numbers of high-quality publications with results reported at the individual patient level and efforts toward a standardization of neuropsychological testing are therefore considered key requirements for developing evidence-based guidelines and subsequently enabling comparisons between various diagnostic methods with respect to their diagnostic accuracy in future research.

ACKNOWLEDGMENTS

This work arose from the project E-PILEPSY, which has received funding from the European Union (grant agreement 20131203), in the framework of the Health Program (2008-2013).

DISCLOSURE OF CONFLICTS OF INTEREST

G.Z. has received financial support for information technology equipment and conference travel from Eisai, not connected to the present work. F.B. has received speaker honoraria from Eisai and PeerVoice, payment for consultancy from Eisai, and travel support from Eisai, ITALFARMACO, and UCB Pharma. C.H. has received honoraria from Eisai, GW Pharmaceuticals, UCB, and Desitin for advisory board activities, talks, and expert testimony, as well as funding from the European Union (E-Pilepsy, EpiCare) not connected to the present work. The remaining authors have no conflicts of interest to report. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

AUTHOR CONTRIBUTIONS

E.T. initiated the study, provided guidance in designing and implementing the study, and supervised the work. E.S. and A.Th. proposed the study plan and implemented the study. M.K., G.Z., F.B., K.B., and T.K. provided help in planning the study and supervising the work. F.B., M.R., and A.Ta. provided help in the literature search. E.S., A.Th., and A.Ta. participated in screening of included studies and performed data extraction. E.S. and M.K. performed quality appraisal of included studies. G.Z. and E.S. carried out statistical analyses. E.S. wrote the first draft of the paper, implemented comments from coauthors, and created the final version of the manuscript. All authors critically revised the manuscript and approved the final version.

ORCID

Georg Zimmermann http://orcid.org/0000-0002-8282-1034
Teia Kobulashvili http://orcid.org/0000-0001-9213-1355
Francesco Brigo http://orcid.org/0000-0003-0928-1577

REFERENCES

1. Mouthaan BE, Rados M, Barsi P, et al. Current use of imaging and electromagnetic source localization procedures in epilepsy surgery centers across Europe. Epilepsia. 2016;57:770–6.
2. Kobulashvili T, Höfler J, Dobesberger J, et al. Current practices in long-term video-EEG monitoring services: a survey among

partners of the E-PILEPSY pilot network of reference for refractory epilepsy and epilepsy surgery. Seizure. 2016;38:38–45.
3. Vogt VL, Äikiä M, Del Barrio A, et al. Current standards of neuropsychological assessment in epilepsy surgery centers across Europe. Epilepsia. 2017;58:343–55.
4. Wiebe S. Effectiveness and safety of epilepsy surgery: what is the evidence? CNS Spectr. 2004;9:120–32.
5. Sherman EM, Wiebe S, Fay-McClymont TB, et al. Neuropsychological outcomes after epilepsy surgery: systematic review and pooled estimates. Epilepsia. 2011;52:857–69.
6. Helmstaedter C, Brosch T, Kurthen M, et al. The impact of sex and language dominance on material-specific memory before and after left temporal lobe surgery. Brain. 2004;127:1518–25.
7. Helmstaedter C. Cognitive outcomes of different surgical approaches in temporal lobe epilepsy. Epileptic Disord. 2013;15:221–39.
8. Lee TM, Yip JT, Jones-Gotman M. Memory deficits after resection from left or right anterior temporal lobe in humans: a meta-analytic review. Epilepsia. 2002;43:283–91.
9. Witt JA, Coras R, Schramm J, et al. Relevance of hippocampal integrity for memory outcome after surgical treatment of mesial temporal lobe epilepsy. J Neurol. 2015;262:2214–24.
10. Wada J, Rasmussen T. Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observations. J Neurosurg. 1960;17:266–82.
11. Bauer PR, Reitsma JB, Houweling BM, et al. Can fMRI safely replace the Wada test for preoperative assessment of language lateralisation? A meta-analysis and systematic review. J Neurol Neurosurg Psychiatry. 2014;85:581–8.
12. Binder JR, Sabsevitz DS, Swanson SJ, et al. Use of preoperative functional MRI to predict verbal memory decline after temporal lobe epilepsy surgery. Epilepsia. 2008;49:1377–94.
13. Lineweaver TT, Morris HH, Naugle RI, et al. Evaluating the contributions of state-of-the-art assessment techniques to predicting memory outcome after unilateral anterior temporal lobectomy. Epilepsia. 2006;47:1895–903.
14. Loddenkemper T, Morris HH, Lineweaver TT, et al. Repeated intracarotid amobarbital tests. Epilepsia. 2007;48:553–8.
15. Ojemann JG, Kelley WM. The frontal lobe role in memory: a review of convergent evidence and implications for the Wada memory test. Epilepsy Behav. 2002;3:309–15.
16. Helmstaedter C, Kurthen M. Validity of the WADA test. Epilepsy Behav. 2002;3:562–3.
17. Kirsch HE, Walker JA, Winstanley FS, et al. Limitations of Wada memory asymmetry as a predictor of outcomes after temporal lobectomy. Neurology. 2005;65:676–80.
18. Jack CR Jr, Nichols DA, Sharbrough FW, et al. Selective posterior cerebral artery injection of amytal: new method of preoperative memory testing. Mayo Clin Proc. 1989;64:965–75.
19. Vulliemoz S, Pegna AJ, Annoni JM, et al. The selective amobarbital test in the anterior choroidal artery: perfusion pattern assessed by intraarterial SPECT and prediction of postoperative verbal memory. Epilepsy Behav. 2008;12:445–55.
20. NHS National Institute for Health Research. PROSPERO: International Prospective Register of Systematic Reviews.https://www.crd.york.ac.uk/PROSPERO/. Accessed: August 28, 2018.
21. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009;62:1006–12.
22. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–36.
23. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64:401–6.
24. Sauro KM, Wiebe S, Perucca E, et al. Developing clinical practice guidelines for epilepsy: a report from the ILAE Epilepsy Guidelines Working Group. Epilepsia. 2015;56:1859–69.
25. Agresti A, Booth JG, Hobert JP, et al. Random-effects modeling of categorical response data. Sociol Methodol. 2000;30:27–80.
26. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation; 2004.
27. Mikuni N, Miyamoto S, Ikeda A, et al. Subtemporal hippocampectomy preserving the basal temporal language area for intractable mesial temporal lobe epilepsy: preliminary results. Epilepsia. 2006;47:1347–53.
28. Hori T, Yamane F, Ochiai T, et al. Selective subtemporal amygdalohippocampectomy for refractory temporal lobe epilepsy: operative and neuropsychological outcomes. J Neurosurg. 2007;106:134–41.
29. Yu HY, Shih YH, Su TP, et al. The Wada memory test and prediction of outcome after anterior temporal lobectomy. J Clin Neurosci. 2010;17:857–61.
30. Pravatà E, Sestieri C, Colicchio G, et al. Functional connectivity MRI and post-operative language performance in temporal lobe epilepsy: initial experience. Neuroradiol J. 2014;27:158–62.
31. Binder JR. Functional MRI is a valid noninvasive alternative to Wada testing. Epilepsy Behav. 2011;20:214–22.
32. Dym RJ, Burns J, Freeman K, et al. Is functional MR imaging assessment of hemispheric language dominance as good as the Wada test?: a meta-analysis. Radiology. 2011;261:446–55.
33. Massot-Tarrús A, Mousavi SR, Mirsattari SM. Comparing the intracarotid amobarbital test and functional MRI for the presurgical evaluation of language in Epilepsy. Curr Neurol Neurosci Rep. 2017;17:54.
34. Szaflarski JP, Gloss D, Binder JR, et al. Practice guideline summary: use of fMRI in the presurgical evaluation of patients with epilepsy: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. Neurology. 2017;88:395–402.
35. Binder JR, Swanson SJ, Sabsevitz DS, et al. A comparison of two fMRI methods for predicting verbal memory decline after left temporal lobectomy: language lateralization versus hippocampal activation asymmetry. Epilepsia. 2010;51:618–26.
36. Dodrill CB, Ojemann GA. An exploratory comparison of three methods of memory assessment with the intracarotid amobarbital procedure. Brain Cogn. 1997;33:210–23.
37. Jokeit H, Ebner A, Holthausen H, et al. Individual prediction of change in delayed recall of prose passages after left-sided anterior temporal lobectomy. Neurology. 1997;49:481–7.
38. Haag A, Knake S, Hamer HM, et al. The Wada test in Austrian, Dutch, German, and Swiss epilepsy centers from 2000 to 2005: a review of 1421 procedures. Epilepsy Behav. 2008;13:83–9.
39. Djordjevic J, Jones-Gotman M. Inquiry on assessments across epilepsy centers in different countries. In: *Neuropsychology in the Care of People with Epilepsy*. Helmstaedter C, Hermann B, Lassonde M, et al, eds. Paris, France: John Libbey Eurotext; 2011:13–26.

40. Witt J, Helmstaedter C. A survey on neuropsychological practice in German-speaking epilepsy centers. In: *Neuropsychology in the Care of People with Epilepsy*. Helmstaedter C, Hermann B, Lassonde M, et al, eds. Paris, France: John Libbey Eurotext; 2011:27–31.
41. Wilson SJ, Baxendale S, Barr W, et al. Indications and expectations for neuropsychological assessment in routine epilepsy care: report of the ILAE Neuropsychology Task Force, Diagnostic Methods Commission, 2013-2017. Epilepsia. 2015;56:674–81.
42. Baxendale S, Thompson PJ, Duncan JS. Neuropsychological function in patients who have had epilepsy surgery: a long-term follow-up. Epilepsy Behav. 2012;23:24–9.
43. Alpherts WC, Vermeulen J, van Rijen PC, et al. Verbal memory decline after temporal epilepsy surgery?: a 6-year multiple assessments follow-up study. Neurology. 2006;67:626–31.
44. Andersson-Roswall L, Engman E, Samuelsson H, et al. Cognitive outcome 10 years after temporal lobe epilepsy surgery: a prospective controlled study. Neurology. 2010;74:1977–85.
45. Loring DW, Hermann BP, Perrine K, et al. Effect of Wada memory stimulus type in discriminating lateralized temporal lobe impairment. Epilepsia. 1997;38:219–24.
46. Binder JR, Swanson SJ, Hammeke TA, et al. A comparison of five FMRI protocols for mapping speech comprehension systems. Epilepsia. 2008;49:1980–97.
47. Avila C, Barrós-Loscertales A, Forn C, et al. Memory lateralization with 2 functional MR imaging tasks in patients with lesions in the temporal lobe. AJNR Am J Neuroradiol. 2006;27:498–503.
48. Benjamin CFA, Li AX, Blumenfeld H, et al. Presurgical language fMRI: clinical practices and patient outcomes in epilepsy surgical planning. Hum Brain Mapp. 2018;39:2777–85.
49. Gross RE, Mahmoudi B, Riley JP. Less is more: novel less-invasive surgical techniques for mesial temporal lobe epilepsy that minimize cognitive impairment. Curr Opin Neurol. 2015;28:182–91.
50. Papanicolaou AC, Rezaie R, Narayana S, et al. On the relative merits of invasive and non-invasive pre-surgical brain mapping: new tools in ablative epilepsy surgery. Epilepsy Res. 2018;142:153–5.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Schmid E, Thomschewski A, Taylor A, et al. Diagnostic accuracy of functional magnetic resonance imaging, Wada test, magnetoencephalography, and functional transcranial Doppler sonography for memory and language outcome after epilepsy surgery: A systematic review. *Epilepsia*. 2018;00:1–13. https://doi.org/10.1111/epi.14588