The Clinical Diagnostic Utility of Electrophysiological Techniques in Assessment of Patients With Disorders of Consciousness Following Acquired Brain Injury: A Systematic Review

S. L. Hauger, Cand Psychol; A.-K. Schanke, PhD; S. Andersson, PhD; C. Chatelle, PhD; C. Schnakers, PhD; M. Løvstad, PhD

**Objective:** To investigate the diagnostic utility of electrophysiological recordings during active cognitive tasks in detecting residual cognitive capacities in patients with disorders of consciousness (DoC) after severe acquired brain injury. **Design:** Systematic review of empirical research in MEDLINE, Embase, PsycINFO, and Cochrane from January 2002 to March 2016. **Main Measures:** Data extracted included sample size, type of electrophysiological technique and task design, rate of cognitive responders, false negatives and positives, and excluded subjects from the study analysis. The Quality Assessment of Diagnostic Accuracy Studies–2 (QUADAS-2) was used for quality appraisal of the retrieved literature. **Results:** Twenty-four studies examining electrophysiological signs of command-following in patients with DoC were identified. Sensitivity rates in healthy controls demonstrated variable accuracy across the studies, ranging from 71% to 100%. In patients with DoC, specificity and sensitivity rates varied in the included studies, ranging from 0% to 100%. Pronounced heterogeneity was found between studies regarding methodological approaches, task design, and procedures of analysis, rendering comparison between studies challenging. **Conclusion:** We are still far from establishing precise recommendations for standardized electrophysiological diagnostic procedures in DoC, but electrophysiological methods may add supplemental diagnostic information of covert cognition in some patients with DoC. **Key words:** attention, cognition, consciousness disorder, diagnostic errors, electroencephalography, electrophysiology, minimally conscious state, persistent vegetative state, sensitivity specificity

**Author Affiliations:** Department of Research, Sunnaas Rehabilitation Hospital, Norway (Mrs Hauger and Drs Løvstad and Schanke); Department of Psychology, University of Oslo, Oslo, Norway (Drs Andersson, Løvstad, and Schanke); Laboratory for NeuroImaging of Coma and Consciousness, Massachusetts General Hospital, Boston, and Acute Neurorehabilitation Unit, Department of Clinical Neurosciences, University Hospital of Lausanne, Switzerland (Dr Chatelle); and Department of Neurosurgery, University of California, Los Angeles (Dr Schnakers).

This study was supported by the Norwegian Extra Foundation for Health and Rehabilitation through EXTRA funds (grant 2012/2/0084). The funding partner has no involvement in the methodological aspects of the project, such as the design of the reviews’ protocol, data extraction, analyses, and interpretation of the data.

Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal’s Web site (www.headtraumahab.com).

The authors declare no conflicts of interest.

**Corresponding Author:** Solveig Lægreid Hauger, Cand Psychol, Department of Research, Sunnaas Rehabilitation Hospital, 1450 Nesoddtangen, Norway (solveig.laegreidhauger@sunnnaas.no/solveig.hauger@gmail.com).

DOI: 10.1097/HTR.0000000000000267

**DEVELOPMENTS** in neuroimaging and electrophysiological methods have allowed both structural and functional studies of the living brain, enabling online monitoring of mental processes, including the neural correlates of human behavior. Hence, much of contemporary evidence and theories of brain processes are informed by neuroimaging techniques, offering insight into age-old questions about brain-behavior relationships, and an emerging understanding of underlying neural mechanisms. Although previously regarded as scientifically intractable, consciousness can now be studied with modern neuroscientific techniques, such as positron emission tomography (PET), functional (fMRI) and structural (diffusion tensor imaging), magnetic resonance imaging (MRI), and electrophysiological techniques.

In parallel with this methodological development, a great increase in scientific interest has taken place...
with respect to patients with disorders of consciousness (DoC) following severe acquired brain injury, that is, patients in either a vegetative (VS) state, also referred to as the “unresponsive wakefulness syndrome” (UWS), or a minimally conscious state (MCS). Whereas the VS is characterized by absence of any behavioral signs of awareness, but regained intermittent wakefulness, the MCS, by contrast, is characterized by the presence of inconsistent, but clearly discernible behavioral evidence of awareness of self or the environment (ie, visual pursuit, localization to pain, or reproducible command-following). Recently, the MCS entity has been suggested to be divided into MCS+ and MCS−, depending on the complexity of behavioral responses. While MCS+ is characterized by more complex cognitive capacities, that is, presence of command-following, MCS− is, on the contrary, characterized by nonlinguistic and simple signs of consciousness. However, consensus on a clear definition of MCS+ and MCS− is currently lacking. Novel neuroimaging and electrophysiological techniques have offered new insight and enhanced theoretical understanding of these patients’ level of consciousness, brain connectivity, and metabolic and cognitive functioning.

The current standard approach to clinical diagnosis of DoC is based upon behavioral assessment strategies, along with patient history and structural brain imaging. Notably, rates of misdiagnosis in DoC have been estimated to be as high as approximately 40%. The lack of a “gold standard” for detection of conscious awareness in DoC is a prominent confounding factor for accurate diagnostic assessment, and it is recommended to apply standardized neurobehavioral rating scales designed to detect subtle, but clinically significant signs of consciousness. In a comprehensive evidence-based review of the psychometric properties of existing assessment scales, the Coma Recovery Scale—Revised (CRS-R) was recommended with minor reservation, while the Sensory Modality Assessment Technique (SMART), Western Neuro Sensory Stimulation Profile (WNSSP), Sensory Stimulation Assessment Measure (SSAM), Wessex Head Injury Matrix (WHIM), and Disorders of Consciousness Scale (DOCS) were recommended with moderate reservations.

CLINICAL DIAGNOSTIC UTILITY OF ELECTROPHYSIOLOGICAL METHODS IN PATIENTS WITH DOC

Advances in neuroscientific methodology have led to optimism regarding potential clinical utility in diagnostic and prognostic considerations in patients with DoC, in part due to several studies indicating that cognitive processing can be detected with imaging techniques in the absence of behavioral signs of consciousness. These studies applied tasks that require subjects to exert mental responses to command, in contrast to merely passive paradigms eliciting only “automatic” responses. Hence, to infer consciousness, it is necessary to include tasks involving active cognitive processing in combination with functional neuroimaging and electrophysiological methods. However, functional imaging methods, such as fMRI and PET, require high levels of technical skills, are expensive, and most often are not readily accessible in rehabilitation facilities. On the contrary, electrophysiological techniques are more readily available by having the benefit of being of low-cost and noninvasive and can be conducted repeatedly at bedside. Herein, event-related potentials (ERPs) represent time-locked electroencephalographic (EEG) activity elicited by external events, thus providing a neurophysiological correlate of cognitive processing at the millisecond level, from early and largely sensory components to later and cognitively mediated waveforms, such as the P3.

Task-related systematic changes in oscillatory variation can also be an index of cognitive effort and can be detected through the analysis of frequency bands, such as, event-related desynchronization (ERD). Such electrophysiological features or activation patterns can also be applied in machine learning systems that allow quantification of differences in neural responses at an individual level. Surface electromyogram (EMG) is, on the contrary, recordings of electrical activity in muscles and is a commonly used tool to study physiological principles of muscles related to movement generation.

OBJECTIVES OF THE SYSTEMATIC REVIEW

Although modern functional imaging and EEG-based techniques have given rise to hopes of improved diagnostic accuracy in DoC, the body of existing systematic reviews and overview articles have various shortcomings in providing a sufficient estimate of the clinical usefulness of neurophysiological measures. A major limitation of existing reviews is the lack of reports regarding rates of responders, meaning subjects showing signs of active mental effort during electrophysiological assessment, in both healthy subjects and patients with DoC, and also an insufficient account of false negatives, that is, the rate of persons who do not display clear signs of cognitive effort in electrophysiological assessments, despite definite voluntary behavioral responses. Some reviews lack a representative body of included studies, either due to overly strict study inclusion criteria regarding sensitivity/specificity, while others have not required the use of active paradigms, rendering degree of consciousness uninterpretable. Yet other papers only provide a topical overview without explicit systematic
Methods of the analysis and inclusion criteria were specified in advance and documented in a protocol, adhering to established recommendations for conducting systematic reviews, including the PRISMA guidelines. The full review protocol can be accessed in the Supplemental Digital Content 1 (available at: http://links.lww.com/JHTR/A193), as well as PRISMA checklist in Supplemental Digital Content 2 (available at: http://links.lww.com/JHTR/A194). Studies were included in the systematic review if they involved electrophysiological methods used in combination with experimental paradigms encompassing active conditions. Furthermore, only English empirical studies with more than 5 subjects were included. Studies were included if they investigated patients who met the diagnostic criteria for VS and MCS after acquired brain injury, where level of consciousness was established with a standardized behavioral assessment tool with acceptable psychometric properties, that is, either the CRS-R, WHIM, SSAM, WNSSP, DOCS, or SMART scales. A further inclusion criterion required publication after the consensus-based criteria for diagnosing MCS, published in 2002. Literature reviews and systematic reviews were excluded.

Search method for identification of studies

We undertook a systematic review of the literature and selected relevant studies published between January 2002 and March 2016 in the following databases: MEDLINE, Embase, PsycINFO, Database of Abstracts of reviews of effects (Cochrane Library), and Cochrane Central Register of Controlled Trials (Cochrane Library). Primary search terms used defining DoC were Consciousness disorder, disorder of consciousness, vegetative state, persistent vegetative state, unresponsive wakefulness syndrome, or minimally conscious state. Primary terms were paired with secondary terms defining aspects of electrophysiological measurement: electrodiagnosis, electrophysiology, neurophysiology, electroencephalography, encephalogram, EEG, myography, or electromyography. These were furthermore paired with third terms related to measure outcome: Event Related Potentials, ERP, evoked potentials, P300, active task/condition/paradigm, residual function, covert attention/awareness/cognition or command-following.

We last searched the electronic databases on March 7, 2016. See Supplemental Digital Content 3 (available at: http://links.lww.com/JHTR/A195) for a full description of MEDLINE search strategy. As studies were identified, researchers also checked for additional relevant articles being cited.

Study selection and analysis

Selection of studies

Titles and abstracts were reviewed first, and when indicating relevance, full-text articles were assessed using the inclusion and exclusion criteria to exclude those papers that were not relevant to this review. The initial selection was conducted by one author (S.L.H.) and double-checked by an independent second author (M.L.). Any disagreements were resolved by consensus, and if no agreement could be reached, it was planned that a third author (S.A.) would decide. One study author was contacted for additional information regarding clarification of the included study sample. Data were extracted by S.L.H. and verified by M.L.

Quality appraisal of retrieved literature

Quality appraisal of the retrieved literature was conducted using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). The initial assessment
was conducted by author S.L.H. and verified by a second author (M.L.). The QUADAS-2 checklist assesses the risk of bias and concerns regarding applicability over 4 domains: patient selection, index test, reference standard, and flow and timing.59 see Supplemental Digital Content 4 for QUADAS-2 questions (available at: http://links.lww.com/JHTR/A196). Patient selection was regarded to be at high risk of bias if the study did not primarily include patients in a medically stable phase, or in cases of insufficient differential diagnosis, that is, from coma or locked-in-syndrome, was not based on a consecutive or random sample, or did not clearly avoid inappropriate exclusion, that is, outpatients or concurrent referrals. Unblinded interpretation of the electrophysiological assessment and lack of detailed descriptions of procedures for processing of EEG data and experimental procedures were considered to represent a high risk of bias concerning the electrophysiological index test. The reference standard was considered to be at high risk if the behaviorally based diagnostic conclusion did not adhere to established consensus-based diagnostic criteria for VS and MCS,10,11 and if the interpretation of the behavioral assessment was not blinded to the results of the electrophysiological assessment. Concerns regarding applicability were related to the representativeness of the studies in relation to the review questions, such as sample representatives, clearness and relevance of processing and interpretation of electrophysiological data in assessing consciousness, and adherence to diagnostic criteria for DoC.

**RESULTS**

**Characteristics of the included studies**

As illustrated in Figure 1, a total of 832 articles were initially identified from the search process and 9 were identified through other sources. Twenty-four studies were finally included for review. The characteristics of these studies are summarized in Supplemental Digital Content Table 1 (available at: http://links.lww.com/JHTR/A199).

**Included study samples**

Of the 24 studies, 7 did not include a healthy control group for the active paradigm,60–66 whereof 4 referred to previously published healthy control data.60,61,63,65 The studies varied considerably with regard to sample sizes, from only 6 included patients60,61 to a total of 167 electrophysiological recordings acquired from 113 patients in the largest study.67 Overall, many studies were characterized by small sample sizes.

**Behavioral assessment tool**

All studies applied the CRS-R as the behavioral assessment scale of choice, except for one, where WHIM was applied.62 Hence, the included studies represented uniform and sound procedures for behavioral diagnosis of consciousness.

**Electrophysiological techniques**

The included studies displayed a wide variation with regard to applied index tests. The majority of the studies applied EEG-based technology, while 2 included studies used experimental tasks with EMG.62,68 Ten studies applied systems using EEG in combination with machine-learning, where algorithms were used to identify “patterns” of brain activity using a classifier method (for a review of classifier methods, see the study69). A subgroup of studies applied complex multivariate classifier methods, integrating data from a variety of electrophysiological features based on recordings during active tasks, for example, ERPs, frequency power, complexity, and connectivity measures.67,70

**Design/task**

The systematic review revealed considerable heterogeneity with regard to types of active experimental paradigms applied. The majority of tasks fell into 2 main categories, either imagery tasks or tasks requiring counting an auditory target stimuli, while only 3 studies involved visual stimuli.71–73 Five imagery tasks included instructions to imagine motor movements, for example, squeezing hand, moving toes, or moving arm toward an object.60,61,63,71,74 Fourteen studies included the active instruction to count either a target name...
Figure 1. Flowchart of the selection of articles.

or word, occurrence of deviant tones, or a target global deviant. The latter has been repeatedly studied in a “local-global” paradigm consisting of a series of tone sequences containing a 2-level structure of occasional irregularities in short-term (“local”) violations within a 5-sound sequence, and long-term (“global”) violations of the expectancies of such sequences. Seven studies included subjectively relevant stimuli, for example, photo of the subject, a customized familiar motor imagery task, or the subject’s own name (SON), where SON was applied in 5 studies. All experimental tasks included verbally delivered instructions.

Excluded subjects

Not all studies provided information of whether subjects were excluded from analysis or not. Notably, some studies reported high rates of excluded subjects in the patient group. For example, Gibson and colleagues reported exclusion of 5 of 11 patients from the EEG analysis. Chennu and colleagues reported exclusion of 9 out of 30 recruited patients, and in the study of Faugeras and colleagues, a total of 35 out of 100 patient EEG recordings were excluded. Data exclusion was mainly due to low quality of EEG recordings, and excessive noise artifacts in patients with DoC, demonstrating one of the intrinsic limitations of this approach. Also, exclusion of EEG data from healthy controls due to artifacts was explicitly reported in 2 studies.

Diagnostic performance

Table 1 illustrates calculated rates of sensitivity and specificity per study in healthy subjects and patients with DoC, except from 5 studies, due to results only confined to the group level, lack of reports on individual patient behavioral responses, or because comparison between EEG responses and behaviorally based diagnosis was not possible. Sensitivity and specificity rates in patients with DoC were calculated with the behavioral assessment as the reference standard, although a true gold standard to confirm consciousness level is nonexistent. In healthy controls, the studies displayed a relatively wide variability with regard to sensitivity rates, ranging from 71% to 100%. A high false negative rate up to 29% showed that the electrophysiological test failed to detect active mental effort in a considerable number of healthy subjects, while other studies identified all control subjects as responders. There was also a wide variety in sensitivity rates in the patient group, ranging from 0% to 100%. Here, a sensitivity rate of 0% showed that none of the included patients with discernible behavioral evidence of command-following (MCS+) were classified as responders in the active task. Of notice, a sensitivity rate of 100%
**TABLE 1**
Calculations of sensitivity and specificity in electrophysiological studies with active tasks

<table>
<thead>
<tr>
<th>Study nr</th>
<th>Author</th>
<th>Group</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>True positive</th>
<th>False positive</th>
<th>False negative</th>
<th>True negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>Bekinschtein et al62</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>78% (±27.93)</td>
<td>1 (100%)</td>
<td>2 (22%)</td>
<td>0 (0%)</td>
<td>7 (78%)</td>
</tr>
<tr>
<td>Study 2</td>
<td>Habbal et al68</td>
<td>P</td>
<td>15% (±17.45)</td>
<td>94% (±14.54)</td>
<td>3 (15%)</td>
<td>1 (6%)</td>
<td>17 (85%)</td>
<td>17 (94%)</td>
</tr>
<tr>
<td>Study 4</td>
<td>Cruse et al74</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>15 (84%)</td>
<td>3 (16%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HC</td>
<td>100% (±47.27)</td>
<td>78% (±27.93)</td>
<td>1 (100%)</td>
<td>2 (22%)</td>
<td>0 (0%)</td>
<td>7 (78%)</td>
</tr>
<tr>
<td>Study 5</td>
<td>Cruse et al63</td>
<td>P</td>
<td>13% (±19.63)</td>
<td>62% (±31.93)</td>
<td>2 (13%)</td>
<td>3 (38%)</td>
<td>13 (87%)</td>
<td>5 (62%)</td>
</tr>
<tr>
<td>Study 7</td>
<td>Gibson et al60</td>
<td>P task 1</td>
<td>100% (±47.27)</td>
<td>80% (±34.53)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>4 (80%)</td>
</tr>
<tr>
<td>Study 8</td>
<td>Horki et al61</td>
<td>P task 1</td>
<td>100% (±47.27)</td>
<td>71% (±32.32)</td>
<td>1 (100%)</td>
<td>1 (100%)</td>
<td>0 (0%)</td>
<td>1 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P task 2</td>
<td>100% (±47.27)</td>
<td>78% (±27.93)</td>
<td>1 (100%)</td>
<td>1 (100%)</td>
<td>0 (0%)</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>Study 9</td>
<td>Pan et al72</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>1 (100%)</td>
<td>0 (0%)</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>Study 10</td>
<td>Li et al73</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 11</td>
<td>Luìè et al77</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 12</td>
<td>Pokorny et al80</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 13</td>
<td>Schnakers et al80</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 14</td>
<td>Risetti et al84</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 15</td>
<td>Chennu et al78</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 16</td>
<td>Schnakers et al80</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 17</td>
<td>Hauget et al76</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HC</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 18</td>
<td>Real et al79</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 19</td>
<td>Bekinschtein et al81</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 20</td>
<td>Faugeras et al82</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 21</td>
<td>Faugeras et al85</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 22</td>
<td>King et al85</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 23</td>
<td>Sitt et al67</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Study 24</td>
<td>Rohaut et al85</td>
<td>P</td>
<td>100% (±47.27)</td>
<td>81% (±20.67)</td>
<td>1 (100%)</td>
<td>2 (20%)</td>
<td>0 (0%)</td>
<td>2 (20%)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; CS, brain-injured but conscious patients; HC, healthy control sample; NE, not estimable; P, patient sample.

*Sensitivity scores were calculated on the basis of the ability of the electrophysiological assessment to positively detect command-following in patients with DoC displaying behavioral responses to command (MCS).

*Specificity scores were calculated on the basis of the ability of the electrophysiological assessment to confirm VS and MCS, by the lack of both electrophysiological and behavioral signs of command-following. Rates of true and false positives, and true and false negatives, are reported as actual numbers of subject and percentages. Sensitivity and specificity were computed by the authors of the review using data from the published articles and calculated with 95% confidence intervals (CIs) per study, according to the efficient-score method (Newcombe99; http://vassarstats.net/index.html).
was in several studies the result of samples consisting of 1 single MCS+/ responder. Specificity rates in the patient groups also ranged from 0% to 100%, the latter again due to 1 single patient. Notably, 8 studies demonstrated specificity rates of 80% or less, illustrating that more than 20% of patients who could not demonstrate response to command behaviorally did so in the electrophysiological assessment.

**Risk of bias**

The QUADAS-2 assessment demonstrated that none of the 24 included studies had a low risk of bias or concerns regarding applicability across all domains (see Table 2). Regarding patient selection, bias concern was found because of inclusion of patients in a very early sub-acute phase after severe acquired brain injury, lack of information regarding time since injury, only 2 studies clearly stated they were based on consecutive sample, and overall lack of clarifications about inappropriate exclusion avoided, that is, outpatients or concurrent referrals. Applicability concerns regarding patient selection was due to potential sample representativity issues. Risk of bias was found with regard to the index and reference tests, as all studies, except one, lacked clear statements of whether or not interpretation of the electrophysiological assessment was blinded to the behavioral assessment, or vice versa. Concern regarding applicability of the index test was thus found in all studies except one, reflecting that there is no tradition of blinding in this field. Furthermore, the domain of flow and timing was overall of bias concern, as 9 studies were scored as unclear or with high bias risk with regard to the time interval between the behavioral and electrophysiological assessments. Accordingly, this implicated a concern for the relation between behavioral and electrophysiological assessments.

**DISCUSSION**

Over the past decade, there has been increasing scientific effort aiming at assessing covert awareness in patients with DoC applying active paradigms during electrophysiological recordings. However, the diagnostic accuracy of electrophysiological methods is still not established. Furthermore, there is no consensus regarding which experimental designs and modes of analysis would be most applicable for clinical use at a single patient level. The aim of this systematic review was to identify existing studies and to explore the clinical utility of electrophysiological methods.

**Task robustness of active paradigms in healthy control subjects**

To evaluate the diagnostic potential of electrophysiological methods to detect remnant cognitive resources in DoC, a main aim was to establish the robustness of active experimental paradigms in healthy conscious subjects. This could not be done in the 7 studies lacking a healthy control group. However, the remaining studies had sensitivity rates in healthy controls varying from 71% to 100%. Of the 3 studies showing sensitivity rates below 80%, 2 included an active condition with the instruction to listen for a change in pitch to the SON. The necessity of including personally relevant stimuli has previously been strongly emphasized, as the probability of electrophysiological responses in patients with DoC increases with salient self-referential stimuli, and the SON has proven promising in this regard. However, these results demonstrate that the cognitive content of the active condition is also of importance, as the instruction to count SON has proven to be more robust, with replicated high sensitivity rates in healthy subjects. While SON is a complex meaningful salient stimulus, other studies have applied simple harmonic tones with the instruction to count a global auditory deviant, denoted as the “local-global” paradigm, where high sensitivity rates in healthy subjects have been repeatedly demonstrated. This review illustrates that far from all electrophysiological studies have shown 100% accuracy in healthy controls. In addition, even if a method is robust in healthy subjects, it remains a question whether the sensitivity will generalize to populations with severe brain injuries.

**Diagnostic accuracy of electrophysiological measures in DoC**

A second aim of this systematic review was to establish the rates of responders in patients with DoC, as well as the number of patients with behavioral command-following that fail to show definite electrophysiological signs of active cognitive effort (false negatives). Sensitivity rates in patients with DoC varied markedly across the included studies, ranging from 0% to 100%, indicating on average that maybe as many as one third of patients who presented with unequivocal behavioral responses to command were not classified as responders based on their electrophysiological activity across studies. It is, however, challenging to disentangle whether lack of responsibility is due to patients’ characteristics or the methodological limitations of the electrophysiological technique. Patients with DoC may suffer from severe underlying perceptual and cognitive impairments, such as deficits in language, working memory, attention, memory, and executive functioning, potentially preventing them from responding in active tasks despite being conscious. Bias due to impaired hearing can be controlled for with auditory evoked potentials and by ensuring presence of the auditory N1 and/or mismatch negativity (MMN) components. Furthermore, the tasks in

www.headtraumarehab.com
<table>
<thead>
<tr>
<th>Study nr</th>
<th>Author</th>
<th>Risk of bias</th>
<th>Reference standard</th>
<th>Flow and timing</th>
<th>Applicability concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Patient selection</td>
<td>Index test</td>
<td>Reference standard</td>
<td>Flow and timing</td>
</tr>
</tbody>
</table>

Abbreviations: +, low risk; −, high risk; ?, unclear risk.
electrophysiological studies may demand higher cognitive abilities than what is required for displaying behavioral command-following, rendering CRS-R and electrophysiological results potentially incomparable. In addition, patients with DoC typically fluctuate in both their level of cognitive functioning and fatigue. Also, active tasks containing verbal instructions to elicit willfully modulated mental processes are limited by the fact that they require language comprehension, constituting a comparable challenge to that inherent in all behavioral scales. Consequently, negative EEG findings in this patient group cannot be interpreted as evidence that the patient lacks awareness any more than a negative behavioral finding does.29,60,90 Specificity rates also varied markedly, ranging from 0% to 100%, implying that some patients show signs of command-following in electrophysiological recordings, despite not doing so behaviorally (false positives). This could be related to small sample sizes or might actually be due to the fact that behavioral measures, in some cases, fail to detect the true level of functioning in the patient. Of note, the 2 largest studies containing 158 and 167 valid patient recordings demonstrated false positive rates of 17% and 33%, respectively.67,70 This highlights that, despite high rates of false negatives, covert signs of command-following have also been demonstrated. Notably, the number of patients showing electrophysiological signs of mental effort despite lack of behavioral command-following is in line with those obtained in fMRI studies using active tasks.38,91 In summary, the 2 large studies applying multivariate EEG classifier systems most likely represent the method with best balance between rates of sensitivity and specificity.

Methodological issues

The review demonstrates heterogeneity with regard to the electrophysiological techniques applied. Even though EEG-based techniques were the most frequently applied method, with only 2 EMG studies, there was a variety in the mode of analysis, such as ERP and ERD, along with diversity in EEG features included in classifier methods, hence complicating comparison of results. Furthermore, the electrophysiological methods are characterized by variations in, for example, choice of EEG equipment, protocols for EEG recordings, and methods for data analysis. Notably, there are studies where data have been reanalyzed, showing diverging results regarding rates of responders in both healthy controls and VS/MCS patients.74,92,93 In addition, studies performed in different scientific laboratories conducting similar experimental paradigms have generated conflicting results. Using a variant of the local-global experiment, a different research group found responses to global deviants in 10/24 comatose patients following cardiac arrest, but only in 6 out of 21 healthy controls,94 thus challenging previous results where the global effect has been interpreted as only being present in conscious subjects.67,70,81,82 These conflicting results have led to a debate about divergences in methodological approaches.95,96

Further methodological challenges are illustrated in the QUADAS-2 assessment, demonstrating a bias concern with regard to whether the interpretation of the electrophysiological assessment was masked to the behavioral assessment and vice versa. In clinical trials, blinding of assessors is a common requirement, while this is not a tradition within electrophysiological research, likely because the electrophysiological recording is not expected to be biased by rater expectations. However, there is a fair amount of subjective evaluations in processing and interpretation of EEG data, rendering reason for bias concern. Also, the QUADAS-2 assessment illustrated that flow of timing between the electrophysiological assessment and behavioral diagnostic measure was a concern in as many as 9 studies, highlighting that the lack of standardized and uniformly accepted methodological approaches is a real concern and a prerequisite for successful clinical translation.

Unfortunately, not all studies reported on the rate of excluded subjects, while others reported relatively high exclusion numbers due to artifacts, even in healthy subjects. In clinical practice, this means that there is a relatively high risk that a time-consuming assessment will not provide interpretable data.

In summary, there are several general and method-specific advantages and disadvantages with electrophysiological techniques applied in the included studies. High levels of artifacts remain an issue of concern in all methods described. In particular, relying on motor responses in EMG tasks is problematic due to frequent severe motor deficits such as paresis, spasticity, and contractures. When it comes to EEG frequency analysis (eg, ERD), this method alone has not per date provided strong evidence of clinical applicability but has been included as one of several components in multivariate feature analysis. Regarding ERP, the P3 is the component of choice in this particular diagnostic context, but as noted, the chance of providing evidence of consciousness is highly dependent on the experimental paradigm applied. In addition, applying multivariate EEG-classifier systems might be less influenced by subjective rater bias.

Conclusions and implications for future studies

Determining where patients lie on the spectrum of conscious awareness, and assessment of residual cognitive resources, is essential in accurate diagnosis of patients with DoC. Electrophysiological methods have the
potential to make important contributions. However, we are still far from establishing precise recommendations for standardized electrophysiological diagnostic measures in DoC. A necessary step in future research is to initiate multicenter studies, as a means to establish comparable data sets with large sample sizes across laboratories and to further establish sensitivity and specificity. Herein, ensuring systematic validation of electrophysiological paradigms in healthy controls is essential. Both false positive and false negative rates may have important implications for clinical decision-making, for example, pain management, intensity of rehabilitation, and sometimes end-of-life decisions. In summary, one needs to cautiously balance the risk of false positive versus false negative diagnostic errors in individual assessments, as it is evident that a patient with discernible signs of behavioral command-following can appear as a false negative electrophysiologically. Thus, standardized behavioral measures still constitute the standard approach to diagnostic assessment. However, in cases where severe motor deficit may mask a patient’s true level of consciousness, or where other factors contribute to diagnostic uncertainty, electrophysiological methods can complement behavioral measures with valuable additional clinical information.

Limitations

The main limitation of this systematic review is the difficulty of study comparison. Subsequently, the review focused on a qualitative synthesis of identified studies, as meta-calculation of pooled sensitivities and specificities across methods and experimental conditions was considered ineffectual. Also, as there is no established veridical benchmark of level of consciousness, precaution should be taken in interpreting results as precise estimates of sensitivity and specificity in patients with DoC.

REFERENCES


