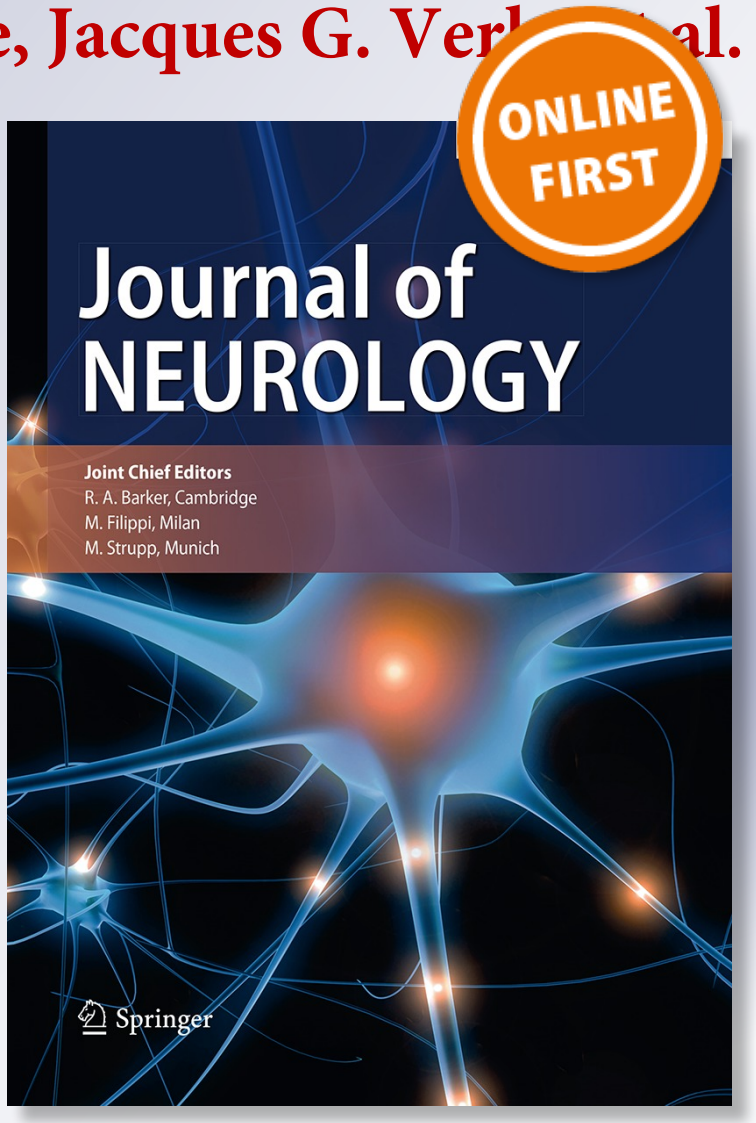
# *Objective assessment of visual pursuit in patients with disorders of consciousness: an exploratory study*

**Sarah Wannez, Thomas Hoyoux, Thomas Langohr, Olivier Bodart, Charlotte Martial, Jérôme Wertz, Camille Chatelle, Jacques G. Ver[illegible]al.**

**Journal of Neurology**

ISSN 0340-5354

J Neurol
DOI 10.1007/s00415-017-8469-0




Springer

J Neurol
DOI 10.1007/s00415-017-8469-0

ORIGINAL COMMUNICATION

# Objective assessment of visual pursuit in patients with disorders of consciousness: an exploratory study

**Sarah Wannez[1,2] · Thomas Hoyoux[3] · Thomas Langohr[3] · Olivier Bodart[1,2] · Charlotte Martial[1,2] · Jérôme Wertz[4] · Camille Chatelle[1,2,5,6] · Jacques G. Verly[3] · Steven Laureys[1,2]**

Received: 11 October 2016 / Revised: 20 March 2017 / Accepted: 21 March 2017


**Abstract** Visual pursuit is a key marker of residual consciousness in patients with disorders of consciousness (DOC). Currently, its assessment relies on subjective clinical decisions. In this study, we explore the variability of such clinical assessments, and present an easy-to-use device composed of cameras and video processing algorithms that could help the clinician to improve the detection of visual pursuit in a clinical context. Visual pursuit was assessed by an experienced research neuropsychologist on 31 patients with DOC and on 23 healthy subjects, while the device was used to simultaneously record videos of both one eye and the mirror. These videos were then scored by three researchers: the experienced research neuropsychologist who did the clinical assessment, another experienced research neuropsychologist, and a neurologist. For each video, a consensus was decided between the three persons, and used as the gold standard of the presence or absence of visual pursuit. Almost 10% of the patients were misclassified at the bedside according to their consensus. An automatic classifier analyzed eye and mirror trajectories, and was able to identify patients and healthy subjects with visual pursuit, in total agreement with the consensus on video. In conclusion, our device can be used easily in patients with DOC while respecting the current guidelines of visual pursuit assessment. Our results suggest that our material and our classification method can identify patients with visual pursuit, as well as the three researchers based on video recordings can.

**Keywords** Disorders of consciousness · Minimally conscious state · Behavioral assessment · Visual pursuit · Automatic detection

## Introduction

Bedside assessment of consciousness using behavioral scales is currently the gold standard to assess patients with disorders of consciousness (DOC), which includes unresponsive wakefulness syndrome/vegetative state (UWS/VS [1]) and minimally conscious state (MCS [2]). Patients in UWS/VS are awake (i.e., with their eyes open), but do not demonstrate any voluntary sign of consciousness of themselves or of their environment, while patients in MCS are characterized by the presence of reproducible and purposeful behaviors. A subcategorization has been proposed among MCS: MCS *plus* and MCS *minus* [3]. Patients in MCS *plus* are defined by the presence of response to command, whereas patients in MCS *minus* are defined by the fact that they only show lower-level non-reflex behaviors, such as localization to pain, object localization, or visual pursuit. When patients are able to functionally

Sarah Wannez and Thomas Hoyoux contributed equally.

✉ Sarah Wannez
coma@chu.ulg.ac.be

1 Coma Science Group, GIGA-Consciousness, University of Liège, Liège, Belgium

2 Neurology Department, University Hospital of Liège, Liège, Belgium

3 Laboratory for Signal and Image Exploitation, Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium

4 Phasya S.A., Liège, Belgium

5 Laboratory for NeuroImaging of Coma and Consciousness, Massachusetts General Hospital, Boston, MA, USA

6 Acute Neurorehabilitation Unit, Department of Clinical Neurosciences, University Hospital of Lausanne, Lausanne, Switzerland

Published online: 31 March 2017

communicate and/or to functionally use objects, they have emerged from MCS (EMCS [2]). For ethical and medical reasons, it is of paramount importance to correctly identify signs of consciousness. For example, it has been shown that patients in MCS are able to process auditory information, and to suffer from pain, unlike patients in UWS/VS [4]. Different scales have been designed to assess post-comatose patients, but the Coma Recovery Scale-Revised (CRS-R [5]) was amongst the few showing strong evidence of reliability and validity for assessment of DOC, based on a recent systematic review completed by the American Congress of Rehabilitation Medicine [6]. As compared to a diagnosis achieved solely by clinical consensus of the medical staff, this scale permits one to avoid 41% of misdiagnosis (i.e., erroneously considering patients in MCS as being in UWS/VS) [7].

Visual pursuit is one of the first signs appearing during recovery of consciousness [2]. According to different studies assessing the evolution from UWS/VS to MCS, around 45% of patients are detected thanks to visual pursuit [8, 9]. In the global population of MCS, however, the prevalence of the visual pursuit is around 70% [10, 11]. This response is, therefore, key for the clinical assessment of patients with DOC. The CRS-R includes the assessment of visual pursuit by means of a mirror (as opposed to a person or an object, which were revealed to be less efficient [10, 11]). Studies on healthy subjects also showed that using a mirror was more efficient than using an object, because it elicits a smoother visual pursuit. The use of a mirror decreases the probability of erroneously considering that the patient did not follow the stimulus [12]. However, the assessment of visual response is highly relying on clinical observation, hence more likely to be biased by human subjectivity and errors. More and more, medical sciences tend to develop objective tools for supplementing the clinical assessment to decrease the caveats of current bedside assessment, e.g., using electrophysiology or neuroimaging for improving the detection of command-following [13–15] or cerebral activity compatible with a residual consciousness [16–18]. For the assessment of visual response, the use of an eye-tracking system could be very useful to offer a new objective way to detect visual pursuit or fixation. However, such devices typically require a calibration phase, where the subject has to fixate different points on a screen to let the computer know, where the eyes are looking at. For patients with DOC, who are by definition non-collaborative and non-communicative, this is not feasible. In a first attempt to provide objective assessment of visual pursuit, some previously published studies used a computerized infrared eye-tracking system [19, 20]. While the calibration phase was eliminated thanks to an a posteriori calibration, the setting still suffered from some limitations. For instance, patients were seated in front on a screen displaying moving stimuli (e.g., a red circle and a color head of a parrot). As demonstrated in the literature, this might be less attractive and efficient than the mirror [10, 11]. Moreover, the sitting position may not only require supplementary time for clinicians, but also not always be feasible with patients, who could suffer from some spasticity, or on the contrary from a lack of tonus. Globally, this eye-tracker device is not really suitable for patients, neither adapted to visual pursuit assessment if one wants to follow current guidelines.

Recently, we developed an easy-to-use device, which proved to be efficient on healthy subjects and on some patients with DOC [21]. Briefly, eye and mirror movements were video-recorded, and a correlation between their positions was calculated throughout the recording. The final value provided by the correlation was called the confidence score (C-score). This device enabled one to assess visual pursuit with a mirror, as recommended by the CRS-R, and did not interfere with the usual, recommended clinical assessment. Moreover, the patient did not need to adopt any specific position.

The current study aims (1) to characterize the reliability of the clinical assessment of visual pursuit as compared to a consensus by researchers on video, (2) to assess the reliability of the C-score to detect visual pursuit in patients with DOC, and (3) to propose a new automatic procedure to identify the presence of visual pursuit in case the accuracy of the C-score is not sufficient, as we know that the population of patients is much more heterogeneous than that of healthy subjects.

## Materials and methods

### Material

The material consisted of a head-mounted device worn by the patient. This device was composed of a fish eye scene camera recording the patient's field of view, and of an infrared (IR) eye camera, recording the movements of one eye through an IR mirror. The IR mirror, transparent to light in the visible part of the spectrum, did not hamper the vision of the patient. The scene camera recorded images of 752 × 480 pixels, at 30 frames per second. The eye camera recorded images of 240 × 160 pixels, at 180 frames per second, and was for some subjects set to record at a lower frame rate of 120 frames per second, but with a higher resolution of 320 × 240 pixels, to get more precise images. Post-processing took into account the differences in image quality, as all of the images were standardized before analysis. Two setups were used. In the first setting, we used a cap-like device, and, in the second, a glasses-like device because of some difficulties encountered with the cap.

Indeed, it was not easy to get the cap fitted on patients with large heads, and when these patients were moving their heads too much, the whole device was displaced. Moreover, the mirror in the cap was mobile, adjusted by the research neuropsychologist to record the eye movements. The mirror was fixed on the glasses, increasing the standardization of the eye position in the video images. The glasses were provided by Phasya S.A. (Angleur, Belgium), in the form of a modified prototype of the Drowsimeter R100 (see Fig. 1).

### Clinical assessments

Patients with DOC were recruited during a 1-week hospitalization in the University Hospital of Liège. They were sent by their treating physician and/or the family, and several clinical examinations were performed. Written informed consent was obtained from the patient surrogate-decision makers in accordance with the research protocol approved by the University Hospital of Liège. The inclusion criteria were: (1) to be at least 18 years and (2) the presence of a severe brain injury leading to a prolonged DOC, as diagnosed by the CRS-R. Exclusion criteria were (1) time since brain injury shorter than three months and (2) presence of premorbid neurological or psychiatric disorders. Patients were not included on the basis of the integrity of their visual functions as, most of the time, clinicians do not know a priori if a patient is able to see before testing visual pursuit. A cohort of healthy subjects being at least 18 years and without neurological or psychiatric disorders was enrolled. The healthy subjects provided written informed consent. These healthy subjects were already published in the C-score article [21].

**Fig. 1** Adapted device recording eye and mirror images and worn by the patient

Patients were tested in a comfortable position (sitting in a chair or installed in a bed), and wore the device as described above. Visual pursuit was tested according to the CRS-R guidelines, namely, with a mirror, held around 10–15 cm from the patient's face, and moving from the front of the patient to right, left, up, and down directions (in different orders) to an angle of 45°. The procedure was repeated once as recommended in the CRS-R administration guidelines, leading to two trials in each direction, and a total of eight trials for the whole assessment. Global visual pursuit was considered to be present when a smooth visual pursuit was observed on two occasions, in any direction, out of the eight trials. After the visual pursuit assessment, the other items of the CRS-R were tested to define the clinical diagnosis for each patient. All the clinical assessments were administered by an experienced research neuropsychologist (SW).

For healthy subjects, assessment was done in a sitting position. They were asked first to follow the mirror, and then to not follow it (i.e., to keep a fixed gaze and/or to perform random eye movements).

After data acquisition, an anonymous data set of videos of all visual pursuit sessions (patients and healthy subjects) was created, where each video was composed to display the sequence of eye movements as they were recorded by the eye camera, side-by-side with a synthetic depiction of the corresponding mirror movements seen by the subject. This synthetic depiction was designed to ensure that the unnecessary details captured by the scene camera (e.g., the face in the mirror reflection, or the surroundings) did not reveal the subject group or identity. An example of such images is reported in Fig. 2. The videos of this anonymous data set were separately presented to three researchers with great experience in clinical assessment of patients with DOC: the experienced research neuropsychologist who did the clinical assessment (SW), another experienced research neuropsychologist (CM), and a neurologist (OB). They scored the presence or absence of visual pursuit for each of the eight movements (or trials). After the scoring, trials without unanimity were identified, and discussed between researchers until a final decision was obtained (consensus by researchers). For each subject, the clinical final score represented the proportion of successful trials. For example, if the subject followed five times out of the eight trials, the score was 0.62 (=5/8). The presence of global visual pursuit was determined according to CRS-R criteria (two or more successful trials out of the eight performed trials $\geq$0.25).

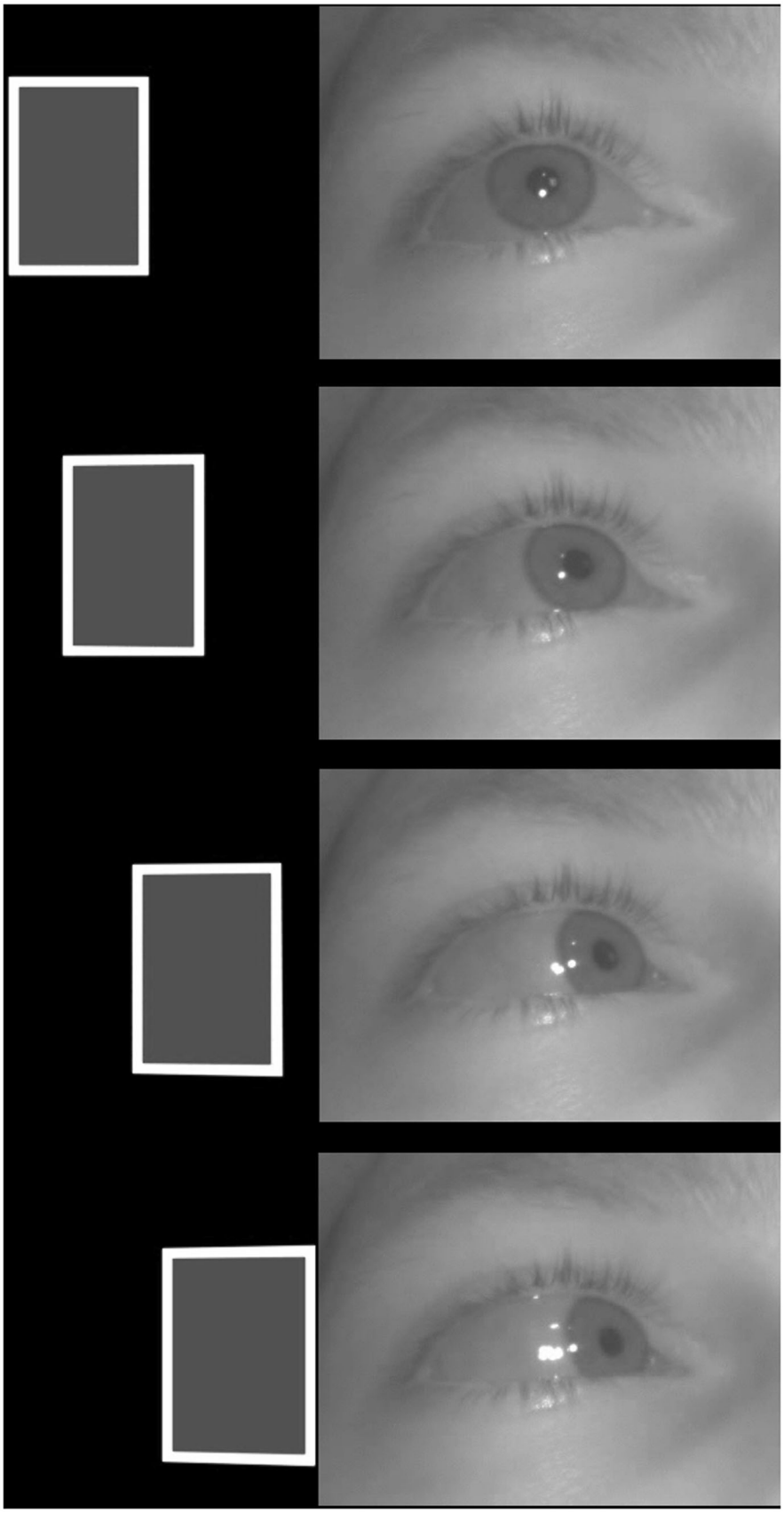


**Fig. 2** Freeze-frame of videos assessed by the three researchers. On the *left side,* synthetic mirror image. On the *right side,* eye image

### Classification

In a first step, the C-score was computed for each subject as previously published [21]. In summary, the 2D pupil trajectory (*X* and *Y*-coordinate sequence) and 3D mirror trajectory (*X*-, *Y*-, and *Z*-coordinate sequence) were extracted from the eye and scene videos, respectively. In both cases, we used Kalman filtering to correct tracking inconsistencies. The eye and scene camera frame rates being different, excess pupil trajectory points were discarded: only one eye image out of six was kept. The remaining pupil trajectory points were matched to the corresponding mirror trajectory points according to their timestamps (the *Z*-coordinate of the mirror trajectory points was discarded). Tags were manually associated to some trajectory points to mark the beginning and end of the mirror movements (i.e., mirror going to the right, coming back to the center, going to the left, and so on). We only kept movements starting in front of the patients, until the farthest point in the movement direction (left, right, up, or down), as described in the CRS-R administration guidelines. The Pearson correlation was then calculated between the remaining pupil and mirror trajectory points, separately for the *X*- and *Y*-axes. Negative correlations were set to zero. Finally, the C-score was obtained by calculating the average between those two correlations values. Arbitrarily, we set the presence of a visual pursuit at a value of C-score equal or superior to 0.25, representing the proportion of visual pursuit needed to reach the CRS-R criteria.

In a second step, as we found that the C-score—which worked very well with healthy subjects and with some patients—was insufficient to deal reliably with more difficult patient cases, we used another method for improving the reliability of the system. We used a supervised machine learning approach known as artificial neural networks (ANNs [22]). We used an ANN classifier to classify the data (400 data points per movement followed or not followed, as defined by the consensus by researchers), and used a leave-one-subject-out method to test the ability of the model to classify new inputs. Each trial was categorized by the classifier as followed or not (individual indices), and a global score was then calculated for each subject (e.g., 5 pursuits out of 8 trials = 0.62). Visual pursuit was considered present from 0.25, as in the clinical assessment. This new index of measurement of visual pursuit was called the modified score (M-score).

### Hypotheses and statistical analyses

We expected a high congruence between the clinical assessment of visual pursuit as assessed by the research neuropsychologist (SW) and the consensus by researchers based on video scoring. We tested this hypothesis using Cohen's kappa, at trial level (i.e., the presence or absence of a following movement for each trial separately) and global level (i.e., the presence or absence of global visual pursuit based on CRS-R criteria of at least two successful trials). Interpretation of kappa statistic was done according to recommendations [23]: agreement was classified as poor (<0), slight (0–0.2), fair (0.21–0.4), moderate (0.41–0.6), substantial (0.61–0.8), or almost perfect (0.81–1).

We also expected a high congruence between the C-score and the consensus by researchers (based on the Cohen's kappa, at global level). We predicted a significant correlation of the C-score with the proportion of visual pursuits identified by the consensus by researchers (Spearman correlation). We calculated the sensitivity and specificity of the C-score, considering the consensus by researchers as the gold standard.

Finally, we expected a high congruence between the M-score and the consensus by researchers (Cohen's kappa, at trial level and global level). We also expected a positive correlation between the proportion of visual pursuits identified by the consensus by researchers and the M-score (Spearman correlation). We calculated the sensitivity and specificity of the M-score at the global level, as we did for the C-score.

The analyses were performed using IBM SPSS Statistics version 20 (IBM Corp., Armonk, N.Y., USA), separately for patients and healthy subjects. Results were considered significant at $p < 0.05$.

## Results

Thirty-one patients with chronic DOC were included in the study (12 in UWS/VS, 11 in MCS-, 3 in MCS +, 5 in EMCS; 13 traumatic, 11 women, mean age = 40.23 ± 13.19 years, mean time since onset = 4.55 ± 4.84 years, see Table 1 for demographic and clinical data). The results for 23 healthy subjects (mean age = 28 ± 7 years, 10 women) were also reported. It is important to remember that they were recorded with and without visual pursuit, increasing the number of videos (50 videos: 23 following, 17 fixed gaze, 10 random gaze). Regarding the consensus by researchers, five trials out of the 400 performed had to be discussed to reach a consensus in the healthy subjects group (1.25%). In the patients group, 28 trials out of 247 had to be discussed (11%). For the other trials, the three researchers had indicated the same appreciation without concerting. Patients and healthy subjects were not matched for age or gender. Eight trials were performed for each subject, except for one patient who only had seven trials due to an error of the research neuropsychologist (SW) during clinical assessment.

**Table 1** Demographic and clinical data of patients

| | Gdr | Age | Etio. | TSO | Diag. | CRS-R | Aud | Vis | Mot | Oro | Com | Ar | Clinical visual pursuit | Offline visual pursuit | Consensus visual pursuit | C-score visual pursuit | M-score visual pursuit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | F | 28 | NTBI | 4 months | UWS | 6 | 1 | 0 | 2 | 1 | 0 | 2 | No | No | No | No | No |
| 2 | M | 41 | TBI | 1 year | UWS | 4 | 0 | 0 | 1 | 1 | 0 | 2 | No | No | No | No | No |
| 3 | M | 73 | NTBI | 9 years | UWS | 6 | 1 | 0 | 2 | 1 | 0 | 2 | No | No | No | Yes | No |
| 4 | M | 58 | NTBI | 2 years | UWS | 3 | 0 | 0 | 1 | 1 | 0 | 1 | No | No | No | No | No |
| 5 | F | 50 | NTBI | 6 months | UWS | 6 | 1 | 0 | 2 | 1 | 0 | 2 | No | No | No | No | No |
| 6 | M | 37 | TBI | 11 months | UWS | 5 | 0 | 0 | 2 | 1 | 0 | 2 | No | No | No | Yes | No |
| 7 | F | 41 | NTBI | 2 years | UWS | 4 | 1 | 0 | 1 | 1 | 0 | 1 | No | No | No | Yes | No |
| 8 | F | 36 | NTBI | 1 year | UWS | 4 | 0 | 0 | 1 | 1 | 0 | 2 | No | No | No | No | No |
| 9 | M | 33 | NTBI | 14.5 years | UWS | 7 | 1 | 1 | 2 | 1 | 0 | 2 | No | No | No | No | No |
| 10 | M | 22 | NTBI | 1 year | UWS | 6 | 0 | 0 | 2 | 2 | 0 | 2 | No | No | No | Yes | No |
| 11 | M | 23 | NTBI | 1 year | UWS | 4 | 0 | 0 | 1 | 2 | 0 | 1 | No | No | No | Yes | No |
| 12 | F | 42 | NTBI | 9 months | UWS | 4 | 1 | 0 | 1 | 1 | 0 | 1 | No | Yes | No | No | No |
| 13 | F | 40 | TBI | 3.5 years | MCS- | 7 | 1 | 3 | 1 | 1 | 0 | 1 | Yes | Yes | Yes | Yes | Yes |
| 14 | M | 45 | TBI | 13 years | MCS- | 8 | 1 | 3 | 2 | 1 | 0 | 1 | Yes | Yes | Yes | Yes | Yes |
| 15 | M | 54 | NTBI | 6 months | MCS- | 9 | 1 | 2 | 5 | 0 | 0 | 1 | No | No | No | No | No |
| 16 | F | 25 | NTBI | 11 months | MCS- | 9 | 1 | 3 | 1 | 2 | 0 | 2 | Yes | Yes | No | Yes | No |
| 17 | M | 26 | TBI | 12 years | MCS- | 8 | 0 | 3 | 2 | 1 | 0 | 2 | Yes | No | No | Yes | No |
| 18 | M | 25 | TBI | 1.5 year | MCS- | 8 | 1 | 3 | 1 | 1 | 0 | 2 | Yes | Yes | Yes | Yes | Yes |
| 19 | M | 34 | NTBI | 12 years | MCS- | 8 | 1 | 3 | 1 | 1 | 0 | 2 | Yes | Yes | Yes | Yes | Yes |
| 20 | M | 34 | NTBI | 12 years | MCS- | 10 | 1 | 1 | 5 | 1 | 0 | 2 | No | No | No | No | No |
| 21 | F | 62 | NTBI | 2 years | MCS- | 8 | 2 | 3 | 1 | 1 | 0 | 1 | Yes | Yes | Yes | Yes | Yes |
| 22 | F | 42 | NTBI | 7 months | MCS- | 7 | 1 | 3 | 1 | 1 | 0 | 1 | Yes | Yes | Yes | Yes | Yes |
| 23 | M | 55 | NTBI | 7 years | MCS- | 10 | 0 | 3 | 5 | 1 | 0 | 1 | Yes | Yes | No | Yes | No |
| 24 | F | 31 | TBI | 13 years | MCS+ | 9 | 3 | 3 | 1 | 1 | 0 | 1 | Yes | Yes | Yes | Yes | Yes |
| 25 | M | 60 | NTBI | 2 years | MCS+ | 6 | 3 | 1 | 1 | 1 | 0 | 0 | No | No | No | No | No |
| 26 | M | 30 | TBI | 4 years | MCS+ | 8 | 3 | 0 | 1 | 2 | 0 | 2 | No | No | No | No | No |
| 27 | M | 24 | TBI | 5 months | EMCS | 23 | 4 | 5 | 6 | 3 | 2 | 3 | Yes | Yes | Yes | Yes | Yes |
| 28 | M | 49 | TBI | 1 year | EMCS | 13 | 4 | 5 | 6 | 1 | 2 | 1 | Yes | Yes | Yes | Yes | Yes |
| 29 | M | 33 | TBI | 11 years | EMCS | 18 | 3 | 4 | 6 | 3 | 0 | 2 | Yes | Yes | Yes | Yes | Yes |
| 30 | F | 37 | TBI | 6 years | EMCS | 23 | 4 | 5 | 6 | 3 | 2 | 3 | Yes | Yes | Yes | Yes | Yes |
| 31 | M | 58 | TBI | 4 years | EMCS | 12 | 3 | 3 | 0 | 3 | 2 | 1 | Yes | Yes | Yes | Yes | Yes |

*Gdr* gender, *Etio* etiology, *TSO* time since onset, *Diag* diagnosis, *CRS-R* CRS-R total score, *Aud* CRS-R auditory subscale, *Vis* CRS-R visual subscale, *Mot* CRS-R motor subscale, *Oro* CRS-R oromotor/verbal subscale, *Com* CRS-R communication subscale, *Ar* CRS-R arousal subscale

### Healthy subjects

At the trial level, an almost perfect agreement was observed between bedside assessment and the research neuropsychologist who did the clinical assessment (SW) video scoring ($\kappa = 0.98$, based on the 400 trials done with subjects). A disagreement was observed in four trials (1%). An almost perfect agreement was also observed between bedside assessment and consensus by researchers ($\kappa = 0.98$). A disagreement was observed in four trials (1%).

At the global level, an almost perfect agreement was observed between bedside assessment and the research neuropsychologist who did the clinical assessment (SW) scoring on video ($\kappa = 0.96$, based on 50 subjects). A disagreement was observed in one subject (2%). An almost perfect agreement was observed between bedside assessment and consensus by researchers ($\kappa = 0.96$). A disagreement was observed in one subject (2%).

An almost perfect agreement was observed between consensus by researchers and C-score decision ($\kappa = 0.92$, based on 50 subjects). A disagreement was observed in two subjects (4%). The proportion of succeeded trials based on the consensus by researchers and the C-score correlated significantly (Spearman $r = 0.891$, $p < 0.001$; see Fig. 3a). The sensitivity of the C-score was 96.1%, and the specificity 95.8%.

Regarding the M-score, a perfect agreement was observed with consensus by researchers at the trial level ($\kappa = 1$, 0% disagreement). A perfect agreement was observed with consensus by researchers at global level ($\kappa = 1$, 0% disagreement). The M-score significantly and perfectly correlated with the proportions of succeeded trials based on the consensus by researchers (Spearman $r = 1$, $p > 0.001$; see Fig. 3b). The sensitivity and the specificity of the M-score both reached 100%.

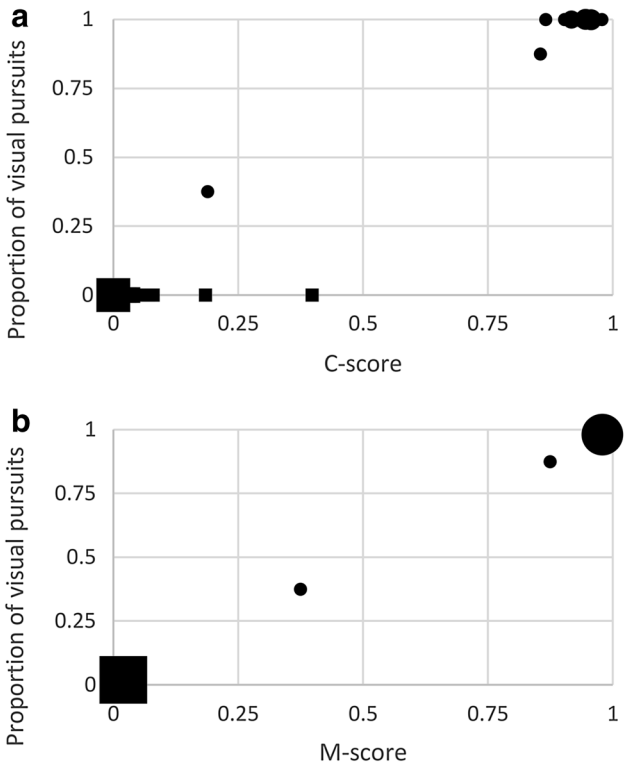



**Fig. 3** Correlation between classification measures and consensus by researchers in healthy subjects. **a** Correlation between C-score and consensus by researchers. **b** Correlation between M-score and consensus by researchers. *Dots* represent the healthy subjects with a visual pursuit according to the consensus by researchers. *Squares* represent the healthy subjects without visual pursuit. The differentiation in *size of the squares and dots* represents the amount of subjects with similar results. There is no complete overlap between the squares and dots

### Patients

At the trial level, an almost perfect agreement was observed between bedside assessment and video scoring ($\kappa = 0.864$, based on the 247 trials done with patients). A disagreement was observed in 14 trials (5.7%). An almost perfect agreement was also observed between bedside assessment and consensus by researchers ($\kappa = 0.859$). A disagreement was observed in 14 trials (5.7%). The kappa's relative to the different diagnostic subgroups were also computed (see Table 2), except in the UWS/VS subgroup, as those patients do not show a visual pursuit at bedside, by definition. Thus, no reliable statistical indices were computable. Moreover, no reliable visual pursuit was detected on video (consensus by researchers).

At the global level, an almost perfect agreement was observed between bedside assessment and video scoring ($\kappa = 0.871$, based on 31 patients). A disagreement was observed in two patients (6.5%). A substantial agreement was observed between bedside assessment and consensus by researchers ($\kappa = 0.805$). A disagreement was observed in three patients (9.7%).

A moderate agreement was observed between consensus by researchers and C-score decision ($\kappa = 0.516$, based on 31 patients). A disagreement was observed in eight patients (25.8%). The proportion of successful trials based on the consensus by researchers and the C-score correlated significantly (Spearman $r = 0.83$, $p < 0.001$; see Fig. 4a). The sensitivity of the C-score reached 100%, while the specificity was limited to 60%.

Regarding the M-score, an almost perfect agreement was observed with consensus by researchers at the trial level ($\kappa = 0.907$). A disagreement was observed in nine trials (3.6% disagreement). A perfect agreement was observed with consensus by researchers at patient level ($\kappa = 1$, 0% disagreement). The M-score significantly correlated with the proportions of successful trials based on the consensus by researchers (Spearman $r = 0.913$,

**Table 2** Kappa, sensitivity, and specificity linked to each comparison between two different measures (including at the bedside, on video, consensus, and M-score)

| Comparison (Kappa) | MCS− | MCS+ | EMCS |
|---|---|---|---|
| Bedside vs. video | 0.815 | 0.864 | 0.944 |
| Bedside vs. consensus | 0.789 | 0.864 | 1 |
| Consensus vs. M-score | 0.902 | 1 | 0.894 |
| **Sensitivity and specificity** | **MCS− (%)** | **MCS+ (%)** | **EMCS (%)** |
| Bedside vs. consensus | | | |
| Sensitivity | 97 | 80 | 100 |
| Specificity | 85 | 100 | 100 |
| Consensus vs. M-score | | | |
| Sensitivity | 94 | 100 | 92 |
| Specificity | 96 | 100 | 100 |

In the absence of a real gold standard, we use the consensus by researchers as our reference, and compared the different scoring systems to this consensus. The sensitivity and the specificity for the bedside and video scoring comparison are not reported as no reference is available (no consensus involved)

$p < 0.001$; see Fig. 4b). At the global level, the sensitivity and the specificity of the M-score both reached 100%. These results (i.e., kappa, sensitivity, and specificity) for each diagnosis subgroups were also computed (see Table 2). Again, the UWS/VS subgroup was not included as the absence of visual pursuit is a necessary criterion for this state. Thus, no reliable statistical indices were computable. Moreover, no reliable visual pursuit was detected by the classifier.

## Discussion

The aim of the study was to assess and improve an objective tool to assess visual pursuit in patients with DOC. Visual pursuit is one of the first signs of consciousness observed over the course of recovery [2], therefore, differentiating UWS/VS (i.e., unconscious) from MCS (i.e., conscious) patients. Although the CRS-R provides clear and precise guidelines for the testing and scoring of visual responses, we showed here that it might lead to potential errors when assessed by a single person. More precisely, around 6% of the trials were scored differently at bedside and when videos were scored afterwards, be it by the experienced research neuropsychologist who did the clinical assessment (SW) or by the three researchers (OB, CM, and SW), leading to up to 10% errors in detecting visual pursuit in such patients. This is of great importance as such errors will likely lead to patients being erroneously diagnosed in an UWS/VS, therefore influencing the course of

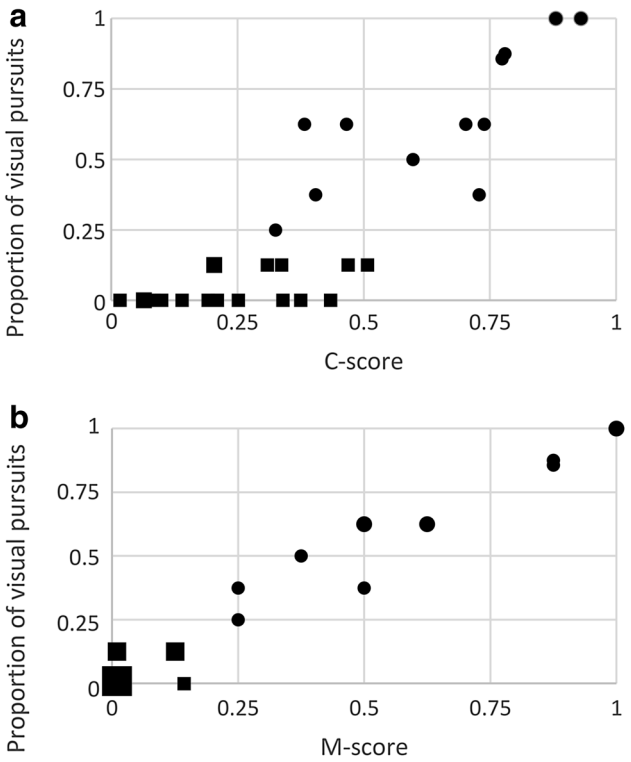



**Fig. 4** Correlation between classification measures and consensus by researchers in patients. **a** Correlation between C-score and consensus by researchers. **b** Correlation between M-score and consensus by researchers. *Dots* represent the patients with a visual pursuit according to the consensus by researchers. *Squares* represent the patients without visual pursuit. The differentiation in *size of the squares and dots* represents the amount of subjects with similar results. There is no complete overlap between the squares and dots

treatment, including rehabilitation plan, pain management, or end-of-life decisions [24–26]. If it is possible to reduce the error rate when several examiners assess the patient (even on video), this is not easy to implement in a clinical setting, because it is time-consuming and it requires a lot of human resources. In this context, an objective, plug-and-play measure might be of particular interest.

We thus conducted an exploratory study to propose an easy-to-use device to assist clinicians in assessing visual pursuit in patients with DOC. The previously proposed score, the C-score, misclassified 26% of the patients, and its main limitation was the lack of specificity, especially in the patient group. In healthy subjects, while distributions were clearly separated, a higher score was observed in one subject even though this subject was not following the stimulus. This suggests that the C-score might be affected by unrelated or tiny eye movements, leading to an increase in the C-score value, despite the absence of visual pursuit. We thus proposed a novel approach based on ANN, classifying each trial as successful or not. The global output, named M-score, represented the proportion of successful

trials for each subject. It correctly classified all the patients and healthy subjects, by producing the same answer as the consensus by researchers, suggesting that it can actually provide a measure equivalent to three researchers examining a patient at the subject-level. When testing the M-score at the trial level (i.e., comparing the classifier individual indices for each trial to the consensus by researchers) in the different diagnostic subgroups, results were very encouraging in all the subgroups, with kappa ranging from 0.815 to 1, while the sensitivity and the specificity ranged from 80 to 100%. The subgroup of patients in UWS/VS was not included, as they did not show visual pursuit at the bedside, by definition. It is important to note that the video scoring and the classifier did not identify any sustained visual pursuit that would have changed the diagnosis to MCS.

Our findings suggest that such a complementary tool may be of great interest for clinicians, as it is adapted for bedside use, without any constraint for the patient, and it allows to assess visual pursuit with a mirror, as recommended by the CRS-R, currently the gold standard for assessing patients with DOC [5, 11]. The fact that we could easily acquire data in our cohort suggests that it would be a good tool to supplement classical bedside assessment in this population.

One could be surprised by the findings reported for healthy subjects. Although they were asked *not* to follow—vs. follow the mirror—some did not comply with the instructions correctly, or did random movements that could be associated with to the mirror positions. These particular trials were detected during the video scoring, and the M-score agreed with the consensus by researchers.

If the method proposed is a very good first step to more objectively assess visual pursuit, further studies are clearly needed to help better understand visual pursuit and define a clear gold standard for clinical assessment. The consensus of three experienced researchers was here considered as the measure that should better approach the truth, even though it is still based on subjective observations. In addition, the videos might also reduce the information available at bedside as the assessor might perceive external information, such as a head movements or facial expressions influencing his/her judgement.

Finally, these results need to be interpreted with caution as several limitations can be raised. First, our small cohort does not allow any generalization to the DOC population. Second, we modified the image quality and the recording device over the course of the study to move from a cap to glasses and, although we limited the impact of this change as much as possible during data processing, it could have influenced the data. However, the use of the glasses was not only easier for patients, but it was also more standardized. The IR mirror was fixed and not mobile, so the angle of view and the distance from the eye were almost the same in each patient, up to a mild variation due to morphological differences. One limitation of using the glasses is that this only enabled to record the right eye, while, with the cap, the clinician could choose which eye to record (but sill, one eye at a time). Third, the device could be further improved to (1) enable recording images of both eyes, to get as much information as possible, as well as decreasing the risk of recording an eye that suffers from some mobility problems and (2) investigate the use of eyelid movements for detecting visual pursuit (particularly for the vertical eye movements). Another improvement of the glasses might be a reduction of their size and salience. Indeed, a possible caveat of our study might be the presence of a visible device, modifying one's appearance, and negatively impacting the results. Indeed, wearing a cap or glasses modifies one's reflection, while the efficiency of the mirror is presumed to be due to the auto-referential aspect, i.e., one's own face in the mirror [11]. We tried to reduce this effect by using a device as discreet as possible, but still it modifies one's appearance. One should note that one patient showed a visual pursuit only when the cap was removed from his head, suggesting that, in a few cases, patients might be affected by the mirror reflection. An alternative explanation for this result might be the vigilance fluctuation or the tiredness.

In addition, the difference of mean age in patients and healthy subjects might represent a bias in our study. However, to our knowledge, the visual abilities to track moving objects should not be different in this age range (28–40 years). Finally, we did not test for the integrity of visual abilities of the patients before testing visual pursuit as, in most cases; clinicians do not have this information before testing visual pursuit. However, future studies would benefit from a better characterization of the visual pursuit through objective assessments (e.g., visual evoked potentials) when feasible. This would allow us to improve our understanding of visual impairment in DOC as well as our tool for assessing visual pursuit.

In conclusion, we here provide evidence that an objective tool for assessing visual pursuit can be reliably used in patients with DOC, without hampering the usual and recommended way to test it through the classical behavioral bedside assessment. It enables one to assess the patients in the usual clinical setting, without any calibration needed, while still using the stimulus that is recommended (i.e., the mirror). Future studies are still needed to assess the reliability of our tool on a larger cohort of patients with various etiologies and various degrees of visual impairments.

**Acknowledgements** The University and University Hospital of Liège, the French Speaking Community Concerted Research Action (ARC 12/17/01), the Belgian National Funds for Scientific Research

(FRS-FNRS), Human Brain Project (EU-H2020-fetflagship-hbp-sga1-ga720270), and Luminous project (EU-H2020-fetopen-ga686764), the Wallonie-Bruxelles International, the James McDonnell Foundation, Mind Science Foundation, IAP research network P7/06 of the Belgian Government (Belgian Science Policy), the European Commission, the Public Utility Foundation 'Université Européenne du Travail', "Fondazione Europea di Ricerca Biomedica", the Bial Foundation. OB and CM are research fellows, CC is post-doctoral fellow, and SL is research director at FRS-FNRS.

**Compliance with ethical standards**

**Conflicts of interest** Phasya S.A. provided the device without any financial compensation.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study, or by patient surrogate-decision makers.

## References

1. Laureys S, Celesia G, Cohadon F et al (2010) Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. BMC Med 8:68
2. Giacino JT, Ashwal S, Childs N et al (2002) The minimally conscious state: definition and diagnostic criteria. Neurology 58(3):349–353
3. Bruno M-A, Majerus S, Boly M et al (2012) Functional neuroanatomy underlying the clinical subcategorization of minimally conscious state patients. J Neurol 259:1087–1098
4. Boly M, Faymonville M-E, Peigneux P et al (2005) Cerebral processing of auditory and noxious stimuli in severely brain injured patients: differences between VS and MCS. Neuropsychol Rehabil 15(3–4):283–289
5. Giacino JT, Kalmar K, Whyte J (2004) The JFK coma recovery scale-revised: measurement characteristics and diagnostic utility. Arch Phys Med Rehabil 85:2020–2029
6. Seel RT, Sherer M, Whyte J et al (2010) Assessment scales for disorders of consciousness: evidence-based recommendations for clinical practice and research. Arch Phys Med Rehabil 91(12):1795–1813
7. Schnakers C, Vanhaudenhuyse A, Giacino J et al (2009) Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. BMC Neurol 9(1):35
8. Bagnato S, Boccagni C, Sant'Angelo A, Fingelkurts AAA, Fingelkurts AAA, Galardi G (2016) Longitudinal assessment of clinical signs of recovery in patients with unresponsive wakefulness syndrome after traumatic or nontraumatic brain injury. J Neurotrauma 33:1–5
9. Estraneo A, Moretta P, Cardinale V et al (2014) A multicentre study of intentional behavioural responses measured using the coma recovery scale-revised in patients with minimally conscious state. Clin Rehabil 29(8):803–808. doi:10.1177/0269215514556002
10. Thonnard M, Wannez S, Keen S et al (2014) Detection of visual pursuit in patients in minimally conscious state: a matter of stimuli and visual plane? Brain Inj 28(9):1164–1170
11. Vanhaudenhuyse A, Schnakers C, Brédart S, Laureys S (2008) Assessment of visual pursuit in post-comatose states: use a mirror. J Neurol Neurosurg Psychiatry 79(2):223
12. Cruse D, Fattizzo M, Owen AM, Fernández-Espejo D (2017) Why use a mirror to assess visual pursuit in prolonged disorders of consciousness? Evidence from healthy control participants. BMC Neurol 17(1):14. doi:10.1186/s12883-017-0798-1
13. Monti MM, Vanhaudenhuyse A, Coleman MR et al (2010) Willful modulation of brain activity in disorders of consciousness. N Engl J Med 362(7):579–589
14. Cruse D, Chennu S, Chatelle C et al (2012) Relationship between etiology and covert cognition in the minimally conscious state. Neurology 78:816–822. doi:10.1212/WNL.0b013e318249f6f0
15. Bekinschtein T, Coleman MR, Niklison J, Pickard JD, Manes FF (2008) Can electromyography objectively detect voluntary movement in disorders of consciousness? J Neurol Neurosurg Psychiatry 79:826–828
16. Stender J, Gosseries O, Bruno M-A et al (2014) Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. Lancet 384(9942):514–522
17. Stender J, Kupers R, Rodell A et al (2014) Quantitative rates of brain glucose metabolism distinguish minimally conscious from vegetative state patients. J Cereb Blood Flow Metab 35:1–8. doi:10.1038/jcbfm.2014.169
18. Gosseries O, Thibaut A, Boly M, Rosanova M, Massimini M, Laureys S (2014) Assessing consciousness in coma and related states using transcranial magnetic stimulation combined with electroencephalography. Ann Fr Anesth Reanim. doi:10.1016/j.annfar.2013.11.002
19. Trojano L, Moretta P, Loreto V, Cozzolino A, Santoro L, Estraneo A (2012) Quantitative assessment of visual behavior in disorders of consciousness. J Neurol 259(9):1888–1895
20. Trojano L, Moretta P, Loreto V, Santoro L, Estraneo A (2013) Affective saliency modifies visual tracking behavior in disorders of consciousness: a quantitative analysis. J Neurol 260(1):306–308. doi:10.1007/s00415-012-6717-x
21. Hoyoux T, Wannez S, Langohr T, Wertz J, Laureys S, Verly J (2016) A new computer vision-based system to help clinicians objectively assess visual pursuit with the moving mirror stimulus for the diagnosis of minimally conscious state. In: IEEE Winter Conference on Applications of Computer Vision (WACV 2016)
22. Collobert R, Kavukcuoglu K, Farabet C (2011) Torch7: A matlab-like environment for machine learning. NIPS Work, BigLearn, pp 1–6
23. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
24. Demertzi A, Ledoux D, Bruno M-A et al (2011) Attitudes towards end-of-life issues in disorders of consciousness: a European survey. J Neurol 258:1058–1065
25. Demertzi A, Schnakers C, Ledoux D et al (2009) Different beliefs about pain perception in the vegetative and minimally conscious states: a European survey of medical and paramedical professionals. Prog Brain Res 177:329–338
26. Bruno M-A, Ledoux D, Vanhaudenhuyse A, Gosseries O, Thibaut A, Laureys S (2011) Pronostic des patients récupérant du coma. In: Schnakers C, Laureys S (eds) Coma et États de Conscience Altérée. Springer, Paris, pp 17–30