

BMJ Open Wisdom of the caregivers: pooling individual subjective reports to diagnose states of consciousness in brain-injured patients, a monocentric prospective study

Bertrand Hermann,^{1,2,3,4} Gwen Goudard,¹ Karine Courcoux,¹ Mélanie Valente,^{2,3,4,5} Sébastien Labat,¹ Lucienne Despois,¹ Julie Bourmaleau,¹ Louise Richard-Gillis,^{1,2,3,4} Frédéric Faugeras,^{2,3,4} Sophie Demeret,¹ Jacobo D Sitt,^{2,3,4} Lionel Naccache,^{1,2,3,4,5,6} Benjamin Rohaut^{1,2,3,4,7}

To cite: Hermann B, Goudard G, Courcoux K, *et al.* Wisdom of the caregivers: pooling individual subjective reports to diagnose states of consciousness in brain-injured patients, a monocentric prospective study. *BMJ Open* 2019;**9**:e026211. doi:10.1136/bmjopen-2018-026211

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-026211>).

BH, GG and KC contributed equally.

Received 23 August 2018
Revised 27 December 2018
Accepted 2 January 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Benjamin Rohaut;
br2529@columbia.edu

ABSTRACT

Objectives The clinical distinction between vegetative state/unresponsive wakefulness syndrome (UWS) and minimally conscious state (MCS) is a key step to elaborate a prognosis and formulate an appropriate medical plan for any patient suffering from disorders of consciousness (DoC). However, this assessment is often challenging and may require specialised expertise. In this study, we hypothesised that pooling subjective reports of the level of consciousness of a given patient across several nursing staff members can be used to clinically detect MCS.

Setting and participants Patients referred to consciousness assessment were prospectively screened. MCS (target condition) was defined according to the best Coma Recovery Scale-Revised score (CRS-R) obtained from expert physicians (reference standard). 'DoC-feeling' score was defined as the median of individual subjective reports pooled from multiple staff members during a week of hospitalisation (index test). Individual ratings were collected at the end of each shift using a 100 mm Visual Analogue Scale, blinded from the reference standard. Diagnostic accuracy was evaluated using area under the receiver operating characteristic curve (AUC), sensitivity and specificity metrics.

Results 692 ratings performed by 83 nursing staff members were collected from 47 patients. Twenty patients were diagnosed with UWS and 27 with MCS. DoC-feeling scores obtained by pooling all individual ratings obtained for a given patient were significantly greater in patients with MCS than with UWS (59.2mm (IQR: 27.3–77.3) vs 7.2mm (IQR: 2.4–11.4); $p < 0.001$) yielding an AUC of 0.92 (95% CI 0.84 to 0.99).

Conclusions DoC-feeling capitalises on the expertise of nursing staff to evaluate patients' consciousness. Together with the CRS-R as well as with brain imaging, DoC-feeling might improve diagnostic and prognostic accuracy of patients with DoC.

INTRODUCTION

Accurate diagnosis of the level of consciousness in a brain-damaged patient is of great

Strengths and limitations of this study

- We designed a new behavioural tool called disorders of consciousness 'DoC-feeling' to help face the clinical challenge of the detection of minimally conscious state in brain-injured patients suffering from DoC.
- 'DoC-feeling' pools the subjective reports of patient's consciousness obtained from multiple caregivers ('wisdom of the crowds').
- The obtained score shows a very good accuracy when compared with the gold standard (repeated expert clinical assessments using the Coma Recovery Scale-Revised (CRS-R)).
- A validation in a separate cohort would help to precise its value in routine consciousness assessment.
- This approach should be further compared to the CRS-R and brain-imaging techniques in detecting covert signs of consciousness.

importance to better predict recovery. Disorder of consciousness (DoC) taxonomy has been recently challenged^{1–3} but schematically includes the unresponsive wakefulness syndrome (UWS, also termed vegetative state) and the minimally conscious state (MCS). The detection of MCS has a huge prognostic impact since the functional outcome is dramatically better for patients with MCS.^{4–8} However, assessing consciousness in patients with DoC can be challenging and in such cases, clinicians may need dedicated clinical tools and brain-imaging techniques specifically designed to probe consciousness.⁹ Even when using dedicated clinical tools such as the Coma Recovery Scale-Revised (CRS-R¹⁰), a unique assessment remains associated with a high frequency of diagnostic error.¹¹ This can

be due to fluctuations of consciousness level over time. To circumvent this limitation, repeated clinical assessments have been proposed, but this can be limited by the availability of trained clinicians.^{12 13}

In this study, we aimed at evaluating the diagnostic accuracy of pooled nursing staff estimations of the level of consciousness in patients with DoC. Through their clinical practice, nursing staff (ie, nurses and nursing assistants) accumulates extended observation time of patient's behaviour. Interacting with patients through standardised procedures (such as nursing care, medication administration, blood sample, etc...), they spontaneously generate a subjective estimation of the level of consciousness of the patient. Pooling opinions of several individuals have been shown to outperform individual judgements in specific settings (effect known as 'wisdom of the crowds').^{14 15} In this study, we hypothesised that pooling individual nursing staff estimations of the level of consciousness can help in the detection of MCS.

METHODS

Patients

All patients referred for evaluation of consciousness at the Department of Neurology of La Pitié-Salpêtrière Hospital, Paris, between February 2016 and October 2017, were screened prospectively. On hospital admission, patients' relatives were approached to give consent for participation to the study. All patients with a UWS or MCS condition and consent were eligible.

Patient and public involvement

No patients or patients' relatives were involved in the study design or the management of this study. Results of the study have been released as a preprint on a public repository¹⁶ and the dataset of this study is available on Dryad (<https://doi.org/10.5061/dryad.1m03145>).

Evaluation of consciousness

Reference standard

Patients were hospitalised in the neurointensive care unit (neuro-ICU) and were observed for at least 1 week during which they encompassed multiple neurological assessments and brain imagery such as high-density electroencephalogram, event-related potentials, magnetic resonance imaging and [¹⁸F]-fluorodeoxyglucose positron emission tomography. Clinical assessments consisted of repeated neurological examinations which included the CRS-R,¹⁷ performed by expert clinicians (BH, BR, FF, LN) belonging to an external expert team in patients with DoC. CRS-R scoring ranges from 0 to 23 and is based on the presence or absence of responses on a set of hierarchically ordered items testing auditory, visual, motor, oromotor, communication and arousal function. State of consciousness (ie, UWS, MCS) is determined by specific key behaviours probed during the CRS-R assessment. For instance, visual pursuit, reproducible movements to command and/or complex motor behaviour scores for

MCS.¹⁷ Since consciousness level can fluctuate over time, we used the highest level of consciousness among all the CRS-R performed on a given patient as the reference standard. Following this procedure, each patient was thus labelled as being in a UWS or MCS. MCS was the target condition.

Index test

Nursing staff members (nurses and nursing assistants) taking care of a DoC patient were asked to fill in a form at the end of their shift containing a scale called 'DoC-feeling'. DoC-feeling was designed as a 100 mm Visual Analogue Scale (VAS) aiming at quantifying the caregiver subjective reports of patient's best consciousness level observed during the shift. We specifically asked caregivers to rate their 'gut feeling' about the best level of consciousness observed during the shift or the 'présence' (presence), using the French idiom 'le patient est-il là?' which is very close to the English one 'Is there anybody home?' (figure 1; see online supplementary material for the original VAS and its English translation). This wording reproduced the commonly used language to communicate observations relative to consciousness level of a patient among caregivers. Individual DoC-feeling ratings were collected prospectively. Caregivers were blinded to the previous caregivers' ratings and to the reference standard (the CRS-R) and expert physicians were blinded to the index test. In order to obtain a final global metric, for each patient, all individual ratings were pooled using the median to obtain the DoC-feeling score that constituted the index test of this study.

Clinical data

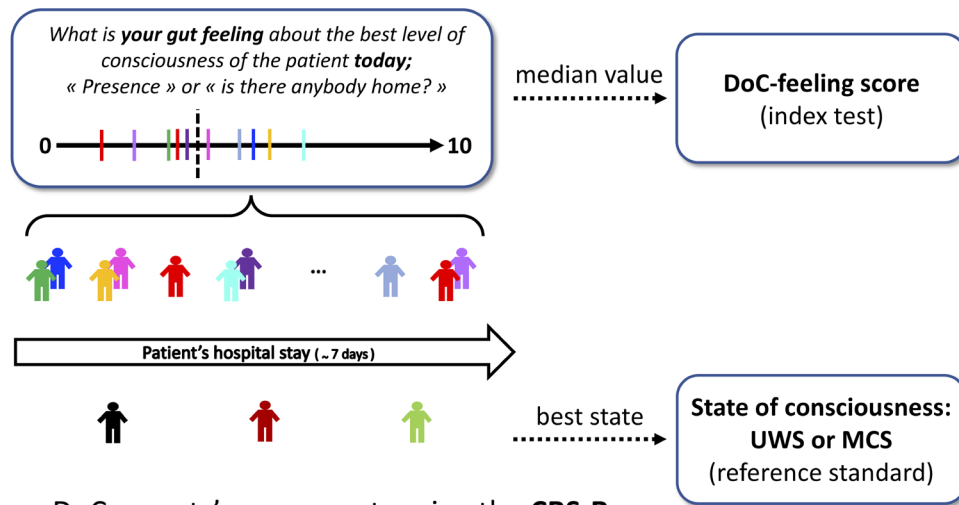
Demographics, aetiology and delay since the acute brain injury (ABI) were collected. In addition to CRS-R and DoC-feeling ratings, we also collected complementary metrics (such as the classical distinction between wakefulness and awareness, interaction during nursing and/or painful care) using the same VAS approach as well as the best FOUR-score observed during each shift¹⁸ (online supplementary material).

Statistics

Our primary objective was to evaluate the diagnostic accuracy of the index test called 'DoC-feeling score' to detect the target condition (MCS) as defined by the standard reference (best CRS-R).

First, to evaluate the association of individuals' DoC-feeling ratings with the standard reference, we computed a linear mixed model (LMM) using DoC-feeling individual ratings as the dependent variable, the state of consciousness as the fixed effect explanatory variable and patients as well as raters as random effects. Normality of residuals distribution was assessed by visual inspection. LMM provides the optimal approach in order to take into account the non-independence between DoC-feeling ratings due to the repeated measurements over time at both the patient level (same patient rated by several raters) and the rater level (several ratings by rater).

- Nursing staff assessments using the **DoC-feeling scale**



- DoC experts' assessments using the **CRS-R**

Figure 1 Disorders of consciousness (DoC)-feeling score. Each patient was evaluated around three times by DoC experts using the Coma Recovery Scale-Revised (CRS-R). In parallel, nursing staff members reported their daily observations using the DoC-feeling Visual Analogue Scale. The reference standard was defined as the best state of consciousness observed during one of the CRS-R and the patient was coded as being in an unresponsive wakefulness syndrome (UWS) or a minimally conscious state (MCS) accordingly (reference standard). All individual DoC-feeling scores obtained during the whole hospital stay were pooled and the median value (represented by the vertical dashed line) of the pooled results was defined as the DoC-feeling score (index test).

We next pooled the individual ratings obtained for each patient using the median to obtain the DoC-feeling score (index test). We, thus, obtained a DoC-feeling score as well as a reference standard label (UWS or MCS) for each patient. We performed a direct comparison of the scores between the two populations using a Wilcoxon-Mann-Whitney test. In order to assess the diagnostic accuracy of DoC-feeling scores to detect MCS (target condition), we computed the area under the receiver operating characteristic (ROC) curve (AUC) and report sensitivities and specificities for several cut-offs of DoC-feeling scores. All statistical tests were two sided with a type I error rate of 5%. Categorical variables were expressed as numbers (percentage), quantitative variables as median (IQR). Analyses were performed using the R statistical software V.3.4.1.¹⁹ LMM was performed using the lme4 package.²⁰ AUC, sensitivity and specificity with their 95% CIs were computed using 2000 stratified bootstrap replicates (AUC) and binomial test (sensitivity and specificity) respectively using the pROC package.²¹

The Standards for Reporting Diagnostic Accuracy was followed thoroughly.²²

RESULTS

Patients characteristics

Flow chart

Seventy-two patients were eligible during the inclusion period, 23 were not included because of a lack of informed consent from a legal representative. Two patients were

excluded because they had been diagnosed as conscious ('Exit-MCS'). Forty-seven patients were included in the analysis (see figure 2).

Median age was 49 (32–62) years, 66% (n=31) were female. Main aetiologies of brain injury included anoxia (53%) and traumatic brain injury (17%). Delay between ABI and the evaluation was 134 (40–762) days (see table 1).

Reference test

One hundred and forty-seven CRS-R assessments were performed, with a median of 3^{2–4} per patient (ranging from 2 to 6). According to the best CRS-R, 27 patients (57%) were diagnosed as being in an MCS and 20 (43%) were classified as being in a UWS. Patients with MCS less frequently suffered from anoxia and had a longer delay between the ABI and the study inclusion (see table 1). No differences were found in the number of CRS-R assessments per patient or brain-imaging explorations between patients with UWS and MCS.

Index test

Six hundred and ninety-two DoC-feeling individual ratings were obtained (median of 12^{9–19} ratings per patient). Eighty-three caregivers, 57 nurses and 26 nurses assistants (composed of 47 neuro-ICU regular staff members and 36 float staff members) participated in the study. Each nursing staff member filled a median of 4^{1–12} evaluations. Median delay between the first and the last individual rating was 6 days.^{5–9} No statistical differences were found

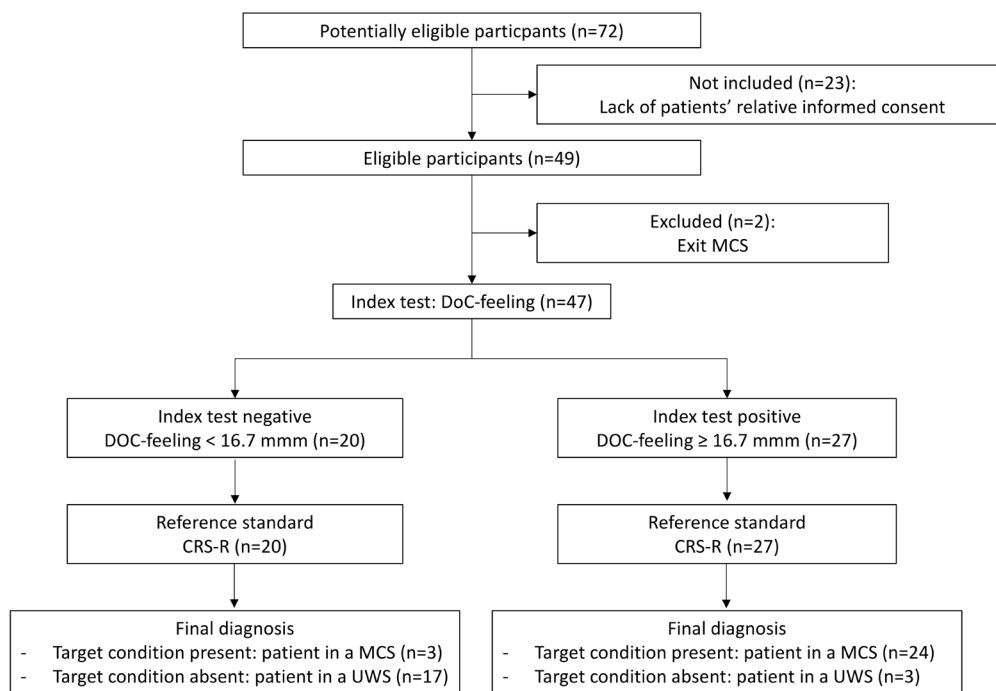


Figure 2 Flow chart. Flow chart representing the repartition of patients while using a disorder of consciousness (DoC)-feeling score (index test) cut-off value of 16.7 mm. Exit-MCS: Patient able to communicate reliably or to use objects functionally. CRS-R, Coma Recovery Scale-Revised; MCS, minimally conscious state; UWS, unresponsive wakefulness syndrome.

between UWS and MCS in the number of DoC-feeling ratings per patient, a number of raters per patient or in terms of number of ratings per rater (table 1).

Analysis of individuals DoC-feeling ratings

Inspection of the 692 DoC-feeling ratings' distribution revealed higher values for patients with MCS than for UWS but with an important variability of ratings for a given patient (figure 3). The LMM analysis revealed a strong significant association between DoC-feeling individuals' ratings and the state of consciousness ($t=6.47$, $df=45$, $p<0.001$).

Diagnostic accuracy of DoC-feeling scores

Overall, patients underwent 12^{9–19} DoC-feeling individual ratings, performed by 7^{5–10} different raters. All DoC-feeling ratings obtained for a given patient were summarised using the median to obtain the pooled metric called DoC-feeling score (index test, figure 4A). DoC-feeling scores were smaller for patients with UWS than for MCS (7.2 mm (2.4–11.4) vs 59.2 mm (27.3–77.3), respectively; $p<0.001$; figure 4B). ROC curve revealed excellent accuracy at detecting MCS (AUC=0.92 (95% CI 0.84 to 0.99); figure 4C) with, for instance, a sensitivity of 89% (95% CI 71% to 98%) and a specificity of 85% (95% CI 62 to 97) when using a DoC-feeling score cut-off at 16.7 mm (figure 4D). Note that this cut-off is only used to give the reader an idea about the diagnostic performances using the more intuitive sensitivity and specificity metrics (see

the Discussion section). The six misclassified patients using this cut-off are described in the online supplementary material. Simulations of AUCs using a various number of ratings per patient suggested that a minimal number of 4 ratings is needed to reach an AUC of 0.9 (online supplementary material). Of note, DoC-feeling score also helped discriminate UWS patients from MCS 'minus' patients (patients with non-reflexive behaviours but absence signs of language at bedside)²³ (see online supplementary material for additional details).

DISCUSSION

In the present study, we developed and assessed a new behavioural tool called DoC-feeling to help diagnose MCS. This score, which pools multiple subjective reports obtained among several caregivers over several days of evaluation, showed a very good accuracy to diagnose MCS.

DoC-feeling is not intended to replace the clinical examination nor the current CRS-R gold standard. However, taking advantage of valuable information collected by all caregivers involved in the care of a patient with DoC, the implementation of DoC-feeling could improve the overall diagnostic accuracy of patients with DoC. Caregivers are trained to evaluate pain and suffering in patients during all delivered procedures. These procedures constitute standardised interactions that can allow the generation of

Table 1 Patient characteristics

	All (n=47)	UWS (n=20)	MCS (n=27)	P value
Demographic characteristics				
Age, years	49 (32–62)	50 (35–65)	47 (30–59)	0.38
Sex ratio (F/M)	0.51	0.33	0.69	0.42
Aetiology				0.02
Anoxia	25 (53)	16	9	
TBI	8 (17)	2	6	
Stroke	6 (13)	0	6	
Other	8 (17)	2	6	
Time from ABI, days	134 (40–762)	57 (27–185)	374 (70–916)	<0.01
Mechanically ventilated	20 (43)	10 (50)	8 (30)	0.26
Neurological evaluation				
Nb of CRS-R/patient	3 (2–4)	3 (2–3)	3 (2–4)	0.10
Best CRS-R score	8 (5–11)	5 (4–6)	11 (9–13)	<0.001
FOUR-score	13 (10–13)	11 (10–11)	13 (13–13)	<0.0001
DoC-feeling assessment				
Nb of raters, (nurses/NAs)	83 (57/26)	59 (40/19)	67 (42/25)	0.13
Nb of ratings, (nurses/NAs)	692 (489/203)	289 (213/76)	403 (276/127)	0.16
Nb of ratings per rater	4 (1–12)	3 (2–7)	2 (1–6)	0.40
Nb of ratings per patient	12 (9–19)	13 (9–20)	12 (9–19)	1.00
Nb of raters per patient	7 (5–10)	6 [5–10]	7 (6–10)	0.86
Time between first and last assessment, days	6 (5–9)	7 (5–9)	6 (5–8)	0.27
Brain imagery assessment				
EEG/ERPs	44 (94)	19 (95)	25 (93)	0.13
MRI	40 (85)/24 (51)	18 (90)/11 (55)	22 (81)/13 (48)	1.00
PET scan*	28 (60)	9 (45)	19 (70)	0.39

Results are expressed in n(%) or median(IQR) as appropriate.

*PET scan was performed only in patients free of mechanical ventilation.

ABI, acute brain injury; CRS-R, coma recovery scale-revised; DoC, disorders of consciousness; EEG/ERPs, electroencephalogram/event related potentials; FOUR, full outline of unresponsiveness; MCS, minimally conscious state; MRI, magnetic resonance imaging; NAs, nursing assistants; Nb, number; PET, positron emission tomography; TBI, traumatic brain injury; UWS, unresponsive wakefulness syndrome.

very reliable heuristic processes to assess one's percept in terms of pain suffering and also consciousness.

Pooling opinions of several individuals have been previously shown to outperform individual judgements in specific settings. Recently, there has been a growing interest for this kind of approach (called collective intelligence or 'wisdom of the crowds') in the medical field, especially in diagnosis procedure (diagnosis of skin cancer, mammography screening, etc...).^{14 24–26}

In that perspective, quantifying expertise that is not restricted to physicians might be of prime interest. Capitalising on assessments of consciousness gathered at any hour of the day and through multiple observers may also potentially increase our ability to detect signs of consciousness in these patients who usually show large fluctuations of cognitive state and arousal.¹² DoC-feeling may also help to better describe and quantify these fluctuations. Additionally, it also enables to acknowledge

the caregiver group expertise and to increase care team attention through a coherent and cumulative set of observational data.

The good accuracy of DoC-feeling obtained in our setting is likely to be generalisable elsewhere. First, as the distribution of CRS-R scores obtained in this cohort spanned most of the possible CRS-R scores, it is unlikely that the good accuracy of DoC-feeling results from two easily discernible patients' clusters. Second, as all the patients included in this study, either in an acute or a chronic stage, were specifically referred to our institution for expertise, it is most likely that our cohort was actually representative of patients for whom the diagnosis is the most difficult. However, we would like to emphasise that the used cut-off in the result section might be variable across teams and across time for a given team. This is why DoC-feeling should only be used in addition and not instead of CRS-R.

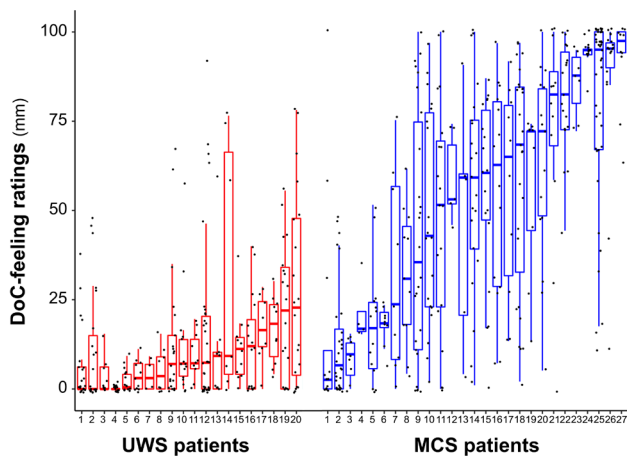


Figure 3 Individual disorders of consciousness (DoC)-feeling ratings. DoC-feeling ratings tended to be smaller in patients with unresponsive wakefulness syndrome (UWS) when compared with patients with minimally conscious state (MCS). All individual ratings are shown (dots, n=692), alongside boxplots helping to visualise the median and the IQR for both UWS (on the left in red) and MCS (on the right in blue) patients.

Our study presents some limitations inherent to the aim of developing a pragmatic and easily implementable tool in daily clinical practice. First, as for all studies on consciousness disorders, we faced a typical situation of an imperfect gold standard. Although CRS-R is still the most widely accepted reference, the optimal number of assessments remains unknown.¹³ According to a recent study, using three CRS-R assessments can lead to a 17% rate of misdiagnoses.¹² It is worth noting that this is exactly the reason why we developed DoC-feeling. CRS-R requires a specialised expertise that is not available everywhere and that can be extremely time-consuming, especially now that multiple assessments are recommended to take into account fluctuations of consciousness over time.¹³ In sharp contrast, DoC-feeling scale could be implemented in any team, is much faster and allows to gather multiple observations per day. Second, caregivers might have been influenced by other factors that would have been very difficult to control. For instance, they might have been influenced by insights from other caregivers or, in case of multiple ratings for a given rater, by their previous ratings. However, the variability of individual ratings for a given patient (that tended to increase over time, see online supplementary material) suggests that caregivers did report their own perception independently from each other and their eventual previous ratings. Moreover, interactions among small groups of people could, in fact, have had a positive effect since the aggregation of small groups' insights have been shown to outperform the overall judgement of the whole group.²⁷ This kind of tool might thus be less prone to individual subjective bias which is frequent during decision-making under a high degree of uncertainty such as assessment of patients with DoC.²⁸ Staff members could also have been biased by classical predictors

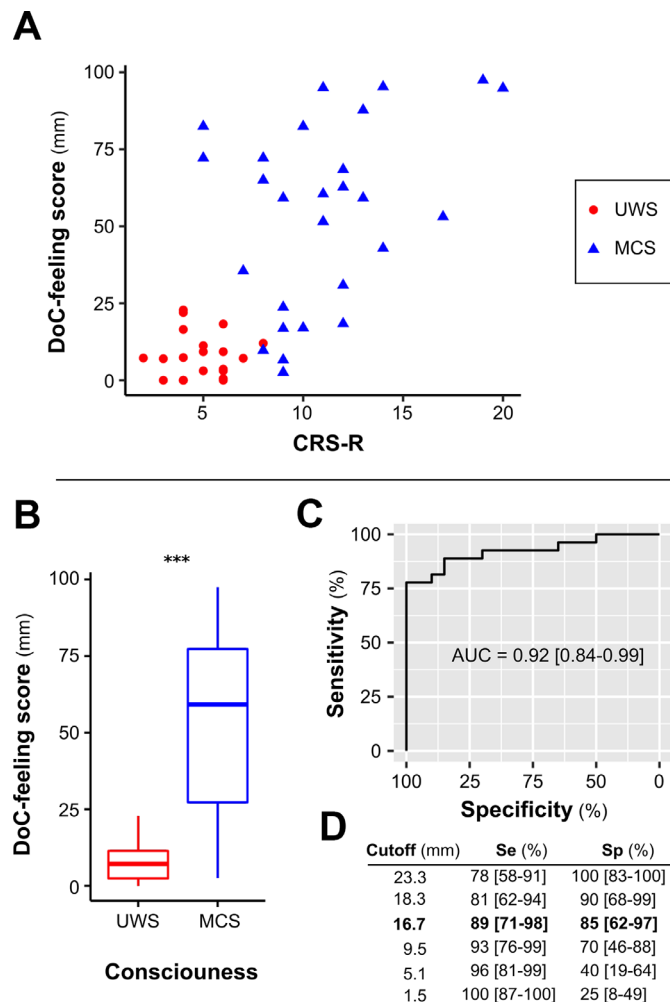


Figure 4 DoC-feeling scores. DoC-feeling scores were obtained by pooling individual ratings obtained for each patient. DoC-feeling scores were smaller for patients with UWS than for MCS (A, B) and also correlated with the CRS-R score (A). Area under the ROC curve (C), sensitivity (Se) and specificity (Sp) for several cut-offs (D) revealed very good performances at identifying the MCS. ***P<0.001. CRS-R, Coma Recovery Scale-Revised; DoC, disorders of consciousness; MCS, minimally conscious state; ROC, receiver operating characteristic; UWS, unresponsive wakefulness syndrome.

of consciousness recovery such as aetiology or delay from ABI or by the perception of patients' relatives, although it is commonly acknowledged that relatives frequently lack objectivity (in both directions) in such dramatic situations.²⁹ Finally, although the number of float staff members involved and the result of a preliminary survey assessing prior knowledge of regular nursing staff on DoC (online supplementary material) suggest together that DoC-feeling should be accurate in other settings, the monocentric design of this study requests external validation.

Despite these limitations, we think that the implementation of DoC-feeling score can significantly improve diagnostic accuracy and confidence in the diagnosis when supporting other metrics (ie, CRS-R and functional brain imaging at rest or during cognitive tasks). Moreover, even

when incongruent with other metrics, DoC-feeling score could be still useful. Indeed, this could either suggest that key clinical elements have been missed by physicians while performing punctual CRS-R assessments, but it could also reveal, in case of discrepancy with all the other elements (clinical and brain imagery), a possible misperception of a patient's consciousness level that needs to be acknowledged and considered in any further medical decision processes. This last point could be crucial in bridging the gap between the caregiver's team and the patient's relatives in situations of conflict.

In conclusion, we propose a new behavioural tool, called DoC-feeling, based on the 'wisdom of the crowds' effect (or, in our case, the 'wisdom of the caregivers'), which can help to improve the diagnostic of MCS and thus to promote a better prognostication and decision-making in patients suffering from DoC.

Author affiliations

¹Department of Neurology, Neuro ICU, Groupe Hospitalier Pitié-Salpêtrière, AP-HP, Paris, France

²Institut du Cerveau et de la Moelle épinière, ICM, PICNIC lab, F-75013, Paris, France

³Inserm U 1127, F-75013, Paris, France

⁴CNRS, UMR 7225, F-75013, Paris, France

⁵Department of Neurophysiology, Groupe Hospitalier Pitié-Salpêtrière, AP-HP, Paris, France

⁶Sorbonne Universités, Faculté de Médecine Pitié-Salpêtrière, Paris, France

⁷Division of Critical Care and Hospitalist Neurology, Columbia University, New York City, New York

Acknowledgements We thank all the members of the Pitié-Salpêtrière hospital Neuro-ICU led by SD (medical director), JB and LR-G (head nurses) who participated in this study (alphabetic order): Jérémie Abitbol, Fatiha Ait Yata Azzi, Fatoumata Bah, Francis Bolgert, Sandrine Briand, Sandra Coelho, Alexia Camuzat, Marie-Chantal Colmar, Flora Cherruault, Cecile Chordi, Véronique Cottin, Bintou Coulibaly, KC, Mélanie Dalibard, LD, Estelle Dumarey, Bouchra El Aouni, Atef El Ouarghi, Helene Espiand, Cécilia Eltebert, Fabrice Fanhan, Agnès Flament, Marie-Suzelle Fontano, Pascale Fournier, Céline Frammezelle, GG, Alexandra Grinéa, Noura Harchaoui, Marie Harmancij, Claire Jacqueminet, Charlotte Janvier, Jamila Kebli, SL, Aurélie Lemoal, Kim Louis-Joseph, Brice Lucas, Valérie Maes, Sophie Maillard, Romain Maurel, Madely Petit, Floriane Pépin, Isabelle Picot, Eva Proneur, Manuela Roselmac, Sylviane Saintini, Mélody Seidel, Yolène Sully, Kelly Tcha, Laura Verbaux, Nicolas Weiss, Kelly Yanganju. We thank all the members of the PICNIC-Lab 'DoC-Team', led by LN and dedicated to the improvement of care of patients suffering from disorder of consciousness (alphabetic order): Athena Demertzi, Denis Engemann, FF, BH, Pauline Pérez, Federico Raimondo, BR; Johan Stender, MV and JDS. We thank Raphael Porcher for his help on statistical issues and Jan Claassen for his final review of our manuscript and finally, the two reviewers for their very constructive comments.

Collaborators Jérémie Abitbol; Fatiha Ait Yata Azzi; Fatoumata Bah; Francis Bolgert; Sandrine Briand; Sandra Coelho; Alexia Camuzat; Marie-Chantal Colmar; Flora Cherruault; Cecile Chordi; Véronique Cottin; Bintou Coulibaly; Mélanie Dalibard; Athena Demertzi; Estelle Dumarey; Bouchra El Aouni; Atef El Ouarghi; Denis Engemann; Helene Espiand; Cécilia Eltebert; Fabrice Fanhan; Agnès Flament; Marie-Suzelle Fontano; Pascale Fournier; Céline Frammezelle; Alexandra Grinéa; Noura Harchaoui; Marie Harmancij; Claire Jacqueminet; Charlotte Janvier; Jamila Kebli; Aurélie Lemoal; Kim Louis-Joseph; Brice Lucas; Valérie Maes; Sophie Maillard; Romain Maurel; Madely Petit; Floriane Pépin; Pauline Pérez; Isabelle Picot; Eva Proneur; Federico Raimondo; Manuela Roselmac; Sylviane Saintini; Mélody Seidel; Johan Stender; Yolène Sully; Kelly Tcha; Laura Verbaux; Nicolas Weiss; Kelly Yanganju.

Contributors Study concept and design: BR, FF, GG, JB, JDS, KC, LD, LN, LR-G, SD and SL. Data collection: BH, GG, KC, LD, MV and SL. Analysis and interpretation of data: BH and BR. Drafting of the manuscript: BH, BR and LN. Critical revision of the manuscript for important intellectual content: BH, BR, JDS, LN and SD. Statistical

analysis: BH, BR and LN. Study supervision: BH, BR, GG, KC and MV had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. BH, GG and KC contributed equally to this work.

Funding This work was supported by: Amicale des Anciens Internes des Hôpitaux de Paris & Syndicat des Chefs de Cliniques et Assistants des Hôpitaux de Paris (AAIHP—SCCAHP; BR), Assistance Publique—Hôpitaux de Paris (AP-HP; BR and LN), Institut National de la Santé et de la Recherche Médicale (Inserm; BH, JDS and LN), Sorbonne Université (LN), the James S. McDonnell Foundation (LN), Académie des Sciences- Lamonica Prize 2016 (LN) and Philippe Foundation (BR). The research leading to these results has received funding from the program 'Investissements d'avenir' ANR-10- IAIHU-06.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The protocol conformed to the Declaration of Helsinki, to the French regulations, and was approved by the local ethic committee (Comité de Protection des Personnes; CPP no 2013-A01385-40; Ile de France 1; Paris, France).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The dataset of this study is available on <https://doi.org/10.5061/dryad.1m03145>.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Naccache L. Minimally conscious state or cortically mediated state? *Brain* 2018;141:949–60.
- Bayne T, Hohwy J, Owen AM. Reforming the taxonomy in disorders of consciousness. *Ann Neurol* 2017;82:866–72.
- Bernat JL. Nosologic considerations in disorders of consciousness. *Ann Neurol* 2017;82:863–5.
- Luauté J, Maucourt-Boulch D, Tell L, *et al*. Long-term outcomes of chronic minimally conscious and vegetative states. *Neurology* 2010;75:246–52.
- Noé E, Olaya J, Navarro MD, *et al*. Behavioral recovery in disorders of consciousness: a prospective study with the Spanish version of the Coma Recovery Scale-Revised. *Arch Phys Med Rehabil* 2012;93:428–33.
- Bruno M-A, Ledoux D, Vanhauzenhuyse A, *et al*. Prognosis of patients with altered state of consciousness. In: *Coma and disorders of consciousness*: Springer, 2012:11–23.
- Klein AM, Howell K, Vogler J, *et al*. Rehabilitation outcome of unconscious traumatic brain injury patients. *J Neurotrauma* 2013;30:1476–83.
- Faugeras F, Rohaut B, Valente M, *et al*. Survival and consciousness recovery are better in the minimally conscious state than in the vegetative state. *Brain Inj* 2018;32:72–7.
- Giacino JT, Fins JJ, Laureys S, *et al*. Disorders of consciousness after acquired brain injury: the state of the science. *Nat Rev Neurol* 2014;10:99–114.
- Giacino JT, Ashwal S, Childs N, *et al*. The minimally conscious state: definition and diagnostic criteria. *Neurology* 2002;58:349–53.
- Schnakers C, Vanhauzenhuyse A, Giacino J, *et al*. Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC Neurol* 2009;9:35.
- Wannez S, Heine L, Thonnard M, *et al*. The repetition of behavioral assessments in diagnosis of disorders of consciousness. *Ann Neurol* 2017;81:883–9.
- Giacino JT, Katz DI, Schiff ND, *et al*. Practice guideline update recommendations summary: Disorders of consciousness. *Neurology* 2018;91:450–60.
- Kurvers RH, Herzog SM, Hertwig R, *et al*. Boosting medical diagnostics by pooling independent judgments. *Proc Natl Acad Sci U S A* 2016;113:8777–82.
- Bang D, Frith CD. Making better decisions in groups. *R Soc Open Sci* 2017;4:170193.
- Hermann B, Goudard G, Courcoux K, *et al*. "DoC-feeling": a new behavioural tool to help diagnose the Minimally Conscious State. *bioRxiv* 2018:370775.



17. Giacino JT, Kalmar K, Whyte J. The JFK Coma recovery scale-revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* 2004;85:2020–9.
18. Wijdicks EF, Bamlet WR, Maramattom BV, et al. Validation of a new coma scale: the FOUR score. *Ann Neurol* 2005;58:585–93.
19. R Development Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput* 2017 <http://www.R-project.org/>
20. Pinheiro J, Bates D, DebRoy S, et al. nlme: Linear and Nonlinear Mixed Effects Models. *R Package Version* 2017 <https://CRAN.R-project.org/package=nlme>
21. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
22. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
23. Bruno MA, Vanhaudenhuyse A, Thibaut A, et al. From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *J Neurol* 2011;258:1373–84.
24. Wolf M, Krause J, Carney PA, et al. Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. *PLoS One* 2015;10:e0134269.
25. Kurvers RH, Krause J, Argenziano G, et al. Detection accuracy of collective intelligence assessments for skin cancer diagnosis. *JAMA Dermatol* 2015;151:1346–53.
26. Hautz WE, Kämmer JE, Schaubert SK, et al. Diagnostic performance by medical students working individually or in teams. *JAMA* 2015;313:303–4.
27. Navajas J, Niella T, Garbulsy G, et al. Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nat Hum Behav* 2018;2:126–32.
28. Rohaut B, Claassen J. Decision making in perceived devastating brain injury: a call to explore the impact of cognitive biases. *Br J Anaesth* 2018;120:5–9.
29. Rohaut B, Faugeras F, Naccache L. Neurology of consciousness impairments. In: Stevens RD, Sharshar T, Ely EW, eds. *Brain disorders in critical illness*: Cambridge University Press, 2013:59–67.