Brain Diffusion Tensor Imaging Predicts Long Term Outcome in Patients after Cardiac Arrest: a multicenter, prospective cohort study

Lionel Velly, Vincent Perlbarg, Thomas Boulier, Nicolas Adam, Sebastien Delphine, Charles-Edouard Luyt, Valentine Battisti, Gregory Torkomian, Charlotte Arbelot, Russell Chabanne, Betty Jean, Carol Di Perri, Steven Laureys, Giuseppe Citerio, Alessia Vargiolu, Benjamin Rohaut, Nicolas Bruder, Nadine Girard, Stein Silva, Vincent Cottenceau, Thomas Tourdias, Olivier Coulon, Bruno Riou, Lionel Naccache, Rajiv Gupta, Habib Benali, Damien Galanaud, Louis Puybasset, for the MRI-COMA Investigators*

*A complete list of investigators in the Multimodal Resonance Imaging for Outcome Prediction on Coma Patients (MRI-COMA) study is provided in the appendix

The authors' affiliations are as follows: the Assistance Publique–Hôpitaux de Paris, Groupe Hospitalier Pitié–Salpêtrière (L Puybasset MD PhD, D Galanaud MD PhD, C-E Luyt MD PhD, V Battisti PhD, G Torkomian MEng, C Arbelot MD, N Adam MD, B Riou MD PhD, and B Rohaut MD) and Laboratoire d'Imagerie Biomédicale (UMR S 1146 / UMR 7371), Université Pierre-et-Marie-Curie-Paris 06, Paris (H Benali PhD, V Perlbarg PhD, S Delphine MEng, T Boulier PhD); Hôpital Universitaire Pellegrin, Bordeaux, (T Tourdias MD PhD, V Cottenceau MD), Centre Hospitalier Universitaire de Clermont-Ferrand (B Jean MD, R Chabanne MD), Clermont-Ferrand; Assistance Publique–Hôpitaux de Marseille, Hôpital Timone Adultes, (N Girard MD, N Bruder MD) and MeCA (L Velly MD PhD, O Coulon PhD) - Institut des Neurosciences (UMR 7289, CNRS), Aix Marseille Univ, Marseille; Neuro-Campus Baudot - Institut des Sciences du Cerveau and Hôpital Universitaire, Toulouse, (S Silva MD, PhD) ; Institut du

Cerveau et de la Moelle Epinière (PICNIC LAB), Université Pierre-et-Marie-Curie-Paris 06, Paris (L Naccache MD PhD) – all in France; Neurology Dept & Cyclotron Research Centre University & University Hospital of Liège Sart Tilman, Liège – Belgium (S Laureys MD PhD, C Di Perri MD PhD); Milan Bicocca University, Milan and Hospital San Gerardo, Monza – Italia (G Citerio MD PhD, A Vargiolu PhD); Massachusetts General Hospital, Harvard Medical School, Boston, MA – USA (R Gupta MD PhD).

Short running head: Predicting Outcome from Cardiac Arrest

Address reprint requests to Dr. Puybasset at the Department of Anesthesiology and Intensive Care, Groupe Hospitalier Pitié–Salpêtrière, Assistance Publique–Hôpitaux de Paris, 47-83 Blvd. de l'hôpital; 75013 Paris, France, or at louis.puybasset@aphp.fr.

Keywords: cardiac arrest, outcome, prognosis, magnetic resonance imaging, diffusion tensor imaging

SUMMARY

Background

Prediction of neurologic outcome after cardiac arrest is a major challenge. The aim of this study was to examine whether quantitative whole brain white matter fractional anisotropy (WWM-FA) measured by diffusion tensor imaging (DTI) between day 7 and day 28 can predict clinical outcome of such patients.

Methods

A prospective, observational cohort study (part of trial MRI-COMA) was conducted in 14 centers in France, Italy, and Belgium. Patients were eligible if they were unconscious at least 7 days after cardiac arrest. WWM-FA values were compared to standard criteria for unfavorable outcome, conventional magnetic resonance imaging (MRI) sequences (fluid-attenuated inversion recovery and diffusion-weighted imaging), and proton magnetic resonance spectroscopy. The outcome evaluated was the best achieved Cerebral Performance Category (CPC) at 6 months, dichotomized as favorable (CPC 1-2) and unfavorable outcome (CPC 3-5), and the performance were compared by area under the receiver operating characteristic (ROC_{AUC}) curves analysis with that of a logistic regression model. This study is registered with ClinicalTrials.gov, number NCT00577954.

Findings

In the derivation cohort, 185 patients were enrolled and 150 had an interpretable multimodal MRI. Thirty-three patients (22%) had a favorable neurologic outcome. Prognostic accuracy, as quantified by the ROC_{AUC}, was significantly higher with the normalized WWM-FA value (ROC_{AUC} 0.95; 95% confidence interval [CI], 0.91 to 0.98) than with the standard criteria for unfavorable outcome or other MRI sequences. In a subsequent validation cohort of 50 patients, a normalized WWM-FA value lower than 0.91, set from the derivation cohort, had a negative predictive value of 71.4% (95% CI, 41.9 to 91.6) and a positive predictive value of 100% (95% CI, 90.0 to 100), with 89.7% sensitivity (95% CI, 75.8 to 97.1) and 100% specificity (95% CI, 69.1 to 100).

Interpretation

In unconscious survivors 7 days after cardiac arrest, normalized WWM-FA value measured by DTI can accurately predict neurologic outcome at six months. This evidence requires confirmation from future large-scale trials with strict protocol of withdrawal or limitation-of-care decisions and time-window for MRI.

Funding French Ministry of Health, French National Agency for Research, Italian Ministry of health and Regione Lombardia.

Introduction

Prognostication of comatose resuscitated cardiac arrest (CA) patients is challenging, particularly during the first week when lingering effects of sedatives and neuromuscular blocking agents,[1] hypothermia,[2] and unstable physiologic status preclude detailed neurologic examinations.[3] Standard early predictors of unfavorable outcome after CA include absence of brain-stem reflexes, absence of motor response other than extensor, status myoclonus, high serum levels of neuron specific enolase (NSE), and absence of cortical responses by somatosensory evoked potentials.[4] All have substantial limitations in terms of reliability.[2] Recently, the publication of the Parisian Region Out of Hospital CA Registry (PROHCAR)[5] and the Save Hearts in Arizona Registry and Education (SHARE)[6] have raised some concerns about the existence of late awakeners and the consecutive risk of inappropriate early prognostication. Last published guidelines on CA recommend delaying prognostication after therapeutic hypothermia, and basing it on multiple prognostic tools.[7,8] Physicians are, therefore, confronted with a difficult decision-making process, mainly in patients who do not recover consciousness by day 7.[9]

Diffusion magnetic resonance imaging (MRI) is emerging as a promising prognostic tool.[10] One such measure is diffusion-weighted imaging (DWI) with the whole brain apparent diffusion coefficient, but the ideal time window for this technique is short (between days 3 and 5 after CA),[11,12] and has a low sensitivity (25% to 30%) for unfavorable prognosis, despite high specificity (95% to 100%).[11-13] An extension of DWI, diffusion tensor imaging (DTI), especially the calculation of fractional anisotropy (FA) allows *in vivo* quantification of white matter injuries that occur in acute, subacute or delayed fashion after global anoxia.[14,15] Another technique, proton magnetic resonance spectroscopy (^{1}H-MRS), allows one to quantify *in vivo* the concentrations of brain metabolites characterizing cellular dysfunction and neuronal loss. However, there is little evidence as to how these techniques might contribute to a better prognostication for outcome after CA.[16]

The goal of this study was to assess and validate the prognostic performance of whole brain white matter FA (WWM-FA), in comatose patients 7 days after CA as compared to clinical symptoms, morphological MRI and ^{1}H-MRS. Optimal cut-off will be defined from a derivation cohort and evaluated in an independent validation cohort.

Methods

Study design

This study is a multicenter, international, prospective, observational cohort study and part of a larger trial named MRI-COMA (assessing outcome with multimodal MRI of comatose patients of various origin). Institutional review board or Ethics committee approval was obtained for each country.

In the derivation cohort, patients were enrolled between October 2006 and June 2014 at 14 intensive care units (ICU) in France, Belgium, and Italy (appendix pp 2). Patients were eligible for inclusion if they were aged 18 years or more at the time of CA with persisting unconsciousness at day 7 defined as the inability to obey verbal commands not attributed to sedation or aphasia. The main exclusion criteria were: (i) contraindication to MRI; (ii) hemodynamic instability or respiratory failure precluding transport and scanning; and (iii) previous central nervous system disease (appendix pp 5). Written informed consent was obtained from patients' appointed proxies before MRI. The following year, between April 2015 and March 2016, we recruited the validation cohort on the same basis. A minimum of five healthy volunteers were also recruited at each center for the normalization procedure. Written informed consent was obtained from the healthy volunteers directly.

Standard criteria, OHCA, and neurologic evaluations

Demographic and clinical information was collected prospectively by participating centers with an Utstein-style[17] data form. Glasgow Coma Scale (appendix pp 28) was assessed at hospital admission, day of inclusion (day 7), and day of MRI acquisition. Out-of-hospital cardiac arrest (OHCA) score[18] was computed for each patient (appendix pp 29). An electroencephalogram (EEG) was recorded within 72 hours after return of spontaneous circulation using a 21-channel digital recorder according to the International 10-20 System. Blinded experienced neurophysiologists categorized at least 30 min of EEG data, according to the five Synek grades (appendix pp 30).[19,20]

Imaging acquisition, qualitative and quantitative image analysis

MRI acquisitions (appendix pp 7) were performed between day 7 and day 28 after CA on 15 scanners from three manufacturers: GE Medical Systems (Milwaukee, WI), Siemens Medical Solutions (Erlangen, Germany), and Philips Medical Systems (Eindhoven, The Netherlands). DTI, ^{1}H-MRS, and several conventional MRI sequences, including fluid-attenuated inversion recovery (FLAIR) and DWI, were acquired (appendix pp 18 for the precise parameters of each sequence from each center). Healthy volunteers underwent the same imaging protocol as that used for the patients. Anonymized magnetic resonance (MR) source images were transferred to the coordinating center for independent interpretation and reviewed by a board-certified neuroradiologist blinded to demographic, socioeconomic and clinical data. The FLAIR and DWI images were qualitatively scored[21] (appendix pp 31). All quantitative processes used for MRI data analysis, including the normalization procedure for DTI values and WWM-FA and whole-brain FA (WB-FA) calculation (appendix pp 19), are described in the appendix (pp 8 and 9). Importantly, the raw value of each derived diffusion measure, was divided by the mean of this measure across healthy control subjects acquired in the same scanner with the same sequence. Results are thus expressed as a percentage of controls. As FA can be normal or increased during cytotoxic edema,[22] patients were also assessed for whole brain mean average diffusion coefficient (WB-aDC). A low WB-aDC (<0.91) was determined in the control population as the mean value minus 3 standard deviation.

Outcome

Physicians in charge of the patients were blinded to ^{1}H-MRS and DTI information but had access to conventional MRI sequences (without FLAIR-DWI scoring) as well as to all clinical, biologic, and electrophysiologic data. All clinical decisions remained at the discretion of the treating team.

Neurologic performance was assessed at hospital discharge and at 6 months according to Glasgow-Pittsburgh Cerebral Performance Categories (CPC)[23] (appendix pp 32) and modified Rankin scale[24,25] (appendix pp 33). Best-achieved CPC was the primary outcome measure for stratifying patients into

favorable (CPC 1–2) or unfavorable outcome (CPC 3–5). For incapacitated patients, a personal consultee completed questionnaires on their behalf.

Statistical analysis

Assuming a specificity of 100% and a sensitivity of 75% of normalized WWM-FA as a reliable marker to predict unfavorable outcome, a 85% prevalence of unfavorable outcome and anticipating an incidence of 20% of exclusion criteria or technical issues,[16] a sample size of 185 patients was estimated[26] to obtain an adequate sensitivity with a 7.5% width of a 2-sided 95% confidence interval (CI). For the sample size calculation of the validation cohort, we considered the null hypothesis: AUC = 0.5 and we assumed the same prevalence of unfavorable outcome. The necessary sample size of 49 patients was calculated to achieve at least 95% statistical power at a type one error probability of 0.1% to detect AUC of 0.9.[27] Baseline characteristics were stratified according to the primary outcome based on CPC scores. Means and standard deviations were used for normal distributions of continuous variables, and medians and inter-quartile ranges (IQR) for non-continuous variables. For categorical variables, numbers and percentages were used. Differences between continuous variables were assessed by an unpaired two-sample t-test (normally distributed) or a Wilcoxon–Mann–Whitney U test (no assumption for distribution). Comparison between continuous variables from two groups were assessed by Fisher's exact test. Univariate logistic regression models were used to examine the predictive value of each variable. For each predictive variable ($P<0.05$), a receiver-operating-characteristic curve was plotted, and the corresponding area under the curve (ROC_{AUC}) was computed using the pROC package[28] for R software (the R Foundation for Statistical Computing). An optimal cutoff was defined as the value leading to the maximal sensitivity at 100% specificity. For this optimal cutoff, specificity, sensitivity, and positive and negative predictive values were computed, with 95% CI computed by bootstrap. The ROC_{AUC} and sensitivity of each variable were compared with those of the normalized WWM-FA, using the method described by DeLong et al.[29] and Fisher's exact test, respectively. All these statistics were also computed for three multivariate logistic-

regression models. For validation phase of the study, the cutoff points of normalized WWM-FA and WB-FA were not derived. Values above the cutoff point were considered indicative of a favorable outcome (CPC 1-2) and values below as being indicative of an unfavorable outcome (CPC 3-5). Predictive performance was assessed in the validation cohort by estimating negative and positive predictive values, sensitivity and specificity, and the ROC_{AUC}, with corresponding 95% CI. Additional details of statistical analyses, odds ratios, and logistic-regression models are available in the appendix (pp 11 and pp 12).

This study is registered with ClinicalTrials.gov, number NCT00577954.

Role of the funding source

The funding source had no role in study design, data collection, data analysis, data interpretation, or preparation of the report.

RESULTS

Of 185 patients who met the inclusion criteria, 150 were included in the analysis (Figure 1); 33 (22%) had a favorable neurologic outcome (12 with a best-achieved CPC score of 1 and 21 with a CPC of 2), of whom four eventually died within 6 months. Unfavorable outcome was reported in 117 patients: 102 (68%) died with a median survival time of 19 days (IQR 13 to 32 days), at 6 months, 10 (7%) had a score of CPC 4, and five (3%) had a score of CPC 3. CPC and Rankin scale scores, and cause of death are shown in Table S3 (appendix pp 21). Withdrawal or limitation-of-care decisions were taken for 72 patients (48%), 66 for neurological reasons, of whom 64 died and two eventually had a CPC score of 2 (appendix pp 23). Patients with a favorable outcome were younger (P=0.032), less likely to have diabetes mellitus (P=0.014), and more often had out of hospital CA (P=0.023), with a shockable first monitored rhythm (P<0.001), and more frequent myocardial infarction (P=0.001) (Table 1).

At baseline, there was no significant difference between patients with favorable or unfavorable outcome in terms of Glasgow Coma Scale, resuscitation time, or return of spontaneous circulation (Table 1). Favorable-outcome patients were less likely to have a motor response no better than extensor on the day of MRI (P<0.001), status myoclonus (P=0.013), and a higher OHCA score (P=0.040) (Table 2). Three (9%) of 33 favorable-outcome patients presented with at least two of the criteria of poor outcome according to the ERC/ESICM guidelines[30] (appendix pp 24).

In total, 135 (90%) EEG were interpretable. Median time to EEG was 3 days (IQR 2 to 5 days). Absence of EEG reactivity was observed in 12 patients with a favorable outcome, and 83 with an unfavorable outcome (P=0.003). Patients with a favorable outcome were more likely to have a favorable EEG pattern on Synek classification, but 62% of the patients were in the "uncertain" zone (Table 2).

Median time to MRI was not significantly different in unfavorable- and favorable-outcome patients (13 (IQR 10 to 18) *v.s.* 13 (IQR 10 to 18) days). Eighty-eight scans were obtained within the second week after the CA, 41 within the third week, and 21 after the third week. FLAIR or DWI images showed mild-to-

severe signal abnormalities in basal ganglia or cortex in 12 patients with a favorable outcome (appendix pp 16). Ninety patients with an unfavorable outcome had FLAIR or DWI hypersignal in basal ganglia and 60 patients in cortex. FLAIR-DWI scores were significantly higher (*i.e.* worse) in the unfavorable outcome group (Table 2). Figure 1 shows examples of FLAIR, DWI, and DTI images from a healthy control subject and three patients.

Normalized WWM-FA and WB-FA by DTI were significantly lower in unfavorable- compared with favorable-outcome patients (0.78±0.13, 0.82±0.11 *v.s.* 0.96±0.03, 0.99±0.05 respectively, see Table 2). Figure S2 (appendix pp 13) shows a scatter plot of normalized WWM-FA *v.s.* WB-aDC for all patients and healthy volunteers. Twenty-five patients presented with severe reductions in mean normalized WB-aDC (<0.91), all between days 7 and 14 after the CA, none of whom had a favorable outcome at 6 months. In total, 131 (87%) ^{1}H-MRS spectra were interpretable. The NAA/Cr ratios in thalami and pons were significantly lower in the unfavorable- *v.s.* favorable-outcome patients (1.09±0.33, 1.88±0.44 *v.s.* 1.53±0.29, 2.31±0.38 respectively, see Table 2).

Figure 3 (A-C) summarizes the results of ROC curves of the principal predictors of unfavorable outcome. Table 3 shows the cutoff value predictive of unfavorable outcome with a specificity of 100%, the sensitivity value corresponding to this cutoff value and ROC_{AUC} for each predictor. The prognostic accuracy, as quantified by the ROC_{AUC}, was significantly higher with the normalized WWM-FA than with standard criteria for unfavorable outcome or the other MRI sequences.

The optimal cutoff value of normalized WWM-FA was <0.91, with a specificity of 100% (sensitivity 89%). The same cutoff value and similar specificity and sensitivity were obtained in adjusted analyses of the subpopulation excluding patients with withdrawal or limitation of care (appendix pp 25). For the whole population, after excluding the patients presenting a diffuse cytotoxic edema as defined by a low WB-aDC, the ROC_{AUC} for normalized WWM-FA improved to 0.99 (95% CI, 0.97 to 1.00), and sensitivity up to 0.95 (95% CI, 0.88 to 0.98). Figure S3 (appendix pp 15) shows the logistic regression estimated curve of

normalized WWM-FA for unfavorable outcome prediction excluding patients with low aDC. The probability (95% CI) of unfavorable outcome with normalized WWM-FA values of ≤0.85, was 100% (99.8 to 100), and the probability (95% CI) of favorable outcome with a normalized WWM-FA value >0.95 was 95% (87 to 100).

Multivariate analyses are shown in Table 3 and specific results in the appendix (pp 26 and pp 27). In the first model, significant selected variables were Synek classification of EEG and motor response no better than extensor at the day of MRI. The ROC_{AUC} was 0.84 (95% CI, 0.76 to 0.93). The second model, which included the FLAIR-DWI overall score, increased the ROC_{AUC} to 0.90 (95% CI, 0.85 to 0.96). Finally, the third model, which included normalized WWM-FA, increased the ROC_{AUC} to 0.98 (95% CI, 0.96 to 1.00), with a specificity of 100% (95% CI, 89% to 100%) and a sensitivity of 82% (95% CI, 73% to 88%).

In the validation cohort, of 58 patients who met the inclusion criteria, 50 were included in the analysis (Figure 1), 11 (22%) had a favorable neurologic outcome (1 with a best-achieved CPC score of 1 and 10 with a CPC of 2), of whom one eventually died within 6 months. Unfavorable outcome was reported in 39 patients: 29 (74.4%) died with a median survival time of 19 days (IQR 16 to 45 days), at 6 months, 6 (15%) had a score of CPC 4, and 4 (10%) had a score of CPC 3. Rankin scale scores, and cause of death are shown in Table S3 (appendix pp 21). A normalized WWM-FA value lower than 0.91 had a negative predictive value of 71.4% (95% CI, 41.9 to 91.6) and a positive predictive value of 100% (95% CI, 90.0 to 100), with 89.7% sensitivity (95% CI, 75.8 to 97.1) and 100% specificity (95% CI, 69.1 to 100).

DISCUSSION

The present study focused on a subset of CA patients who did not regain consciousness at day 7. This is, in our opinion, the patient group for whom major ethical issues arise, and in whom systematic withdrawal of care would definitely alter the likeliness of late awakening in a subset of patients. In our series, 22% of patients not responding to simple orders by day 7 after CA had a CPC 1 or 2 at 6 months. The normalized WWM-FA measurement requires standardized post-processing steps from nearly conventional MRI sequences (DTI and 3D T1-weighted images), that can be safely acquired during the second week after CA, and is more accurate than other measures.

The last guidelines have recommended delaying CA outcome prognostication for at least 72 hours after rewarming,[7,8] which indeed corresponds to day 5 to 7 after CA in most cases. However, "late-awakeners" after a CA have been described as late as 25 days after arrest,[31-33] Our data clearly show that a proportion of patients can still recover after this period of time, despite a lack of response to simple orders after day 7. For these "late-awakeners" remaining unresponsive at day 7, our results also indicate that evaluation by means of clinical symptoms, EEG and conventional MRI are not specific or sensitive enough to be clinically usable.

As early as 2001, FLAIR[34] and DWI images were shown to add prognostic value to the clinical examination and computed tomography. Hirsch et al. reported that a combination of FLAIR and DWI lesion counts leads to a sensitivity of 80% for a specificity of 100% in predicting unfavorable outcome, but in a population which included conscious patients scanned before day 8 after CA.[21] In our selected population, however, the FLAIR-DWI overall score had a sensitivity of only 40% for a specificity of 100%. This lower sensitivity might be explained by the inclusion of only unresponsive patients on one hand, and because of a difference in the delay of MRI on the other hand. Indeed, in our selected population some survivors had a favourable neurological outcome (especially patients with a CPC 2) despite mild-to-severe signal abnormalities in basal ganglia or cortex. Our finding comforts the results of a large

retrospective multicentre study, where 15% of patients with favourable outcome had DWI signal abnormalities on MRI.[35] Our study also showed that thalamic and pons ^{1}H-MRS had a predictive accuracy similar to FLAIR and DWI scores.

The dramatic decrease in normalized WWM-FA observed in survivors with unfavorable outcome can be linked to the occurrence of delayed white matter injuries after CA.[14] Recently, using DTI, Laitio *et al.*,[36] confirmed that demyelination occurred in comatose survivors of cardiac arrest and was mainly responsible for the decrease in FA. In our study, the clinically relevant values of normalized WWM-FA are 0.85, a threshold below which the probability of unfavorable outcome is close to 100%, and 0.95, a threshold above which the likelihood of favorable outcome is above 95%. Finally, we observed low WB-aDC in 25 (16%) of patients, a feature specific to MRI performed between 7 and 14 days after CA in patients with diffuse cytotoxic edema. None of these patients had a favorable outcome, and 32% had FA above the threshold of 0.91, resulting in false negative scores if we consider normalized WWM-FA only. FA depends on the ratio between axial and radial diffusivity, which are both decreased in low aDC syndrome, explaining why FA measures can be unreliable in this context. Another potential limitation of the technique is the requirement for transportation of a ventilated patient to the MRI suite for 20 to 40 min, with total immobilization, including potential use of sedation or even neuromuscular blockers. Currently, the FA measure also needs to be standardized for each scanner.

Our findings should be interpreted in light of the following potential limitations. First, we cannot exclude the possibility of unmeasured confounding factors in this observational study. Second, our validation and derivation cohorts were consecutive rather than simultaneous. Therefore, we cannot eliminate a change in clinical care during the two periods. However, it is very unlikely that it would be of such an extent to change prognostic and even though, it should not affect the relationship between anatomical changes of the white matter as assessed by normalized WWM-FA and outcome at six months. Third, we based our analysis on comparison of DTI with tools validated for an earlier prognostication. However, our point is

that the clinical issue of prognostication arises in patients unconscious after 7 days not matching the usual criteria of poor outcome (grey zone patients). Another limitation is our inclusion criteria and the timing of MRI, such that our results cannot be applied to MRI performed before Day 7. Finally, we cannot exclude misclassification of outcome due to "self-fulfilling" prophecies arising from premature decisions of WLST. However, even if we exclude all patients who died from care withdrawal, the cutoff for normalized WWM-FA and AUC did not vary significantly (appendix pp 25).

In conclusion, our data indicate that quantitative MRI, specifically normalized WWM-FA derived from DTI, is a tool that can be used to predict outcome with a high degree of accuracy in patients with altered consciousness at day 7 after CA. This evidence requires confirmation from future large-scale trials with strict protocol of withdrawal or limitation-of-care decisions and time-window for MRI.

Contributors

LP was the lead intensivist and contributed to the protocol development, study design and concept, conduct, data acquisition and coordination, data interpretation, writing and revising it critically, and reviewing the report. He takes responsibility for the integrity of the data and the accuracy of the data analysis. LV was involved in in study design, concept, conduct, the coordination and supervision of data collection, data interpretation, DTI pipeline development, writing and coordinating drafts of the report and reviewing the report. VP was involved in study design and conduct, DTI pipeline development, data interpretation and quality check, writing and reviewing the report. TB was involved in data analysis, contribution to the statistical analysis plan, statistical analysis, data interpretation, editing of the report and approved the final report as submitted. SB was involved in data interpretation, DTI pipeline development and quality check. DG, HB, were involved in study design and concept and revising the report critically. NA, GT and VB were involved in data collection and analysis, contribution to the statistical analysis plan, and preparation of the report. AV, BJ, BRo, CA, CDP, CEL, LN, NB, NG, OC, RC, SS, TT, VC were involved in study conduct, data acquisition, and revising the report. BRi and RG were involved in data interpretation, and revising the report critically. SL and GC were involved in study conduct, data acquisition and coordination, and revising the report critically. All authors approved the final version.

REFERENCES

1. Samaniego EA, Mlynash M, Caulfield AF, Eyngorn I, Wijman CA. Sedation confounds outcome prediction in cardiac arrest survivors treated with hypothermia. *Neurocritical care* 2011; **15**(1): 113-9.
2. Kamps MJ, Horn J, Oddo M, et al. Prognostication of neurologic outcome in cardiac arrest patients after mild therapeutic hypothermia: a meta-analysis of the current literature. *Intensive care medicine* 2013; **39**(10): 1671-82.
3. Ben-Hamouda N, Taccone FS, Rossetti AO, Oddo M. Contemporary approach to neurologic prognostication of coma after cardiac arrest. *Chest* 2014; **146**(5): 1375-86.
4. Rossetti AO, Rabinstein AA, Oddo M. Neurological prognostication of outcome in patients in coma after cardiac arrest. *Lancet Neurol* 2016.
5. Paul M, Bougouin W, Geri G, et al. Delayed awakening after cardiac arrest: prevalence and risk factors in the Parisian registry. *Intensive care medicine* 2016; **42**(7): 1128-36.
6. Irisawa T, Vadeboncoeur TF, Karamooz M, et al. Duration of Coma in Out-of-Hospital Cardiac Arrest Survivors Treated With Targeted Temperature Management. *Ann Emerg Med* 2017; **69**(1): 36-43.
7. Cronberg T, Brizzi M, Liedholm LJ, et al. Neurological prognostication after cardiac arrest--recommendations from the Swedish Resuscitation Council. *Resuscitation* 2013; **84**(7): 867-72.
8. Sandroni C, Cariou A, Cavallaro F, et al. Prognostication in comatose survivors of cardiac arrest: an advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. *Resuscitation* 2014; **85**(12): 1779-89.
9. Sandroni C, Cariou A, Cavallaro F, et al. Prognostication in comatose survivors of cardiac arrest: an advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. *Intensive care medicine* 2014; **40**(12): 1816-31.
10. Els T, Kassubek J, Kubalek R, Klisch J. Diffusion-weighted MRI during early global cerebral hypoxia: a predictor for clinical outcome? *Acta neurologica Scandinavica* 2004; **110**(6): 361-7.
11. Mlynash M, Campbell DM, Leproust EM, et al. Temporal and spatial profile of brain diffusion-weighted MRI after cardiac arrest. *Stroke; a journal of cerebral circulation* 2010; **41**(8): 1665-72.
12. Wijman CA, Mlynash M, Caulfield AF, et al. Prognostic value of brain diffusion-weighted imaging after cardiac arrest. *Annals of neurology* 2009; **65**(4): 394-402.
13. Greer DM, Scripko PD, Wu O, et al. Hippocampal magnetic resonance imaging abnormalities in cardiac arrest are associated with poor outcome. *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association* 2013; **22**(7): 899-905.
14. Gerdes JS, Walther EU, Jaganjac S, Makrigeorgi-Butera M, Meuth SG, Deppe M. Early detection of widespread progressive brain injury after cardiac arrest: a single case DTI and post-mortem histology study. *PloS one* 2014; **9**(3): e92103.
15. Newcombe VF, Williams GB, Scoffings D, et al. Aetiological differences in neuroanatomy of the vegetative state: insights from diffusion tensor imaging and functional implications. *Journal of neurology, neurosurgery, and psychiatry* 2010; **81**(5): 552-61.
16. Luyt CE, Galanaud D, Perlbarg V, et al. Diffusion tensor imaging to predict long-term outcome after cardiac arrest: a bicentric pilot study. *Anesthesiology* 2012; **117**(6): 1311-21.
17. Jacobs I, Nadkarni V, Bahr J, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: update and simplification of the Utstein templates for resuscitation registries: a statement for healthcare professionals from a task force of the International Liaison Committee on Resuscitation (American Heart Association, European Resuscitation Council, Australian Resuscitation Council, New Zealand Resuscitation Council, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Councils of Southern Africa). *Circulation* 2004; **110**(21): 3385-97.

18. Adrie C, Cariou A, Mourvillier B, et al. Predicting survival with good neurological recovery at hospital admission after successful resuscitation of out-of-hospital cardiac arrest: the OHCA score. *Eur Heart J* 2006; **27**(23): 2840-5.
19. Synek VM. Prognostically important EEG coma patterns in diffuse anoxic and traumatic encephalopathies in adults. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 1988; **5**(2): 161-74.
20. Synek VM. EEG abnormality grades and subdivisions of prognostic importance in traumatic and anoxic coma in adults. *Clinical EEG* 1988; **19**(3): 160-6.
21. Hirsch KG, Mlynash M, Jansen S, et al. Prognostic Value of A Qualitative Brain MRI Scoring System After Cardiac Arrest. *J Neuroimaging* 2014.
22. Sotak CH. The role of diffusion tensor imaging in the evaluation of ischemic brain injury - a review. *NMR Biomed* 2002; **15**(7-8): 561-9.
23. Randomized clinical study of thiopental loading in comatose survivors of cardiac arrest. Brain Resuscitation Clinical Trial I Study Group. *The New England journal of medicine* 1986; **314**(7): 397-403.
24. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scottish medical journal* 1957; **2**(5): 200-15.
25. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke; a journal of cerebral circulation* 1988; **19**(5): 604-7.
26. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996; **3**(9): 895-900.
27. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; **148**(3): 839-43.
28. 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.
29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**(3): 837-45.
30. Nolan JP, Soar J, Cariou A, et al. European Resuscitation Council and European Society of Intensive Care Medicine 2015 guidelines for post-resuscitation care. *Intensive care medicine* 2015; **41**(12): 2039-56.
31. Rittenberger JC, Popescu A, Brenner RP, Guyette FX, Callaway CW. Frequency and timing of nonconvulsive status epilepticus in comatose post-cardiac arrest subjects treated with hypothermia. *Neurocritical care* 2012; **16**(1): 114-22.
32. Al Thenayan E, Savard M, Sharpe M, Norton L, Young B. Predictors of poor neurologic outcome after induced mild hypothermia following cardiac arrest. *Neurology* 2008; **71**(19): 1535-7.
33. Greer DM. Unexpected good recovery in a comatose post-cardiac arrest patient with poor prognostic features. *Resuscitation* 2013; **84**(6): e81-2.
34. Wijdicks EF, Campeau NG, Miller GM. MR imaging in comatose survivors of cardiac resuscitation. *AJNR American journal of neuroradiology* 2001; **22**(8): 1561-5.
35. Ryoo SM, Jeon SB, Sohn CH, et al. Predicting Outcome With Diffusion-Weighted Imaging in Cardiac Arrest Patients Receiving Hypothermia Therapy: Multicenter Retrospective Cohort Study. *Crit Care Med* 2015; **43**(11): 2370-7.
36. Laitio R, Hynninen M, Arola O, et al. Effect of Inhaled Xenon on Cerebral White Matter Damage in Comatose Survivors of Out-of-Hospital Cardiac Arrest: A Randomized Clinical Trial. *JAMA* 2016; **315**(11): 1120-8.

FIGURE LEGENDS

Figure 1: CONSORT flow chart of the derivation and validation cohorts. Assessment, analysis populations, and follow-up of the patients in the MRI-COMA trial.

Figure 2. Maps of fluid-attenuated inversion recovery (FLAIR) magnetic resonance imaging (MRI) and color-coded raw values of whole white matter fractional anisotropy (WWM-FA) map of a 60-year-old healthy control and three cardiac arrest (CA) patients, performed with the same scanner. Regions of low WWM-FA are depicted as a blue and regions of high WWM-FA are in red. Patient 1: 21-year-old man who had a prolonged out-of-hospital cardiac arrest (OHCA) before being put on extracorporeal life support. The OHCA score was 56 (initial recorded rhythm, ventricular fibrillation; no flow, 0.5 min; low flow, 170 min; serum creatinine, 275 μmol per liter; arterial lactate 26 mmol per liter) with an estimated mortality of 98%. The protein S100B was elevated (5.91 μg per liter) at day 2. The conventional MRI at day 12 found acute lesions in the basal ganglia and occipital cortex, with an FLAIR-DWI overall score of 33. However, the normalized WWM-FA was preserved at 0.92. The patient left the intensive care unit after 25 days (Cerebral Performance Categories (CPC) of CPC of 3). At 6 months, he was still alive with dysarthria but he was fully independent in daily life (CPC of 2; Rankin 3). Patient 2: 37-year-old man with a ventricular fibrillation associated with congenital dilated cardiomyopathy. The conventional MRI at day 7 found acute lesions in the parietal and occipital cortex, with an FLAIR-DWI overall score of 39. However, the normalized WWM-FA was preserved at 0.95. The patient left the intensive care unit after 17 days (CPC of 3). At 6 months, he was still alive with motor spasticity but he was fully independent in daily life (CPC 2; Rankin 3). Patient 3: 37-year-old man with in-hospital CA. The OHCA score was equal to 20 (initial recorded rhythm, ventricular fibrillation; no flow, 0.5 min; low flow, 25 min; serum creatinine, 660 μmol per liter; arterial lactate 2.8 mmol per liter). The protein S100B was elevated (1.70 μg per liter) at day 2. The conventional MRI at day 12 found acute lesions in the basal ganglia and occipital cortex, with an FLAIR-DWI

overall score of 40. The normalized WWM-FA was dramatically decreased at 0.66 and the patient eventually died.

Figure 3. Receiver-Operating-Characteristic Curves for standard criteria (Panel A), qualitative magnetic qualitative resonance imaging (MRI; Panel B), and quantitative MRI biomarkers (Panel C) for unfavorable outcome in the derivation cohort. DNG denotes Deep Gray Nuclei; DTI, Diffusion Tensor Imaging; DWI, Diffusion Weighted Imaging; EEG, Electroencephalography; FA, Fractional Anisotropy; FLAIR, Fluid-Attenuated Inversion Recovery; ^{1}H-MRS, Proton Magnetic Resonance Spectroscopy; NAA/Cr, N-acetyl aspartate over creatinine ratios. OHCA score calculation, FLAIR-DWI scoring system and EEG Synek classification are described in additional files 2, 3 and 4 of the appendix, respectively.

Evidence before this study

We searched MEDLINE for reports on the use of diffusion measures on magnetic resonance imaging (MRI) in cardiac arrest (CA) published up to November 2017 with the following search terms for the patient category of CA: “heart arrest”, “cardiac arrest”, “cardiopulmonary resuscitation”, “ischemic-hypoxic encephalopathy”, “hypoxia-ischemia” or “post anoxic coma’; the following search terms for the diffusion category: "magnetic resonance", “MRI”, "MR", “neuroimaging”, “apparent diffusion coefficient”, "diffusion-weighted imaging", "diffusion tensor imaging"; and search terms for outcome: “prediction”, “predictors”, "prognosis", “prediction model”, or “outcome”. The search yielded 21 observational cohorts of which 17 reported outcomes. Eight studies predominantly examined apparent diffusion coefficient (ADC), 7 diffusion-weighted imaging (DWI) and 2 diffusion tensor imaging (DTI). Except one, all the studies were based on MRI acquired during the first week after cardiac arrest. Diffuse DWI abnormalities were highly specific but only modestly sensitive of poor outcome. Being a qualitative technique, DWI is prone to interobserver variability, but it can be standardized using semi-quantitative methods like ADC. Poor outcome patients exhibited a nadir in ADC values at 3–5 days after CA, which therefore appeared to be the optimal time window for prognostication using ADC. Thresholds have been determined either regionally or globally. One study reported prognostic thresholds for percentage of brain volume below specific values of ADC. However, this study was performed in a small cohort (51 patients) and was monocentric. Indeed, all published multicenter studies on ADC had to use semi-quantitative analysis since ADC metrics highly depend on diffusion acquisition parameters. Two particular bicentric studies reported the use of DTI derived parameters. Outcome was not the primary endpoint in the one including 97 patients in which data were not normalized. The second one from our group proposed a predictive composite score derived on 57 patients that was not externally validated and did not show direct clinical transferability.

Added value of this study

To our knowledge, this is the largest, prospectively followed cohort of CA patients reporting the predictive value of fractional anisotropy (FA), derived from DTI, in a multicenter cohort of patients who were still comatose 7 days after CA. Our findings provide evidence that normalized whole white matter FA (WWM-FA) can be useful in prediction of poor outcome. WWM-FA showed the highest predictive power with an AUC of 0.95 (95% CI, 0.91 to 0.98), over-performing FLAIR-DWI scores and proton magnetic resonance spectroscopy. WWM-FA less than 0.91 of controls predicted poor outcome, with 0% FPR (95% CI, 0 to 14%). Interestingly, we observed FLAIR-DWI abnormalities in 12 out of 36 patients with a good outcome. This study also resolved two methodological issues: the requirement of measurements calibration in each center by acquiring healthy controls to establish measurements of reference and the dramatic need of extensive quality check of MRI acquisitions, that led to 19% exclusion rate in our cohort.

Implications of all the available evidence

Our results are relevant in the clinical setting because they might provide reliable outcome predictors and could possibly improve diagnosis of late awakeners in survivors after CA who still unresponsive to simple orders after 7 days. The findings of our study support the use of quantitative MRI (DTI) for proxies information and management of care withdrawal decisions in this selected population of CA patients.