



Increased task-relevant fMRI responsiveness in comatose cardiac arrest patients is associated with improved neurologic outcomes

Kiran Dhakal^{1,*} , Eric S Rosenthal^{2,*},
Annelise M Kulpanowski¹, Jacob A Dodelson¹, Zihao Wang¹,
Gaston Cudemus-Deseda³, Marjorie Villien¹, Brian L Edlow^{1,2},
Alexander M Presciutti⁴, James L Januzzi⁵, MingMing Ning²,
W Taylor Kimberly², Edilberto Amorim²,
M Brandon Westover⁶, William A Copen⁷,
Pamela W Schaefer⁷, Joseph T Giacino⁸, David M Greer⁹ and
Ona Wu¹ 

Abstract

Early prediction of the recovery of consciousness in comatose cardiac arrest patients remains challenging. We prospectively studied task-relevant fMRI responses in 19 comatose cardiac arrest patients and five healthy controls to assess the fMRI's utility for neuroprognostication. Tasks involved instrumental music listening, forward and backward language listening, and motor imagery. Task-specific reference images were created from group-level fMRI responses from the healthy controls. Dice scores measured the overlap of individual subject-level fMRI responses with the reference images. Task-relevant responsiveness index (Rindex) was calculated as the maximum Dice score across the four tasks. Correlation analyses showed that increased Dice scores were significantly associated with arousal recovery ($P < 0.05$) and emergence from the minimally conscious state (EMCS) by one year ($P < 0.001$) for all tasks except motor imagery. Greater Rindex was significantly correlated with improved arousal recovery ($P = 0.002$) and consciousness ($P = 0.001$). For patients who survived to discharge ($n = 6$), the Rindex's sensitivity was 75% for predicting EMCS ($n = 4$). Task-based fMRI holds promise for detecting covert consciousness in comatose cardiac arrest patients, but further studies are needed to confirm these findings. Caution is necessary when interpreting the absence of task-relevant fMRI responses as a surrogate for inevitable poor neurological prognosis.

Keywords

Cardiac arrest, coma, consciousness, functional MRI, quantitative imaging biomarker

Received 14 April 2023; Revised 27 June 2023; Accepted 29 June 2023

¹Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA, USA

²Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

³Department of Cardiac Anesthesiology and Critical Care Medicine, Massachusetts General Hospital, Boston, MA, USA

⁴Department of Psychiatry, Center for Health Outcomes and Interdisciplinary Research, Massachusetts General Hospital, Boston, MA, USA

⁵Department of Medicine, Cardiology Division, Massachusetts General Hospital and Baim Institute for Clinical Research, Boston, MA, USA

⁶Beth Israel Deaconess Medical Center, Boston, MA, USA

⁷Department of Radiology, Neuroradiology Division, Massachusetts General Hospital, Boston, MA, USA

⁸Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Harvard Medical School, Charlestown, MA, USA

⁹Department of Neurology, Boston University School of Medicine, Boston Medical Center, Boston, MA, USA

*These authors contributed equally to this work.

Corresponding author:

Ona Wu, Athinoula A. Martinos Center for Biomedical Imaging, 149 13th Street, CNY 2301, Charlestown, MA 02129, USA.
Email: ona.wu@mgh.harvard.edu

Introduction

Cardiac arrest is a public health crisis associated with high mortality and severe neurologic morbidity among survivors,¹ many of whom fail to regain consciousness due to anoxic brain injury.^{2–4} Neurocognitive dysfunction following cardiac arrest can be progressive and challenging for prognostication, especially in patients who do not show early neurological responses.^{4,5} Inaccurate prognostication may lead to inappropriate premature withdrawal of life-sustaining treatments (WLST) in patients who might otherwise have favorable neurologic outcomes.^{6,7} In recent years, magnetic resonance imaging (MRI), particularly task-based functional MRI (fMRI),⁸ has been recognized as a potential imaging biomarker to be used in combination with other clinical measures for prognostication and outcome prediction of comatose cardiac arrest patients in acute settings.^{9–12} However, the accuracy of utilizing task-based fMRI in predicting recovery of consciousness on an individual patient basis has not been adequately investigated.

We conducted a prospective pilot study to explore the feasibility and utility of task-based fMRI among comatose cardiac arrest patients in acute care settings. In addition, we developed an objective metric to quantify the degree of task-relevant responsiveness on an individual patient basis that can ultimately be used to detect covert consciousness. We compared responses between patients who achieved signs of arousal recovery (eye-opening to noxious or verbal stimulation), patients without arousal recovery, and healthy controls. We also investigated associations of these early metrics with long-term functional outcomes, with good long-term functional outcomes defined as achieving emergence from the minimally conscious state¹³ (EMCS, i.e., ability to reliably communicate yes or no responses or use familiar objects in a functional manner) as measured on the Coma Recovery Scale-Revised (CRS-R), a widely used disorders of consciousness research measure.¹⁴

Materials and methods

Participants

Data collection and analysis were performed under approval by the Partners Human Research Committee (protocol # 2013P001795) which operates in compliance with all applicable federal, state, and local laws and regulations and with the Federalwide Assurances and is guided by the “Belmont Report.”¹⁵ All participants or their legally authorized representative provided

written informed consent. Twenty-two adult cardiac arrest patients who remained comatose after the restoration of spontaneous circulation (ROSC) were prospectively enrolled between December 2013 and April 2018 if meeting the following inclusion criteria: (i) age ≥ 18 years, (ii) post-ROSC Glasgow Coma Scale (GCS) score ≤ 8 , (iii) no MRI contraindications, (iv) no pre-arrest neurocognitive disorders including dementia, and (v) availability of a legally authorized representative to provide written informed consent. Since this was a pilot study, formal sample size calculation was not performed and patients were not randomized. A detailed description of protocol amendments is provided in the Supplementary Methods. Our study included non-English speakers to allow for a diverse patient population. A patient’s ability to speak English was determined based on information from the family and the patient’s medical record. A cohort of five healthy control (HC) English-speaking adults (age ≥ 18 years) with no history of either cognitive disorders or MRI contraindications was also enrolled. Informed written consent was obtained prior to the performance of all research procedures. As part of the study protocol, with the agreement of the legally authorized representative at enrollment, the patient-level fMRI results were allowed to be shared with the clinical team. When sharing results, potential confounds, such as patient motion or sedation, were discussed with the clinical team.

Neurological testing and outcome assessments

Demographics, clinical characteristics (e.g., initial cardiac rhythm, hospital admission GCS), and targeted temperature management treatment details were extracted from the patients’ medical records. The initial arrest was classified as either shockable (ventricular tachycardia or ventricular fibrillation) or non-shockable (pulseless electrical activity or asystole).

Favorable short-term neurological outcome was defined by arousal recovery (AR), i.e., eye-opening to stimulation, by hospital discharge. Modified Rankin Scale (mRS) and Cerebral Performance Category (CPC) scores were obtained at discharge, and 3-months, 6-months, and 1-year post-cardiac arrest. Missing values for CPC and mRS scores were carried forward from previous visits. Early WLST was defined as dying from WLST within seven days after cardiac arrest. For survivors, we assessed whether patients achieved at least EMCS as measured by the CRS-R¹⁴ (score of 2 on the Communication Scale or score of 6 on the Motor Function Scale) at in-person follow-up visits. Good long-term functional outcome was defined as achieving

EMCS within one year after cardiac arrest as measured on the CRS-R scale.

Functional MRI data acquisition and task paradigms

MRI was acquired with a 32-channel head coil or 20-channel head-neck coil on a 3-Tesla scanner (Skyra, Siemens Medical Solution). Task-based fMRI was collected using 40 gradient-echo echo-planar imaging measurements with the following parameters: echo time (TE) = 25 ms, repetition time (TR) = 3000 ms, flip-angle = 90°, field-of-view (FOV) = 192 mm, matrix = 94 × 94, 2x acceleration and 49 (32-channel head coil) or 46 (20-channel head-neck coil) interleaved slices with a thickness of 2 mm and slice distribution factor of 25% and total acquisition time of 2:09. 3D T1-weighted images (T1WI) were acquired for anatomical references and registration purposes using a multi-echo magnetization-prepared rapid gradient-echo (MEMPRAGE) sequence¹⁶ with TR = 2230 ms; TE = 1.69, 3.55, 5.41, and 7.27 ms, flip-angle = 7°, FOV = 256 mm, matrix = 256 × 256, 2–3x acceleration, TI = 1100–1300 ms and 176 sagittal slices of 1 mm thickness. Five patients were scanned using an MEMPRAGE sequence with built-in motion correction.¹⁷ This sequence was not available for all patients due to scanner upgrades.

Task-based fMRI involved three tasks consisting of alternating 24 s blocks of audio and rest, with two runs for each task. Tasks included listening to instrumental music, listening to spoken language (forward and backward), and following commands through motor imagery (imagine squeezing the right hand). All task stimuli were contrasted against silence, except motor imagery which was contrasted against task-off. Details of the task-based fMRI paradigm have been previously described.⁸ Pre-recorded verbal instructions were played before each task's presentation and start of the fMRI sequence.⁸ The scanner's sound system delivered all auditory signals via MRI-compatible earphones (Newmatic Medical; Caledonia, MI). The same audio files were played to healthy controls and all patients.

Data analysis

Task-based fMRI analysis. Task-based fMRI analyses were performed with the FMRI Expert Analysis Tool (FEAT) Version 6.00, part of FMRIB's Software Library (FSL, www.fmrib.ox.ac.uk/fsl) using techniques previously described.⁸ The 3D T1WI brain was extracted using the Robust Brain Extraction (<https://www.nitrc.org/projects/robex/>)¹⁸ tool, the time-averaged fMRI brain was extracted using the Brain Extraction Tool¹⁹ from FSL, and the resultant

skull-stripped images were used for registration. The time-averaged fMRI data were registered to the 3D T1WI using the boundary-based registration algorithm.²⁰ The 3D T1WI was registered to standard space using FLIRT.^{21,22} The following standard FEAT preprocessing pipeline was applied: motion correction,²¹ slice-timing correction, spatial smoothing (Gaussian kernel full-width half-maximum = 10 mm), grand-mean intensity normalization, and high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 24.0 s). Confound matrices were calculated (fsl_motion_outliers tool from FSL, <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLMotionOutliers>) to account for the effects of excessive motion not compensated for by motion correction. Time-series statistical analysis of individual runs was performed using FMRIB's Improved Linear Model with local autocorrelation correction.²³ The model included fMRI task variables, corresponding contrasts, temporal derivatives, six standard motion parameters, and confound matrices.

The results of the two individual runs for each task were combined using a fixed effects model with FMRIB's Local Analysis of Mixed Effects (FLAME).^{24–26} Group-level comparisons were similarly executed using a fixed effects model with FLAME in FSL.^{24–26} The sample size of the study was insufficient to use mixed-effects models. All Z-statistic images were thresholded using clusters determined by $Z > 3.1$ and corrected cluster significance threshold of $P = 0.05$ ²⁷ to create activation maps. The statistical threshold for cluster significance ($Z > 3.1$) and the size of the Gaussian kernel used for smoothing were selected to decrease false positive cluster activations.²⁸ We also explored Activation Mapping as a Percentage of Local Excitation (AMPLE) methods.²⁹ Comparison of AMPLE to non-AMPLE results was performed by an investigator blinded to patient outcomes. However, the findings were not statistically significantly different compared to non-AMPLE normalized methods (see Supplementary Methods and Supplementary Results). Therefore, we limited analyses to the group-level Z-statistic activation maps for the remainder of the study. Group-level analysis between HC and patients was repeated, excluding the non-English-speaking patients, to assess the influence of non-English-speaking patients on the group-level language task-based fMRI results.

All significant brain responses are shown as Z-statistic images thresholded at cluster-corrected Z-statistic values of 3.1 with $P < 0.05$ and overlaid on MNI 152 T1-weighted 2 mm image at MNI coordinates corresponding to the location of the HC cluster maxima for each task: Music (56, –20, 12), Forward

language (−54, −20, 6), Backward language (58, −14, 8), and motor imagery (−38, 48, 18). Color bars represent the Z-statistic value. Individual subject-level fMRI analyses were performed by an investigator blinded to arousal recovery status and long-term outcomes.

Quantifying task-relevant fMRI responses on an individual patient basis. To quantify the degree of task-relevant fMRI responses of individual subjects, we measured the spatial overlap of significant voxels in subject-wise brain activation maps compared to reference images using the Dice score:

$$Dice = \frac{2 \times V_{overlap}}{V1 + V2}$$

for which $V1$ is the number of voxels in the reference image, $V2$ is the number of voxels in the individual subject's activation map, and $V_{overlap}$ is the number of overlapping voxels in both images. The Dice score ranges from 0 to 1, with 0 indicating no overlap and 1 indicating complete overlap. The Dice score has been used by other studies to measure the reproducibility of fMRI responses.^{30–34} We calculated a responsiveness index (Rindex) as the maximum Dice score over the four tasks to measure overall task-relevant responsiveness. We chose this metric to maximize sensitivity for detecting indications of covert consciousness while being able to reflect the strength of responsiveness as a continuous variable.

For the reference image, we evaluated two options: (1) an anatomical reference image and (2) group-level activation maps (Z-statistics) from healthy controls. Anatomical reference images were generated by combining an *a priori* defined set of regions of interest (ROIs) based on previous fMRI studies of music, language, and motor imagery (See details in Supplementary Methods). ROIs were selected on a task-wise basis (Supplementary Table 1) and combined to generate a single binarized task-specific reference image (Supplementary Fig. 1). Binarized group-level brain activation maps (Z-statistics) of healthy participants were used to define functional reference images (Supplementary Fig. 2). The reference image that performed best in terms of the Dice score for healthy controls (2-tailed paired t-tests) was then used for the remainder of the analysis. Choice of reference image was made by an investigator blinded to patient outcomes.

Statistical analyses. Statistical analyses were performed using JMP 16.1 (SAS Institute Inc., Cary, NC, USA). P-values <0.05 were considered statistically significant. Univariable analyses were performed for categorical

variables using the Fisher's Exact Test, and for continuous variables, the Wilcoxon Rank Sum two-sample Test with normal approximation. For each task, Dice scores for the task-relevant responses were compared between HC and all patients. Differences in Dice scores between the following patient subgroups were evaluated: sedated versus non-sedated, intubated versus non-intubated, and early-WLST versus non-early-WLST (i.e., survivors, patients with WLST > 7 days, and patients whose cause of death was brain death). Subset analyses were performed between HC, patients with AR, and patients without AR (NoAR). Group differences between Dice scores were evaluated by one-way analysis of variance (ANOVA) followed by post hoc Wilcoxon Rank Sum tests. Analyses were repeated comparing the Dice scores between HC, EMCS patients, and non-EMCS patients. The Rindex's utility for predicting EMCS across all patients was evaluated in terms of sensitivity and specificity. We repeated analyses limited to patients who survived to discharge to control for potential bias from WLST. Correlation analyses using Kendall's τ_b test were used to measure the associations between increased Dice scores or Rindex with improved neurological status (ordered NoAR, AR, HC) and better long-term outcomes (ordered non-EMCS, EMCS, HC).

Data availability

The data underlying this article cannot be shared publicly without restrictions due to privacy concerns of individuals that participated in the study. The data will be shared on reasonable request to the corresponding author, pending approval of the local Institutional Review Board.

Results

Demographics and clinical characteristics

A control group of five healthy individuals (two men (40%), mean age \pm standard deviation (SD) = 37 \pm 18 years) was enrolled. Twenty-two patients were prospectively enrolled following cardiac arrest (see Figure 1). However, three did not undergo fMRI exams (two patients had WLST prior to MRI, and one patient could not tolerate lying supine for an extended period). Demographic, clinical, and treatment characteristics of the remaining 19 patients are shown in Table 1. Eleven patients (58%) experienced AR, of whom five (45%) died before discharge, and four (36%) achieved EMCS. No between-group differences were found for age, sex, and admission clinical characteristics. However, NoAR patients were more likely to have undergone MRI earlier, had lower GCS scores at

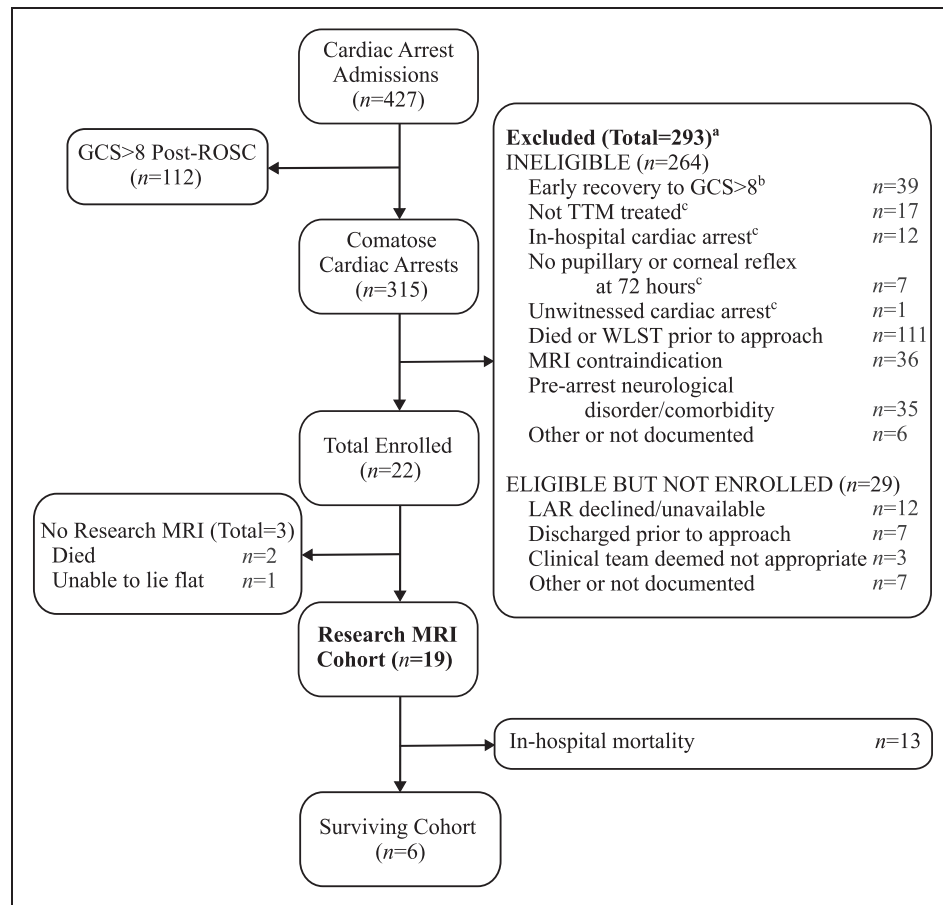


Figure 1. Flow diagram illustrating patient screening, enrollment, and retention. ^aSome patients were not eligible for more than one reason. ^bGCS > 8 before 72 hours or before 24 hours, prior to study protocol amendments. ^cExclusion criteria prior to protocol amendments (see Supplementary Methods). GCS = Glasgow Coma Scale. LAR = Legally authorized representative. ROSC = Restoration of spontaneous circulation. TTM = Targeted temperature management. WLST = Withdrawal of life sustaining treatment.

the time of the MRI, and were more often intubated. In-hospital mortality was 68%, with 92% being due to WLST. 6/12 (50%) of the WLST cases died within seven days of cardiac arrest. None of the NoAR patients survived to discharge.

Seven patients had been enrolled prior to the protocol amendment expanding follow-up visits to 6-months and one-year (see Supplementary Methods). Two of these seven patients survived to discharge; their 6-month and 1-year CPC and mRS scores were carried forward from the 3-month visits. The remaining 12 patients were enrolled with prospective 6-months and 1-year telephone follow-up visits. Four of the 12 patients survived and completed all visits. Among the six survivors, four patients achieved EMCS, and two were in a minimally conscious state¹³ at 1-year post-cardiac arrest. 6-month and 1-year CPC and mRS scores were carried forward for two of the six survivors. For the EMCS patients, the carried forward one-year CPC scores ranged from 1 to 3, with three

patients achieving CPC scores ranging from 1 to 2 and one with CPC = 3. Similarly, the one-year mRS scores ranged from 0 to 4, with two patients achieving mRS 0 to 1 and the other two having mRS = 4. Two of the four EMCS patients had only 3-month follow-up visits and thus might have demonstrated improvement if followed longer.

Task-based fMRI responses

Group-level imaging comparisons between healthy controls and patients. Group-level task-based fMRI results for HC, patients, and their comparisons are shown in Figure 2. Coordinates for the local maxima and cluster size of significant clusters are reported in Supplementary Table 2. Both music and language (forward and backward) tasks induced responses in the auditory cortices of HC and patients. Responses were significantly larger in the HC group compared to the patient cohort, especially in the planum temporale (PT), superior temporal

Table 1. Demographics and clinical characteristics of patients with or without arousal recovery (AR) prior to discharge. Median (IQR) time to eye-opening was 5 (3–7) days, with the minimum being three days and the maximum 22 days.

	Patients (n = 19)	AR (n = 11)	No AR (n = 8)	P-value
Age, years (mean ± SD)	48 ± 23	45 ± 26	53 ± 17	0.41
Sex, male (%)	8 (42%)	2 (18%)	6 (75%)	0.024
Hospital transfer, yes (%)	10 (53%)	6 (55%)	4 (50%)	1.00
Witnessed, yes (%)	10 (53%)	5 (46%)	5 (63%)	0.65
Non-shockable rhythm, yes (%) (n = 18)	10/18 (56%)	4/10 (40%)	6 (75%)	0.19
Primary cardiac etiology, yes (%)	9/19 (47%)	4 (36%)	5 (63%)	0.37
TTM, yes (%)	18/19 (95%)	10 (91%)	8 (100%)	1.0
33°C	17/18 (94%)			
36°C	1/18 (6%)			
Admission GCS, median (IQR)	3 (3–4)	3 (3–4)	3 (3–4.5)	0.73
Admission Pupillary Light Reflex, present (%)	8 (42%)	6 (55%)	2 (25%)	0.35
Admission Corneal Reflex, present (%)	6 (32%)	3 (27%)	3 (38%)	1.00
Admission Cough Reflex, present (%) (n = 18)	8/18 (44%)	4/10 (40%)	4 (50%)	1.00
Admission Gag Reflex, present (%) (n = 17)	7/17 (41%)	4/10 (40%)	3/7 (43%)	1.00
Imaging				
Time-to-MRI, days, median (IQR)	5 (4–8)	7 (5–11)	4 (3–5)	0.007
GCS on day of MRI, median (IQR)	4 (3–6)	5 (4–8)	3 (3–3.75)	0.005
CRS-R on day of MRI, median (IQR) (n = 18)	1 [0–3]	3 [1–4]	0 [0–0]	<0.001
Sedation at MRI, yes (%)	10 (53%)	3 (27%)	7 (88%)	0.020
Intubated at MRI, yes (%)	14 (74%)	6 (55%)	8 (100%)	0.045
Outcomes				
In-hospital death, yes (%)	13 (68%)	5 (45%)	8 (100%)	0.018
Brain death, yes (%)	1/13 (8%)	0 (0%)	1/8 (13%)	
Cardiac death, yes (%)	0/13 (0%)	0 (0%)	0 (0%)	
WLST, yes (%)	12/13 (92%)	5/5 (100%)	7/8 (88%)	
Time-to-death, days median (IQR) (n = 13)	7 (6–12.5)	9 (7.5–18) (n = 5)	6 (5.3–10.8) (n = 8)	0.055
EMCS, yes (%)	4 (21%)	4 (36%)	0 (0%)	0.10
Discharge CPC, median (IQR)	5 (4–5)	4 (3–5)	5 (5–5)	0.018
Discharge mRS, median (IQR)	6 (5–6)	5 (5–6)	6 (6–6)	0.018
90-day CPC, median (IQR)	5 (3–5)	4 (3–5)	5 (5–5)	0.018
90-day mRS, median (IQR)	6 (5–6)	5 (4–6)	6 (6–6)	0.019
6-month CPC, median (IQR)	5 (3–5)	3 (2–5)	5 (5–5)	0.018
6-month mRS, median (IQR)	6 (5–6)	5 (4–6)	6 (6–6)	0.019
12-month CPC, median (IQR)	5 (3–5)	3 (2–5)	5 (5–5)	0.018
12-month mRS, median (IQR)	6 (5–6)	5 (4–6)	6 (6–6)	0.019

AR: Arousal Recovery; TTM: Targeted Temperature Management; GCS: Glasgow Coma Scale; CPC: Cerebral Performance Category; CRS-R: Coma Recovery Scale Revised; EMCS: Emergence from the Minimally Conscious State measured on CRS-R; WLST: Withdrawal of Life Sustaining Treatment; mRS: Modified Rankin Scale. $P < 0.05$ are in bold.

gyrus (STG), and Heschl's gyrus (HG). These regions have been shown to be responsible for auditory feature extraction, comprehension, and semantic and syntactic processing.^{35–38} In addition, both language tasks included the involvement of the bilateral inferior frontal gyri (IFG) and middle temporal gyri (MTG) in the HC cohort but not the patient group. These are brain regions that are responsible for lexical and semantic processing and speech comprehension (MTG) and for phonological processing, context, and semantic memory (IFG) based on prior language processing models.^{37,39,40} Limiting the analysis to English-speaking patients resulted in similar attenuated

responses with no IFG involvement (Supplementary Fig. 3). Motor imagery induced significantly greater activations in HC than in patients, with widespread activation involving the frontal lobe, precentral gyrus, and supplemental motor areas. No significant brain responses were observed in the group-level patient results for motor imagery.

Group-level imaging comparisons between patients with and without arousal recovery. Group-level imaging differences were evident between AR and NoAR patients, with no task-relevant responses observed during language tasks for the NoAR group (Figure 3). Coordinates for the

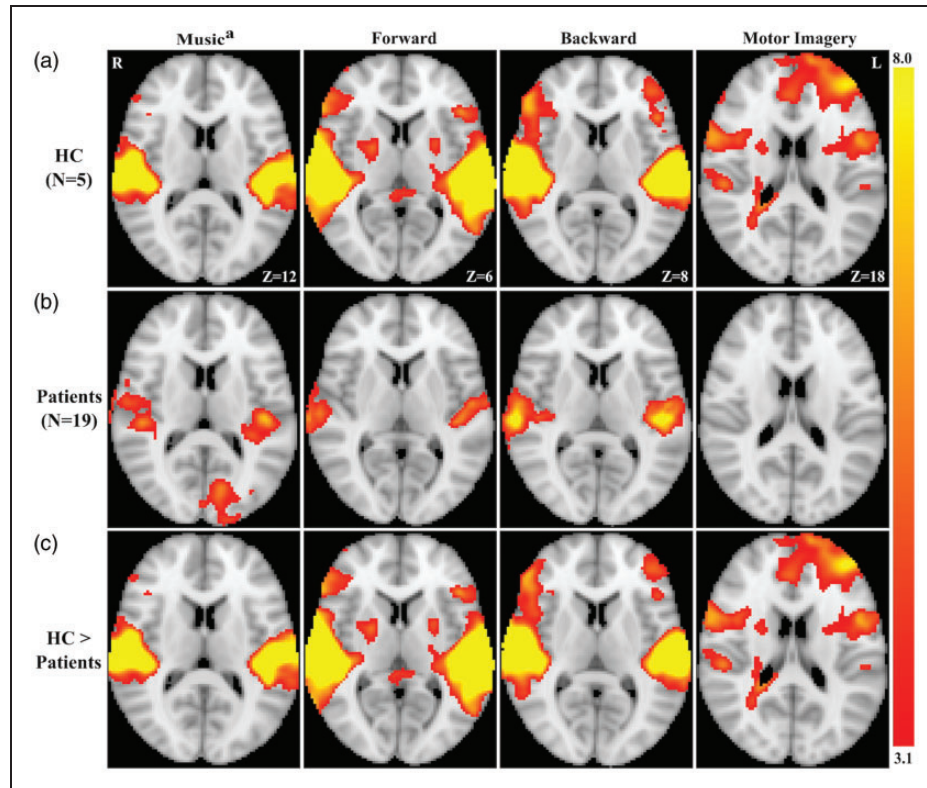


Figure 2. Group-level task-based fMRI responses for healthy controls, patients, and their comparison. The Group-level fMRI results for (a) healthy controls (HC) and (b) patients are shown. (c) For all tasks, group-level comparisons show significantly higher brain responses in healthy controls than in patients. Patients have no significant brain responses during motor imagery, whereas healthy controls show responses across frontal and motor areas. HC = Healthy Control. ^a $n = 18$ for music: one patient was excluded because of incomplete data.

local maxima and cluster size of significant clusters for AR, NoAR, and AR>NoAR are summarized in Supplementary Table 3. The music and language tasks induced significantly greater brain responses in auditory cortices (HG, PT, and STG) in patients with AR compared to the NoAR group. Involvement of IFG or MTG was not evident in either group for both language tasks. Additionally, neither group demonstrated significant task-relevant responses during the motor imagery task (Figure 3).

Group-level imaging comparisons between patients with good and poor long-term outcomes. The group-level responses of patients achieving EMCS (Figure 4) exhibited brain activation patterns involving primary auditory and auditory association cortices, similar to HC (Supplementary Fig. 4) for the music and language tasks (including the IFG), although weaker and to a lesser spatial extent in some regions like the PT, MTG and IFG. In comparison (Figure 4), the non-EMCS outcome group demonstrated substantially reduced activations for the music task in the auditory cortices and no task-relevant activation patterns for the language tasks. Significant

group-level differences between EMCS and non-EMCS cohorts for music and language tasks were observed. No significant group-level responses were noted during motor imagery tasks for either EMCS or non-EMCS groups (Figure 4). Coordinates for the local maxima and cluster size of significant clusters for patients with EMCS, non-EMCS, and their comparisons are summarized in Supplementary Table 4. Coordinates for the local maxima and cluster size of significant clusters for HC compared to EMCS patients are summarized in Supplementary Table 5.

Reference image selection. Dice scores for individual healthy control subjects' brain responses (i.e., Z-statistic maps) against anatomical reference images and group-level functional reference images are provided in Supplementary Table 6. Overall, Dice scores based on group-level functional reference images outperformed scores based on anatomical reference images. Thus, the group-level functional images were selected as the reference images to quantify the degree of task-relevant responses in individual patients.

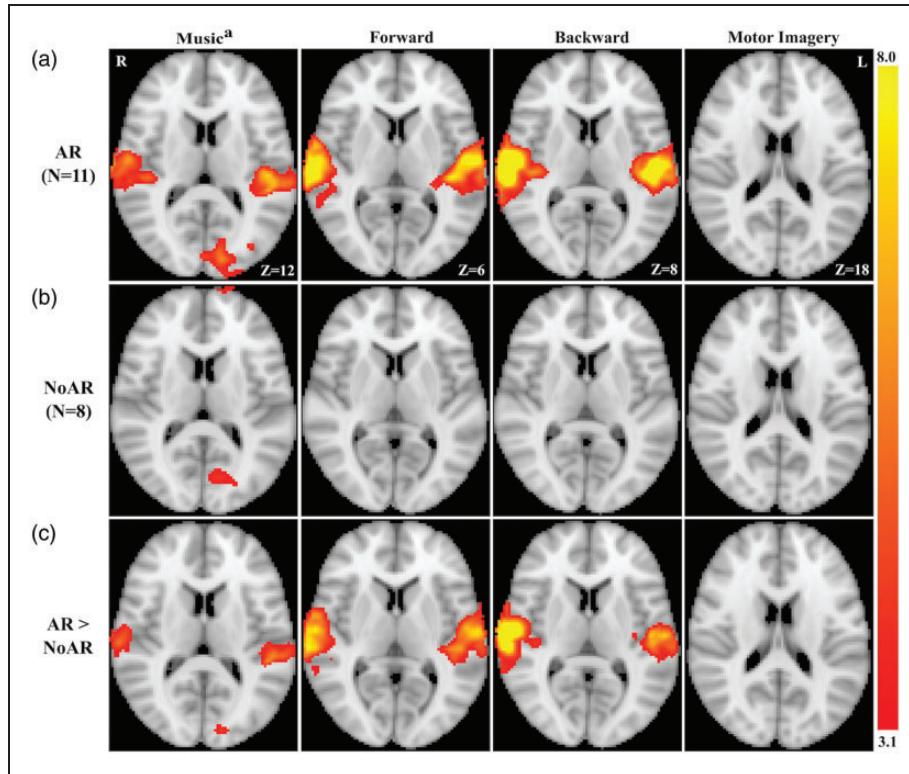


Figure 3. Group-level task-based fMRI responses to music, language, and motor imagery in patients with and without arousal recovery and their comparison. The Group-level fMRI results are shown for (a) patients with AR and (b) without AR (NoAR). (c) Group-level comparisons showed greater activity in patients with AR than in the NoAR patients for music and language tasks. AR = Arousal Recovery. NoAR = No Arousal Recovery. ^a $n = 10$ for music for AR: one patient was excluded because of incomplete data.

Quantification of task-relevant fMRI responses on an individual subject basis. Dice scores for individual HCs and patients are provided in Supplementary Tables 6 and 7, respectively. Individual brain responses for HCs (Supplementary Fig. 5), AR (Supplementary Fig. 6), and NoAR (Supplementary Fig. 7) cohorts are shown in the Supplementary material. No statistically significant differences in Dice scores were observed for all tasks between sedated versus non-sedated patients, intubated versus non-intubated patients, and early-WLST versus non-early-WLST patients (Table 2). Therefore, for the remainder of the analyses, no adjustments were made for sedation, intubation, or early WLST. Higher Dice scores were noted in HCs for all tasks compared to all patients (music: 0.43 (0.18–0.49) vs. 0.0 (0.0–0.045), $P = 0.003$; forward: 0.42 (0.33–0.59) vs. 0.0 (0.0–0.04), $P < 0.001$; backward: 0.48 (0.36–0.54) vs. 0.0 (0.0–0.04), $P < 0.001$; motor: 0.25 (0.02–0.35) vs. 0.0 (0.0–0.02), $P = 0.02$).

Not surprisingly, significant group differences (HC, AR, NoAR) in Dice scores were noted for all tasks (One-way ANOVA music: $F(2,22) = 8.92$, $P = 0.0017$; forward: $F(2,23) = 26.38$, $P < 0.001$, backward: $F(2,23) = 25.43$, $P < 0.001$; motor imagery: $F(2,23) = 4.53$,

$P = 0.023$) as shown in Figure 5. Dice scores for HC (0.43 (0.18–0.49)) were greater than AR (0.0 (0.0–0.26), $P = 0.019$) and NoAR (0.0 (0.0–0.02), $P = 0.005$) for music. Similar results were found for forward language for HC (0.42 (0.33–0.59)) compared to AR (0.0 (0.0–0.1), $P = 0.005$) and to NoAR (0.0 (0.0–0.03), $P = 0.003$) and for backward language for HC (0.48 (0.36–0.54)) compared to AR (0.0 (0.0–0.14), $P = 0.005$) and compared to NoAR (0.0 (0.0–0.0), $P = 0.001$). For motor imagery, Dice scores for HC (0.25 (0.02–0.35)) were noted to be significantly larger than AR (0.0 (0.0–0.0), $P = 0.023$) but not significantly greater than NoAR (0.0 (0.0–0.10), $P = 0.089$). There were no statistically significant differences in Dice scores between AR ($n = 11$) and NoAR ($n = 8$) groups for all tasks (music: $P = 0.69$, forward: $P = 0.36$, motor imagery: $P = 0.39$) except for the backward language task ($P = 0.038$). Except for motor imagery ($\tau_b = 0.24$, $P = 0.19$), Dice scores for all tasks were significantly correlated with neurologic recovery (music: $\tau_b = 0.46$, $P = 0.01$; forward language: $\tau_b = 0.56$, $P = 0.002$; backward language: $\tau_b = 0.70$, $P < 0.001$).

Similar results were found comparing Dice scores between HC ($n = 5$), patients who achieved EMCS

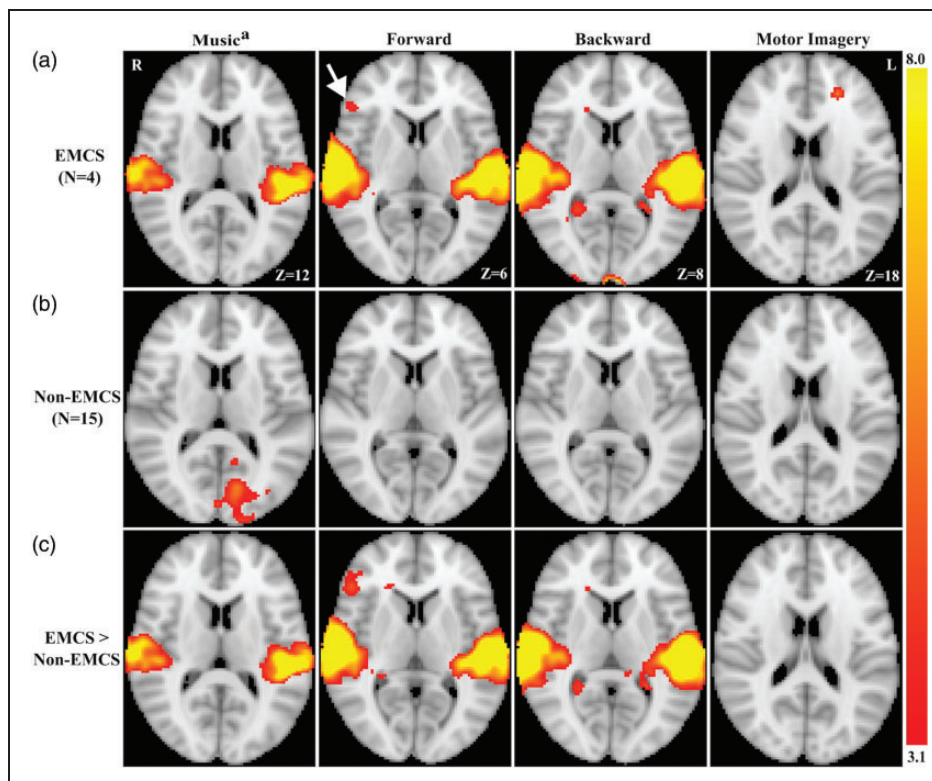


Figure 4. Group-level task-based fMRI responses to music, language, and motor imagery tasks in patients with EMCS, non-EMCS, and their comparison. The Group-level fMRI results are shown for patients with (a) EMCS (emergence from the minimally conscious state) and (b) non-EMCS. (c) Group-level comparisons between the groups show greater responses in patients with EMCS than in patients with non-EMCS. Interestingly, the EMCS forward language results show involvement of the IFG (arrow) which was not evident in the AR group-level findings (Figure 3). ^a $n = 3$ for music: one patient was excluded because of incomplete data. EMCS = Emergence from the minimally conscious state.

Table 2. Univariable Dice score analysis by potentially confounding factors.

	Yes	No	P-value
Sedated ($n = 10$) vs Non-sedated ($n = 9$)			
Music, median (IQR)	0.0 (0.0–0.045)	0.0 (0.0–0.17)	0.10
Forward Language, median (IQR)	0.0 (0.0–0.055)	0.0 (0.0–0.015)	0.67
Backward Language, median (IQR)	0.0 (0.0–0.0)	0.0 (0.0–0.09)	0.14
Motor Imagery, median (IQR)	0.0 (0.0–0.1)	0.0 (0.0–0.0)	0.27
Intubated ($n = 14$) vs Non-Intubated ($n = 5$)			
Music, median (IQR)	0.0 (0.0–0.045)	0.0 (0.0–0.33)	0.86
Forward Language, median (IQR)	0.0 (0.0–0.04)	0.0 (0.0–0.23)	0.63
Backward Language, median (IQR)	0.0 (0.0–0.01)	0.0 (0.0–0.27)	0.44
Motor Imagery, median (IQR)	0.0 (0.0–0.10)	0.0 (0.0–0.0)	0.15
Early-WLST ($n = 6$) vs Non-Early-WLST ($n = 13$)			
Music, median (IQR)	0.005 (0.0–0.045)	0.0 (0.0–0.17)	0.83
Forward Language, median (IQR)	0.0 (0.0–0.04)	0.0 (0.0–0.06)	0.80
Backward Language, median (IQR)	0.0 (0.0–0.0)	0.0 (0.0–0.09)	0.10
Motor Imagery, median (IQR)	0.0 (0.0–0.13)	0.0 (0.0–0.01)	0.65

All comparisons were calculated using a two-sample Wilcoxon Rank Sum Test with normal approximation. WLST: Withdrawal of Life Sustaining Treatment.

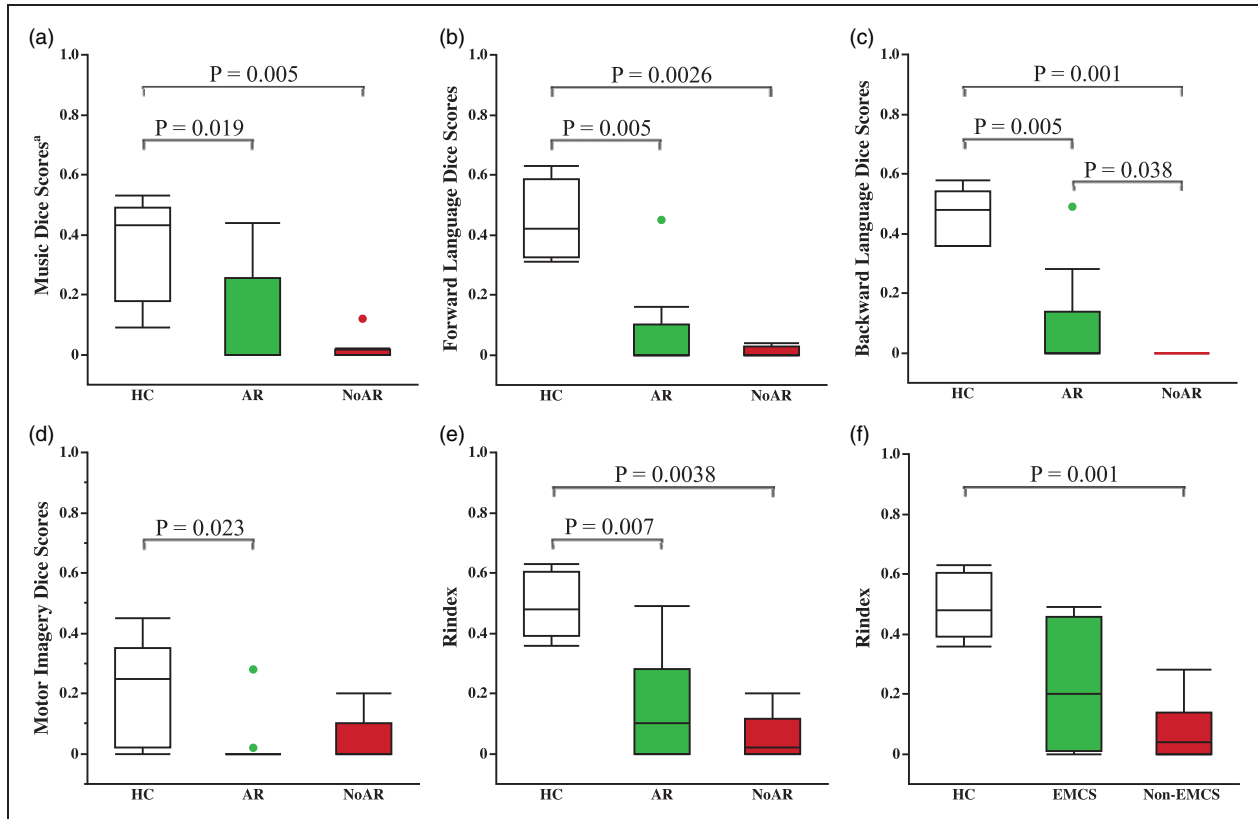


Figure 5. Dice Scores and Rindex for Healthy Controls and Patients measuring responsiveness to tasks during fMRI. Boxplots comparing Dice scores and Rindex for healthy controls (HC) and patients. Dice scores for HC, patients with arousal recovery (AR), and without AR (NoAR) for (a) music, (b) forward language, (c) backward language, and (d) motor imagery. Rindex comparisons (e) between HC, AR, and NoAR groups, and (f) between HC, patients who achieved EMCS and patients who did not (Non-EMCS) by one year. Statistically significant differences between groups are shown. AR = Arousal Recovery. EMCS = Emergence from the minimally conscious state. NoAR = No Arousal Recovery. ^a $n = 10$ for music AR: one patient was excluded because of incomplete data.

($n = 4$), and patients who did not ($n = 15$). Significant group differences in Dice scores were noted for all tasks (One-way ANOVA music: $F(2,22) = 16.45$, $P < 0.001$; forward: $F(2,23) = 34.02$, $P < 0.001$; backward: $F(2,23) = 37.76$, $P < 0.001$; motor imagery: $F(2,23) = 4.70$, $P = 0.021$). For music, forward language, and motor imagery, HC Dice scores were greater than the non-EMCS group's Dice scores (music: 0.0 (0.0–0.01), $P = 0.001$; forward: 0.0 (0.0–0.02), $P < 0.001$; motor imagery: 0.0 (0.0–0.10), $P = 0.034$) but were comparable to the EMCS group's Dice scores (music: 0.36 (0.0–0.44), $P = 0.55$; forward: 0.08 (0.0–0.38), $P = 0.11$; motor imagery: 0.0 (0.0–0.015), $P = 0.072$). For backward language, HC Dice scores were larger than non-EMCS (0.0 (0.0–0.0), $P < 0.001$) but comparable to EMCS (0.16 (0.01–0.44), $P = 0.11$) outcomes groups. No statistically significant differences in Dice scores between EMCS and non-EMCS outcomes were found for music ($P = 0.11$), forward language ($P = 0.27$), and motor imagery ($P = 0.80$), but for the backward language task, the EMCS group had significantly greater Dice scores ($P = 0.012$). Dice scores for all tasks except

for motor imagery ($\tau_b = 0.33$, $P = 0.07$) were significantly correlated with improved consciousness (music: $\tau_b = 0.62$, $P < 0.001$; forward language: $\tau_b = 0.61$, $P < 0.001$; backward language: $\tau_b = 0.77$, $P < 0.001$).

Association of Rindex with arousal recovery and long-term outcomes. The Rindex for HC (0.48 (0.39–0.61)) was significantly greater than those of the AR (0.1 (0.0–0.28), $P = 0.007$) and NoAR (0.02 (0.0–0.12), $P = 0.0038$) groups. There was no significant difference in Rindex between AR and NoAR groups ($P = 0.28$). Interestingly, there was no significant difference in the Rindex between HC and the EMCS group (0.2 (0.01–0.046), $P = 0.27$). HC had a significantly greater Rindex than the non-EMCS group (0.04 (0–0.14), $P = 0.001$). There was no significant difference in the Rindex between the EMCS and non-EMCS outcomes groups ($P = 0.14$). The Rindex was significantly correlated with improved arousal ($\tau_b = 0.54$, $P = 0.002$) and consciousness ($\tau_b = 0.62$, $P = 0.001$).

For all patients, Rindex > 0 was 75% [95% CI: 19–99%] sensitive and 47% [95% CI: 21–73%] specific for

predicting EMCS. Limiting the analysis to patients who survived to discharge ($n=6$), the sensitivity of $R_{index} > 0$ for predicting EMCS ($n=4$) was 75% [95% CI: 19–99%] with 100% [16–100%] specificity. The specificity of absent task-relevant responses (i.e., $R_{index} = 0$) for predicting poor outcomes (i.e., non-EMCS) ($n=2$) was only 75% [95% CI: 19–99%].

Discussion

This prospective study investigated brain responses in comatose cardiac arrest patients during instrumental music, language (forward and backward speech), and motor imagery (imagine squeezing the right hand) tasks using functional MRI. The brain activation patterns we observed in HC were consistent with previously reported language^{39,40} and music^{35,36} processing models. Group-level results showed that language, music, and motor imagery tasks induced significantly greater responses in healthy controls than patients, with activations primarily localized to primary auditory and auditory association cortices during music and language. During motor imagery tasks, although more widespread, involving frontal regions and motor cortices, responses were weaker compared to music and language stimuli, even for healthy controls. The attenuated brain responses involving HG, STG, PT, MTG, and IFG in patients during music and language tasks suggest disruptions in auditory perception, auditory information processing, and semantic processing. The HG and STG are key brain regions responsible for auditory perception and information processing. On the other hand, MTG and PT are essential for lexical and semantic processing and speech comprehension, and IFG for phonological processing, contextual processing, and semantic memory.^{37,39,40} Overall, group-level results in the patients suggest preservation of auditory perception but impairment of semantic and phonological processing, including language.

We investigated whether there were differences in task-based fMRI between AR and NoAR patients. Group-level differences showed that the AR group had significantly greater activations for music and language tasks than the NoAR group but not for the motor imagery task (Figure 3). Furthermore, for the NoAR group, there were no task-relevant responses for either music or language tasks. However, even for the AR group, no IFG involvement was noted for the language tasks. Interestingly, activation of the left occipital lobe was observed for the AR group (and to a lesser extent for the NoAR group) for the music task but not for language. Previous studies have shown that in healthy volunteers, the occipital lobe was involved in music perception and harmonic processing but not language.^{41,42} We did not observe this for our HC

group-level results, likely due to the small sample size ($n=5$). In addition, for the NoAR group, task-relevant activity was only observed for music (although to a lesser extent than the AR group) but not for the language tasks, perhaps due to the greater complexity involved in speech processing. Our findings are consistent with a previous study of patients with disorders of consciousness that showed musical stimuli elicited better responses than verbal commands.⁴³

We also performed group-level analyses of patients with EMCS ($n=4$) versus non-EMCS ($n=15$) (Figure 4). Limiting the group-level analysis to EMCS patients, despite involving a smaller number of subjects, resulted in an increased spatial extent of activation that included auditory association cortices compared to AR group-level findings. The auditory responses in the EMCS cohort demonstrated more widespread activation patterns in the primary auditory cortex (HG) and association cortices (PT and STG) and some frontal IFG responses to language, suggesting residual functionality with auditory processing and semantic processing. In contrast, the group-level results for the patients who did not achieve EMCS were similar to those of the NoAR group-level findings, except for the music task, which showed greater involvement of the occipital pole for the non-EMCS group. Decisions about WLST may have confounded the NoAR and Non-EMCS results; 13 of 15 patients died before discharge, with only one attributed to brain death and the rest due to WLST.

Perhaps more clinically relevant than group-level differences, we quantified the strength and spatial extent of results on an individual patient basis using task-relevant Dice scores. Since the reference images are binarized maps derived from the HC group-level Z-statistic images (Supplementary Fig. 2A), one would expect perfect Dice scores for the healthy controls. Instead, the median HC Dice scores across tasks ranged from 0.25 to 0.48, with the poorest performance attributed to the motor imagery task. The large IQR of the Dice scores is due to the variance of fMRI results, as can be qualitatively noted across the individual HC subjects' brain activation maps (Supplementary Fig. 5). Indeed, the known variability of fMRI results^{30,44} is a barrier to its routine clinical use.⁴⁵ Another potential explanation for the low Dice score is the relatively small number of HC datasets used to create each task-related reference image. A larger number of datasets may lead to more robust reference images. In addition, the Dice score penalizes for both false positives (non-task-relevant activations) and false negatives (absent task-relevant activations) and depends on the statistical threshold and choice of reference images. The Dice score may thus be an overly strict measure of reproducibility. Other measures have been proposed

when discussing fMRI reliability.^{33,46–48} However, these metrics are affected by the same factors as the Dice score, such as the statistical threshold for determining significant clusters and the need for reference images. In addition, these alternative measures, e.g., fMRI signal stability or shifts in the location of cluster centroids, may be less clinically useful.

Despite these potential limitations, increasing Dice scores were associated with better outcomes, measured by either arousal recovery or achieving EMCS (Figure 5). Although some of the patients with AR (Supplementary Fig. 6) have activation patterns in the primary auditory and auditory association cortices comparable to healthy controls (Supplementary Fig. 5), many did not, as reflected in the Dice scores. The lack of statistically significant differences between AR and NoAR groups is likely due to the small sample size and large variance in Dice scores across individuals. The exception was the backward language task, for which none of the NoAR patients exhibited task-relevant responses. A previous study has reported that reverse speech is associated with increased attention,⁴⁹ which may have led to more robust activation in patients with covert consciousness. Studies with larger numbers of patients are needed to test this hypothesis.

Even for patients who achieved EMCS, Dice scores varied considerably. Two patients (P09 and P10) with EMCS had minuscule Dice scores, whereas the other two patients (P03 and P08) had Dice scores comparable to those of healthy controls. For patients P09 and P10, high levels of motion were detected during the fMRI, which diminished signal quality and might have contributed to the failure to detect any task-relevant responses (see Supplementary Fig. 6). The variability of task-based fMRI results amongst patients is not surprising⁵⁰, considering the wide range of Dice scores observed in healthy individuals. Indeed, one healthy control, C01 (Supplementary Table 6), showed weak responses to a passive music listening task (Dice=0.09) and to an active motor imagery task (Dice=0.0). Although larger sample sizes may mitigate the effects on a group-wise basis, our primary focus is to assess the accuracy of task-based fMRI in predicting recovery of consciousness on an individual patient basis. Our results demonstrate that task-based fMRI results are highly variable even in healthy controls. Therefore, caution must be used in interpreting the absence of task-relevant fMRI responses as specific indicators of poor neurologic outcomes.

To compensate for the variability of individual responses, we calculated the Rindex, which is the maximum Dice score for the four tasks. The Rindex provides an intuitive measure of participants' responsiveness without penalizing patients for weak performance on some tasks. For patients who survived

to discharge ($n=6$), the Rindex was 100% specific for identifying patients who achieved EMCS using $Rindex > 0$. However, amongst non-survivors whose cause of death was WLST, there were many patients in our study with $Rindex > 0$. Therefore, it is difficult to determine whether the Rindex is genuinely 100% specific in predicting which patients will achieve EMCS if allowed to survive, as these patients might or might not have neurologically recovered. Further research is needed.

Limitations

There were several potential limitations to our study. The primary limitation was the small sample size, with only five healthy controls and 19 patients. Because of the small sample size, we used a fixed-effect analysis for group-level brain responses instead of the more conservative mixed-effects analyses typically employed when generalizing group-level inferences to the population level.⁵¹ The varying task-relevant responses in patients could have been confounded by multiple factors, including behavioral and physiological conditions and ongoing treatment and medications.^{52–56} Although we did not find statistically significant between-group differences in Dice results for sedated, intubated, or early-WLST patients, it may be due to the small sample sizes. On the other hand, some patients exhibited high Dice and Rindex scores even though they were sedated, consistent with other studies in healthy controls⁵⁷ and patients.^{8,58,59}

Other potential confounders could be due to variable image acquisition in clinical MRI scanners compared to dedicated research systems.^{60,61} Although the same MRI protocol and scanner were used for all participants, scanner upgrades occurred during the study. In addition, different head coils were used depending on the patient's head size, which led to an inconsistent number of slices (49 slices for the 32-channel head coil and 46 slices for the 20-channel head and neck coil) and fMRI signal-to-noise ratio. However, robust task-relevant responses were observed for the healthy control scanned with the 20-channel coil (Supplementary Table 6, C03), suggesting that usage of the 20-channel coil did not interfere with the detection of task-relevant fMRI responses. Another potential confounder was the timing of the fMRI scans (5 [4–8] days post-cardiac arrest), which depended on the patient's clinical condition, such as hemodynamic stability. The evolution of brain injury in comatose cardiac arrest patients is still not well understood. Several serial studies suggest the time course of hypoxic-ischemic injury is highly dynamic, especially in the first few days after cardiac arrest.^{5,12} Thus, the differences observed between AR

patients, who were scanned later, and NoAR patients could have been related to differences in timing.

Including non-English speakers ($n = 4$) in the patient cohort may have contributed to attenuated responses. However, similar activation patterns in group-level results after excluding non-English speakers were found (Supplementary Fig. 3). We included non-English speakers to ensure that the demographics of our research patient sample reflected the general cardiac arrest patient population. Furthermore, previous studies have shown the presence of task-relevant brain responses in the auditory cortices in response to non-native phonemes.⁶² In our study, non-English speaking patients exhibited less task-relevant activations than the English-speaking patients (Supplementary Fig. 3). However, three of these patients did not exhibit arousal recovery, which could explain the reduced group-level responses for this group. Future studies should compare task-based fMRI results from speech paradigms in the patient's native language to responses in English to better investigate the effects of speech comprehension.

Another possible source of bias is that patients included in the study needed to be able to undergo an MRI, excluding hemodynamically unstable patients and patients with MRI contraindications such as implantable cardiac devices. This might have led to a healthier population enrolled in our study. In addition, performing fMRI in comatose cardiac arrest patients can be challenging since many patients are hemodynamically unstable and mechanically ventilated, limiting the widespread use of task-based fMRI for consciousness evaluation.

Lastly, the results of our study are likely confounded by WLST, which makes studies of comatose cardiac arrest survivors difficult in general. Indeed, the primary reason for screened patients being excluded from the study was WLST (111/293, 38%). Patients with severe injury on admission CT may have been excluded due to premature WLST, thereby skewing our enrolled patient sample. In addition, sharing individual patient-level activation maps with the clinical team may have led to self-fulfilling prophecy bias.⁶³ While only fMRI Z-statistic images were shared in these settings, these images are associated with Dice scores and Rindex values.

Conclusions

Our task-based fMRI results suggest it is a promising tool for detecting covert consciousness in comatose cardiac arrest patients. The degree of task-relevant responses appears to increase with an improved level of consciousness, especially during backward spoken language. While task-based fMRI findings combined with other modalities in assessing post-cardiac arrest

patients will provide greater insight into the extent of neurologic injury, as a single modality, they appear more suited to predicting potentially good outcomes rather than predicting poor outcomes. Caution is necessary because many factors may confound the results, including motion artifacts in encephalopathic but conscious patients or sedation in comatose patients. Replication of these pilot findings will be necessary to confirm that they generalize to diverse patients, communities, and clinical environments.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institutes of Health [R01NS102574-01A1, DP2HD101400]; American Heart Association [19A1ML35170037]; and Massachusetts General Hospital [Institute for Heart, Vascular and Stroke Care SPARK award, Executive Committee on Research].

Acknowledgements

The authors thank Brittany Mills, the clinical staff of the Massachusetts General Hospital intensive care units, and MRI radiology technologists for their assistance in data acquisition. We acknowledge Andre van der Kouwe and Dylan Tisdall (Athinoula A. Martinos Center for Biomedical Imaging) and Himanshu Bhat (Siemens Medical Center) for the provision of WIP711D (vNav Motion-Corrected Multiecho MPRAGE) and MEMPRAGE sequences used to acquire the 3D T1-weighted anatomical images. We are grateful to the patients and their families for their generous participation in this study.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' contributions

Kiran Dhakal: Study design, data analysis, interpretation of data, drafting and revising the manuscript for intellectual content

Eric S. Rosenthal: Study design, data acquisition, interpretation of data, revising the manuscript for intellectual content

Annelise M Kulpanowski: Data acquisition, interpretation of data, revising the manuscript for intellectual content

Jacob A. Dodelson: Data acquisition, interpretation of data, revising the manuscript for intellectual content

Zihao Wang: Data analysis, interpretation of data, revising the manuscript for intellectual content

Gaston Cudemus-Deseda: Data acquisition, interpretation of data, revising the manuscript for intellectual content

Marjorie Villien: Data acquisition, interpretation of data, revising the manuscript for intellectual content

Brian L. Edlow: Study design, data acquisition, interpretation of data, revising the manuscript for intellectual content

Alexander M. Presciutti: Interpretation of data, revising the manuscript for intellectual content

James L. Januzzi: Interpretation of data, revising the manuscript for intellectual content

MingMing Ning: Data acquisition, interpretation of data, revising the manuscript for intellectual content

W. Taylor Kimberly: Data acquisition, interpretation of data, revising the manuscript for intellectual content

Edilberto Amorim: Data acquisition, interpretation of data, revising the manuscript for intellectual content

M Brandon Westover: Data acquisition, interpretation of data, revising the manuscript for intellectual content

William A. Copen: Interpretation of data, revising the manuscript for intellectual content

Pamela W. Schaefer: Data acquisition, interpretation of data, revising the manuscript for intellectual content

Joseph T. Giacino: Data acquisition, interpretation of data, revising the manuscript for intellectual content

David M. Greer: Study design, interpretation of data, revising the manuscript for intellectual content

Ona Wu: Study concept, study design, study supervision, data acquisition, data analysis, interpretation of data, drafting and revising the manuscript for intellectual content

All authors critically revised and approved the final manuscript for publication.

ORCID iDs

Kiran Dhakal  <https://orcid.org/0000-0002-9502-6419>

Ona Wu  <https://orcid.org/0000-0002-5509-9461>

Supplementary material

Supplemental material for this article is available online.

References

1. Tsao CW, Aday AW, Almarzoq ZI, et al. Heart disease and stroke statistics-2022 update: a report from the American heart association. *Circulation* 2022; 145: e153–e639.
2. Geocadin RG, Callaway CW, Fink EL, et al. Standards for studies of neurological prognostication in comatose survivors of cardiac arrest: a scientific statement from the American heart association. *Circulation* 2019; 140: e517–e542.
3. Panchal AR, Bartos JA, Cabanas JG, et al. Part 3: adult basic and advanced life support: 2020 American Heart Association Guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation* 2020; 142: S366–S468.
4. Giacino JT, Katz DI, Schiff ND, et al. Comprehensive systematic review update summary: disorders of consciousness: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. *Neurology* 2018; 91: 461–470.
5. Greer D, Scripko P, Bartscher J, et al. Serial MRI changes in comatose cardiac arrest patients. *Neurocrit Care* 2011; 14: 61–67.
6. Katz DI, Polyak M, Coughlan D, et al. Natural history of recovery from brain injury after prolonged disorders of consciousness: outcome of patients admitted to inpatient rehabilitation with 1–4 year follow-up. *Prog Brain Res* 2009; 177: 73–88.
7. Nakase-Richardson R, Whyte J, Giacino JT, et al. Longitudinal outcome of patients with disordered consciousness in the NIDRR TBI model systems programs. *J Neurotrauma* 2012; 29: 59–65.
8. Edlow BL, Chatelle C, Spencer CA, et al. Early detection of consciousness in patients with acute severe traumatic brain injury. *Brain* 2017; 140: 2399–2414.
9. Greer DM and Wu O. Neuroimaging in cardiac arrest prognostication. *Semin Neurol* 2017; 37: 66–74.
10. Wijdicks EF, Campeau NG and Miller GM. MR imaging in comatose survivors of cardiac resuscitation. *AJNR Am J Neuroradiol* 2001; 22: 1561–1565.
11. Wu O, Batista LM, Lima FO, et al. Predicting clinical outcome in comatose cardiac arrest patients using early noncontrast computed tomography. *Stroke* 2011; 42: 985–992.
12. Wu O, Sorensen AG, Benner T, et al. Comatose patients with cardiac arrest: predicting clinical outcome with diffusion-weighted MR imaging. *Radiology* 2009; 252: 173–181.
13. Giacino JT, Ashwal S, Childs N, et al. The minimally conscious state: definition and diagnostic criteria. *Neurology* 2002; 58: 349–353.
14. Giacino JT, Kalmar K, and Whyte J. and The JFK coma recovery scale-revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* 2004; 85: 2020–2029.
15. Department of Health E, Welfare, National Commission for the Protection of Human Subjects of B, et al. The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. *J Am Coll Dent* 2014; 81: 4–13.
16. van der Kouwe AJW, Benner T, Salat DH, et al. Brain morphometry with multiecho MPRAGE. *NeuroImage* 2008; 40: 559–569.
17. Tisdall MD, Hess AT, Reuter M, et al. Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magn Reson Med* 2012; 68: 389–399.
18. Iglesias JE, Liu CY, Thompson PM, et al. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Trans Med Imaging* 2011; 30: 1617–1634.
19. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002; 17: 143–155.
20. Greve DN and Fischl B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* 2009; 48: 63–72.
21. Jenkinson M, Bannister P, Brady M, et al. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 2002; 17: 825–841.
22. Jenkinson M and Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* 2001; 5: 143–156.

23. Woolrich MW, Ripley BD, Brady M, et al. Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage* 2001; 14: 1370–1386.
24. Beckmann CF, Jenkinson M and Smith S. General multi-level linear modeling for group analysis in fMRI. *NeuroImage* 2003; 20: 1052–1063.
25. Woolrich M. Robust group analysis using outlier inference. *NeuroImage* 2008; 41: 286–301.
26. Woolrich MW, Behrens TEJ, Beckmann CF, et al. Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage* 2004; 21: 1732–1747.
27. Worsley KJ. Statistical analysis of activation images. In: Jezzard P, Matthews PM and Smith SM (eds), *Functional Magnetic Resonance Imaging: An Introduction to Methods*. Oxford: Oxford Academic, 2001.
28. Eklund A, Nichols TE and Knutsson H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci U S A* 2016; 113: 7900–7905.
29. Voyvodic JT. Activation mapping as a percentage of local excitation: fMRI stability within scans, between scans and across field strengths. *Magn Reson Imaging* 2006; 24: 1249–1261.
30. Bennett CM and Miller MB. How reliable are the results from functional magnetic resonance imaging? *Ann N Y Acad Sci* 2010; 1191: 133–155.
31. Clement F and Belleville S. Test-retest reliability of fMRI verbal episodic memory paradigms in healthy older adults and in persons with mild cognitive impairment. *Hum Brain Mapp* 2009; 30: 4033–4047.
32. Fröhner JH, Teckentrup V, Smolka MN, et al. Addressing the reliability fallacy in fMRI: similar group effects may arise from unreliable individual effects. *NeuroImage* 2019; 195: 174–189.
33. Morrison MA, Churchill NW, Cusimano MD, et al. Reliability of task-based fMRI for preoperative planning: a test-retest study in brain tumor patients and healthy controls. *PLoS One* 2016; 11: e0149547.
34. Raemaekers M, Vink M, Zandbelt B, et al. Test-retest reliability of fMRI activation during prosaccades and antisaccades. *Neuroimage* 2007; 36: 532–542.
35. Koelsch S. Toward a neural basis of music perception – a review and updated model. *Front Psychol* 2011; 2: 110.
36. Warren J. How does the brain process music? *Clin Med (Lond)* 2008; 8: 32–36.
37. Xu Y, Gandour J, Talavage T, et al. Activation of the left planum temporale in pitch processing is shaped by language experience. *Hum Brain Mapp* 2006; 27: 173–183.
38. Nakada T, Fujii Y, Yoneoka Y, et al. Planum temporale: where spoken and written language meet. *Eur Neurol* 2001; 46: 121–125.
39. Friederici AD. The brain basis of language processing: from structure to function. *Physiol Rev* 2011; 91: 1357–1392.
40. Fujii M, Maesawa S, Ishiai S, et al. Neural basis of language: an overview of an evolving model. *Neurol Med Chir (Tokyo)* 2016; 56: 379–386.
41. Schmithorst VJ. Separate cortical networks involved in music perception: preliminary functional MRI evidence for modularity of music processing. *Neuroimage* 2005; 25: 444–451.
42. Satoh M, Takeda K, Nagata K, et al. Activated brain regions in musicians during an ensemble: a PET study. *Brain Res Cogn Brain Res* 2001; 12: 101–108.
43. O’Kelly J and Magee WL. The complementary role of music therapy in the detection of awareness in disorders of consciousness: an audit of concurrent SMART and MATADOC assessments. *Neuropsychol Rehabil* 2013; 23: 287–298.
44. Gorgolewski KJ, Storkey AJ, Bastin ME, et al. Single subject fMRI test–retest reliability metrics and confounding factors. *NeuroImage* 2013; 69: 231–243.
45. Voyvodic JT. Reproducibility of single-subject fMRI language mapping with AMPLE normalization. *J Magn Reson Imaging* 2012; 36: 569–580.
46. Caceres A, Hall DL, Zelaya FO, et al. Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage* 2009; 45: 758–768.
47. Wei X, Yoo SS, Dickey CC, et al. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *Neuroimage* 2004; 21: 1000–1008.
48. Nettekoven C, Reck N, Goldbrunner R, et al. Short- and long-term reliability of language fMRI. *Neuroimage* 2018; 176: 215–225.
49. Brown EC, Muzik O, Rothermel R, et al. Evaluating reverse speech as a control task with language-related gamma activity on electrocorticography. *Neuroimage* 2012; 60: 2335–2345.
50. Bardin JC, Fins JJ, Katz DI, et al. Dissociations between behavioural and functional magnetic resonance imaging-based evaluations of cognitive function after brain injury. *Brain* 2011; 134: 769–782.
51. Friston KJ, Holmes AP, Price CJ, et al. Multisubject fMRI studies and conjunction analyses. *NeuroImage* 1999; 10: 385–396.
52. Frölich MA, Banks C and Ness TJ. The effect of sedation on cortical activation: a randomized study comparing the effects of sedation with midazolam, propofol, and dexmedetomidine on auditory processing. *Anesth Analg* 2017; 124: 1603–1610.
53. Gross WL, Lauer KK, Liu X, et al. Propofol sedation alters perceptual and cognitive functions in healthy volunteers as revealed by functional magnetic resonance imaging. *Anesthesiology* 2019; 131: 254–265.
54. Huang Z, Zhang J, Wu J, et al. Disrupted neural variability during propofol-induced sedation and unconsciousness. *Hum Brain Mapp* 2018; 39: 4533–4544.
55. Kirsch M, Guldenmund P, Ali Bahri M, et al. Sedation of patients with disorders of consciousness during neuroimaging: effects on resting state functional brain connectivity. *Anesth Analg* 2017; 124: 588–598.
56. Liu X, Lauer KK, Ward BD, et al. Propofol disrupts functional interactions between sensory and high-order processing of auditory verbal memory. *Hum Brain Mapp* 2012; 33: 2487–2498.
57. Huang Z, Vlisides PE, Tarnal VC, et al. Brain imaging reveals covert consciousness during behavioral unresponsiveness induced by propofol. *Sci Rep* 2018; 8: 13195.

58. Tzovara A, Rossetti AO, Juan E, et al. Prediction of awakening from hypothermic postanoxic coma based on auditory discrimination. *Ann Neurol* 2016; 79: 748–757.
59. Claassen J, Doyle K, Matory A, et al. Detection of brain activation in unresponsive patients with acute brain injury. *N Engl J Med* 2019; 380: 2497–2505.
60. Botvinik-Nezer R, Holzmeister F, Camerer CF, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 2020; 582: 84–88.
61. Wurnig MC, Rath J, Klinger N, et al. Variability of clinical functional MR imaging results: a multicenter study. *Radiology* 2013; 268: 521–531.
62. Wilson SM and Iacoboni M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* 2006; 33: 316–325.
63. Greer DM, Rosenthal ES and Wu O. Neuroprognostication of hypoxic-ischaemic coma in the therapeutic hypothermia era. *Nat Rev Neurol* 2014; 10: 190–203.