doi:10.1093/brain/aww063 BRAIN 2016: 139; 1–4 | e32

BRAIN
A JOURNAL OF NEUROLOGY

LETTER TO THE EDITOR

Reply: Replicability and impact of statistics in the detection of neural responses of consciousness

Marzia De Lucia[1] and Athina Tzovara[1,2,3]

1 Laboratoire de Recherche en Neuroimagerie (LREN), Department of Clinical Neuroscience, Lausanne University and University Hospital, Lausanne, CH-1011, Switzerland
2 Department of Psychiatry, Psychotherapy, and Psychosomatics, University of Zurich, CH-8032, Switzerland
3 Neuroscience Centre Zurich University of Zurich, CH-8032, Switzerland

Correspondence to: Marzia De Lucia, Laboratoire de Recherche en Neuroimagerie (LREN), Department of Clinical Neuroscience, Lausanne University and University Hospital, Lausanne, CH-1011, Switzerland
E-mail: Marzia.De-Lucia@chuv.ch

Sir,

The investigation of preserved neural functions in comatose patients and their link to long-term outcome is largely based on the analysis of electrophysiological measurements at the scalp (Morlet and Fischer, 2014; Juan *et al.*, 2015). Because the analysis and interpretation of these data can influence treatments and the patients' recovery, the introduction of quantitative methods for the analysis of EEG has been increasingly recognized as a major advancement for a systematic use of evoked activity measurements in the clinical domain (Lodder and van Putten, 2013; Noirhomme *et al.*, 2014, 2015; Rossetti *et al.*, 2014; Hermans *et al.*, 2016). In this context, the letter by Gabriel *et al.*, (2016) raises an important point of discussion about the reliability and replicability of the methods for measuring the neural correlates of violation detection in mismatch negativity paradigms (Garrido *et al.*, 2009). In particular the letter highlights the inconsistency between different approaches in detecting such neural correlates at the single subject level using the same dataset. These results question the reliability of previous studies in detecting differential brain activity in response to different sensory stimuli and stimulate a debate around the 'best' analysis method for EEG recordings, particularly in a clinical setting. In our reply we would like to contribute to this debate by drawing attention to the possible sources of such discrepancies and how they could be taken into account for allowing a fair comparison. First, these various analyses stem from different hypotheses about the source of the neural response to sensory stimuli. These hypotheses lead to different choices of the EEG features undergoing the statistical analysis and may impact the final results. Within the methods selected in the letter by Gabriel and co-authors, the vast majority rely on the analysis of the average event-related potentials (ERPs) (Fischer *et al.*, 1999; Naccache *et al.*, 2005; Qin *et al.*, 2008; Daltrozzo *et al.*, 2009). This approach mainly relies on the assumption that the neural response to external sensory stimuli is the result of a series of transient post synaptic activities elicited at fixed latencies from stimulus onset and focuses on differential effects between conditions in terms of amplitude modulation. More specifically, clinical studies based on mismatch negativity protocols rely on the occurrence of specific components (significant amplitude modulation with respect to baseline or zero), and how they are modulated in specific experimental conditions in terms of polarity, peak values and latencies (Fischer *et al.*, 1999, 2004, 2008; Naccache *et al.*, 2005; Daltrozzo *et al.*, 2009; Faugeras *et al.*, 2012). Of note, these previous studies take implicitly into consideration the supposed similarity of the component features (i.e. N100 and mismatch negativity components) between control subjects and patients. In addition the polarity of relevant components at predefined latencies and electrode locations may be considered as patient inclusion criteria before assessing the occurrence of a mismatch negativity (Fischer *et al.*, 1999). All methods relying on average ERP analysis are therefore similar in the principles but differ on the strategy for assessing the presence of a significant modulation and also on the statistical analysis and its acceptance threshold. In the comparison between these different methods, it is

Advance Access publication March 26, 2016

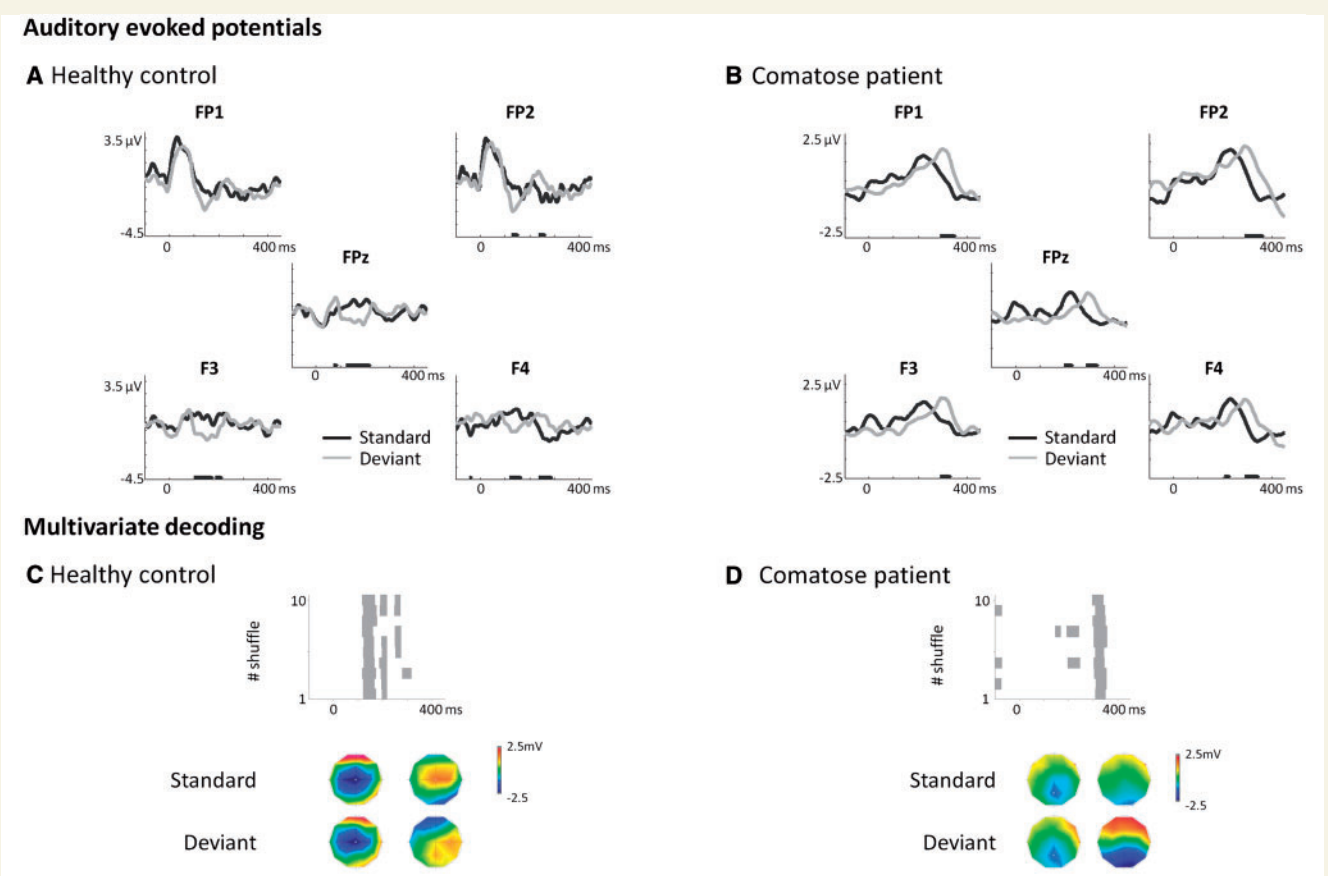

Figure 1 Single-electrode EEG responses to standard and deviant sounds and statistical analysis in a healthy control subject and an exemplar coma patient. (**A** and **B**) Average auditory evoked potentials in response to standard (black lines) and duration deviant (grey lines) sounds. The thick lines on the *x*-axes highlight periods of significant difference between the two responses (unpaired time-point by time-point *t*-tests; $P < 0.05$). (**C** and **D**) Periods of significant difference and corresponding voltage topographies, revealed by the single-trial topographic analysis, taking into account the distribution of the response across all sensors. The periods of significant difference refer to 10 different splits of the data, evaluated in a cross-validation procedure (Tzovara *et al.*, 2012*a*).

important to consider that the statistical analysis on components-related features should be evaluated separately from the technique that is implemented for assessing the EEG component [i.e. cross correlation in Fischer *et al.* (1999) or t-continuous wavelet transform in Daltrozzo *et al.* (2009)]. In this direction we would expect a high similarity between the results across individuals by adjusting the sensitivity of the statistical analysis across studies. One way to compare on a fair basis these different approaches is to rank the subjects based on the presence of the mismatch negativity at different levels of statistical threshold for each method. The analysis of the consistency of this ranking would give a better estimate of the reproducibility of the results across methods.

The last method included in the study (Tzovara *et al.*, 2012*a*) stems from a different model of the relevant EEG features for ERP estimation and the comparison with the previous results is therefore less straightforward. This method aims at extracting voltage topographies and time intervals that are mostly discriminative between conditions in terms of classification performance. Importantly, this type of decoding method circumvents the limitation of defining *a priori* inclusion criteria for subjects/patients data analysis, and privileges data-driven features selection (Blankertz *et al.*, 2011). In particular in the case of electrophysiological studies of comatose patients it is well known that the ERPs at the single patient level exhibit striking differences in the stereotypical ERP responses to basic sensory stimuli (Fischer *et al.*, 1999) and especially under hypothermia and sedation (Fig. 1B and Madhok *et al.*, 2012). That said, as a sanity check, we have previously shown that the discriminative intervals extracted in a data-driven manner by the single-trial topographic analysis do typically overlap with the differential time periods estimated at the average ERP level (Tzovara *et al.*, 2012*a*). In our experience data collected in comatose patients showed that this multivariate analysis does not always convey the same information as univariate statistics at the single electrode level (Tzovara *et al.*, 2013; De Lucia and Tzovara, 2015*b*). We provide here the results of the analysis of the EEG response to standard and duration deviant sounds with two types of analyses, the time-point by

time-point *t*-tests and the single-trial topographic analysis (Tzovara *et al.*, 2012*a*, *b*) for a healthy control subject and an exemplar patient in the first day of coma (Fig. 1). In the case of the control subject, a typical mismatch negativity-like response can be observed in most frontal electrodes, starting already at ~150 ms post-stimulus onset (Fig. 1A). However, the exact latency of this response strongly depends on the electrode that is examined. When considering the single-trial topographic analysis, the periods of significant difference are consistent with the effect observed in most electrodes as they contain temporal information that can be conveyed across all electrodes (Fig. 1C). A similar image is also seen in the case of an exemplar coma patient, with the difference that the discriminant latencies appear much later, starting at ~300 ms post-stimulus onset, possibly as a result of coma or hypothermic treatment (Fig. 1B and D).

In light of these findings, we agree with Gabriel and co-authors that inconsistences in statistical results can be puzzling, especially when these results might be used in clinical practice and in the detection of consciousness. Our proposition is that the selection of a given method (and all the fine tuning of the methods parameters) should be considered 'context specific' and should be based on an extensive and systematic validation. The single-trial topographic analysis has been proven successful in the assessment of violation detection in post-anoxic comatose patients treated with therapeutic hypothermia. The reliability of these results is supported by their highly significant predictive power of patients' outcome when looking at the improvement over 2 days of the decoding standard and deviant sounds in the same patients (Tzovara *et al.*, in press). These results have encouraged further application of the same decoding analysis to EEG data recorded in the patients with the same aetiologies (Cossy *et al.*, 2014; Tzovara *et al.*, 2015) and particularly using similar experimental protocols that have been proposed by other groups for the detection of conscious processing in healthy and clinical populations (Bekinschtein *et al.*, 2009; Faugeras *et al.*, 2011, 2012). Particularly, the results obtained by this decoding approach using the local/global paradigms have suggested that violation detection in reduced consciousness might be a dynamic phenomenon that is preserved during the first days of coma and degenerates in patients with poor outcome (Piarulli *et al.*, 2015; Tzovara *et al.*, 2015). The validation of the single-trial topographic analysis in different contexts of disorders of consciousness and different coma aetiologies is the goal of future studies.

Many other solutions for multivariate analysis of ERPs at the single trial level have been proposed, particularly in the field of Brain Computer Interface (Muller *et al.*, 2008), in fundamental neuroscience research for detailing the relation between single trial EEG and behaviour in decision-making studies (Ratcliff *et al.*, 2009) and in the field of disorders of consciousness (Noirhomme *et al.*, 2015). Because of the richness and variety of the proposed methods, the use of a generic terminology, such as multivariate analysis, should be considered with care. Indeed in previous studies from our group, the application of other decoding analysis to the EEG data recorded in post-anoxic patients during the same mismatch negativity protocol did not provide the same predictive value (De Lucia and Tzovara, 2015*a*).

In summary, in light of recent evidence questioning the reliability and replicability of human research findings in general (Open Science, 2015), we would like to emphasize the importance of validating results in the same clinical populations across laboratories and encouraging meta-analysis of published results based on similar experimental protocols (Uttal, 2014).

Acknowledgements

We thank Dr Andrea Rossetti and Dr Mauro Oddo for exceptional clinical support in all conducted studies on comatose patients.

Funding

This work was supported by the EUREKA's Eurostars Programme (Project title and number: ComAlert E!9361 to M.D.L.).

References

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L. Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA 2009; 106: 1672–7.

Blankertz B, Lemm S, Treder M, Haufe S, Muller KR. Single-trial analysis and classification of ERP components–a tutorial. Neuroimage 2011; 56: 814–25.

Cossy N, Tzovara A, Simonin A, Rossetti AO, De Lucia M. Robust discrimination between EEG responses to categories of environmental sounds in early coma. Front Psychol 2014; 5: 155.

Daltrozzo J, Wioland N, Mutschler V, Lutun P, Calon B, Meyer A, et al. Cortical information processing in coma. Cogn Behav Neurol 2009; 22: 53–62.

De Lucia M, Tzovara A. Decoding auditory EEG responses in healthy and clinical populations: a comparative study. J Neurosci Methods 2015a; 250: 106–13.

De Lucia M, Tzovara A. Prognostic use of cognitive event-related potentials in acute consciousness impairment. In: Rossetti AO, Laureys S, editors. Clinical neurophysiology in disorders of consciousness. Vienna, Austria: Springer; 2015b. p. 81–93.

Faugeras F, Rohaut B, Weiss N, Bekinschtein T, Galanaud D, Puybasset L, et al. Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. Neuropsychologia 2012; 50: 403–18.

Faugeras F, Rohaut B, Weiss N, Bekinschtein TA, Galanaud D, Puybasset L, et al. Probing consciousness with event-related potentials in the vegetative state. Neurology 2011; 77: 264–8.

Fischer C, Dailler F, Morlet D. Novelty P3 elicited by the subject's own name in comatose patients. Clin Neurophysiol 2008; 119: 2224–30.

Fischer C, Morlet D, Bouchet P, Luaute J, Jourdan C, Salord F. Mismatch negativity and late auditory evoked potentials in comatose patients. Clin Neurophysiol 1999; 110: 1601–10.

Fischer C, Morlet D, Luaute J. Sensory and cognitive evoked potentials in the prognosis of coma. Suppl Clin Neurophysiol 2004; 57: 656–61.

Garrido MI, Kilner JM, Stephan KE, Friston KJ. The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol 2009; 120: 453–63.

Gabriel D, Muzard E, Henriques J, Mignot C, Pazart L, Andre'-Obadia A, et al. Replicability and impact of statistics in the detection of neural responses of consciousness. Brain 2016; 139: e30.

Hermans MC, Westover MB, van Putten MJ, Hirsch LJ, Gaspard N. Quantification of EEG reactivity in comatose patients. Clin Neurophysiol 2016; 127: 571–80.

Juan E, Kaplan PW, Oddo M, Rossetti AO. EEG as an indicator of cerebral functioning in postanoxic coma. J Clin Neurophysiol 2015; 32: 465–71.

Lodder SS, van Putten MJ. Quantification of the adult EEG background pattern. Clin Neurophysiol 2013; 124: 228–37.

Madhok J, Wu D, Xiong W, Geocadin RG, Jia X. Hypothermia amplifies somatosensory-evoked potentials in uninjured rats. J Neurosurg Anesthesiol 2012; 24: 197–202.

Morlet D, Fischer C. MMN and novelty P3 in coma and other altered states of consciousness: a review. Brain Topogr 2014; 27: 467–79.

Muller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J Neurosci Methods 2008; 167: 82–90.

Naccache L, Puybasset L, Gaillard R, Serve E, Willer JC. Auditory mismatch negativity is a good predictor of awakening in comatose patients: a fast and reliable procedure. Clin Neurophysiol 2005; 116: 988–9.

Noirhomme Q, Brecheisen R, Lesenfants D, Antonopoulos G, Laureys S. "Look at my classifier's result": disentangling unresponsive from (minimally) conscious patients. Neuroimage 2015, http://dx.doi.org/10.1016/j.neuroimage.2015.12.006.

Noirhomme Q, Lehembre R, Lugo Zdel R, Lesenfants D, Luxen A, Laureys S, et al. Automated analysis of background EEG and reactivity during therapeutic hypothermia in comatose patients after cardiac arrest. Clin EEG Neurosci 2014; 45: 6–13.

Open Science C. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science 2015; 349: aac4716.

Piarulli A, Charland-Verville V, Laureys S. Cognitive auditory evoked potentials in coma: can you hear me? Brain 2015; 138: 1129–37.

Qin P, Di H, Yan X, Yu S, Yu D, Laureys S, et al. Mismatch negativity to the patient's own name in chronic disorders of consciousness. Neurosci Lett 2008; 448: 24–8.

Ratcliff R, Philiastides MG, Sajda P. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad Sci USA 2009; 106: 6539–44.

Rossetti AO, Tzovara A, Murray MM, De Lucia M, Oddo M. Automated auditory mismatch negativity paradigm improves coma prognostic accuracy after cardiac arrest and therapeutic hypothermia. J Clin Neurophysiol 2014; 31: 356–61.

Tzovara A, Murray M, Plomp G, Herzog M, Michel C, Lucia MD. Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. Pattern Recog 2012a; 45: 2109–22.

Tzovara A, Murray MM, Michel CM, De Lucia M. A tutorial review of electrical neuroimaging from group-average to single-trial event-related potentials. Dev Neuropsychol 2012b; 37: 518–44.

Tzovara A, Rossetti AO, Juan E, Suys T, Viceic D, Rusca M, et al. Prediction of awakening from hypothermic post anoxic coma based on auditory discrimination. Ann Neuro. Accepted Article, doi: 10.1002/ana.24622

Tzovara A, Rossetti AO, Spierer L, Grivel J, Murray MM, Oddo M, et al. Progression of auditory discrimination based on neural decoding predicts awakening from coma. Brain 2013; 136: 81–9.

Tzovara A, Simonin A, Oddo M, Rossetti AO, De Lucia M. Neural detection of complex sound sequences in the absence of consciousness. Brain 2015; 138: 1160–6.

Uttal WR. Reliability in cognitive neuroscience: a meta-meta-analysis. Chicago: University of Chicago Press; 2014.