Perspectives on functional and hyperkinetic movement disorders
Phenomenology & pathophysiology
van der Salm, S.M.A.

Citation for published version (APA):
Chapter 2

THE EYE OF THE BEHOLDER:
INTER-RATER AGREEMENT AMONG
EXPERTS ON PSYCHOGENIC JERKY
MOVEMENT DISORDERS

S.M.A. van der Salm
R.J. de Haan
D.C. Cath
A.F. van Rootselaar
M.A.J. Tijssen

ABSTRACT

OBJECTIVE
The current criteria for conversion disorder in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) rely on the assumption that neurological disorders can be distinguished from conversion disorders through clinical assessment. This study aims to assess inter-rater agreement of clinicians with experience in the diagnosis of various hyperkinetic jerky movements, including psychogenic jerks.

METHODS
Sixty patients with either psychogenic jerks, myoclonus, or tics were rated by international experts using a standardized survey resembling daily clinical practice. The survey included the following diagnostic steps; a short video offering a visual impression of the patients and their jerky movements, medical history, neurological examination (on video), additional investigations and the findings of a standardized psychiatric interview. The diagnosis and diagnostic certainty were scored after each step.

RESULTS
After all clinical information was given, moderate inter-rater agreement was reached (kappa=0.56 ± 0.1) with absolute agreement (100%) on the diagnosis in 12 (20%) patients and reasonable agreement (>75%) in 43 (72%) patients. Psychiatric evaluation did not contribute to inter-rater agreement or diagnostic certainty.

CONCLUSIONS
Our findings illustrate that experienced movement disorder specialists moderately agree with the clinical diagnosis of jerky movements. Clinical assessment, especially by a team of clinicians in challenging individual cases, might improve diagnostic agreement.
ABSTRACT

OBJECTIVE

The current criteria for conversion disorder in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) rely on the assumption that neurological disorders can be distinguished from conversion disorders through clinical assessment. This study aims to assess inter-rater agreement of clinicians with experience in the diagnosis of various hyperkinetic jerky movements, including psychogenic jerks.

METHODS

Sixty patients with either psychogenic jerks, myoclonus, or tics were rated by international experts using a standardized survey resembling daily clinical practice. The survey included the following diagnostic steps; a short video offering a visual impression of the patients and their jerky movements, medical history, neurological examination (on video), additional investigations and the findings of a standardized psychiatric interview. The diagnosis and diagnostic certainty were scored after each step.

RESULTS

After all clinical information was given, moderate inter-rater agreement was reached (kappa=0.56 ± 0.1) with absolute agreement (100%) of experts on the diagnosis in 12 (20%) patients and reasonable agreement (>75%) in 43 (72%) patients. Psychiatric evaluation did not contribute to inter-rater agreement or diagnostic certainty.

CONCLUSIONS

Our findings illustrate that experienced movement disorder specialists moderately agree with the clinical diagnosis of jerky movements. Clinical assessment, especially by a team of clinicians in challenging individual cases, might improve diagnostic agreement.

INTRODUCTION

Conversion disorders encompass symptoms or deficits that affect voluntary motor or sensory function suggestive of a neurological or general medical condition. Psychogenic movement disorders (PMD) are a specific subtype of conversion disorder which cannot be explained by any known movement disorder nor attributed to a structural lesion of the nervous system. PMDs are generally considered to be associated with psychological stressors, although not always demonstrated. Symptoms can mimic the full range of organic involuntary movements. The phenomenological similarities between PMDs and organic movement disorders such as myoclonus and tics carry the risk of misdiagnosis and inadequate treatment of patients.

Establishing the diagnosis of PMD is an intricate diagnostic process since there is no ‘gold standard’ diagnostic test. Over two decades ago, Fahn and Williams proposed the first PMD-diagnostic criteria. Since then, the criteria have been critically revised. Clinical evaluation focuses on positive criteria for PMD such as atypical progression, improvement of motor function by distraction, entrainment with voluntary movements, and suggestibility of symptoms. In theory, based on the current criteria a distinction between psychogenic movement disorders and their organic counter parts can be made. However, a survey amongst movement disorder specialists demonstrated a range of opinions regarding the diagnosis of PMDs and a lack of consensus on differentiating criteria with other movement disorders. Thus in clinical cases encountered in daily practice, the distinction may, to varying degrees, be subject to the opinion of the clinician.

Jerky movement disorders can range from tics and myoclonus to psychogenic jerks. In contrast to myoclonus and psychogenic jerks, most patients with motor tics appreciate some degree of urge and volition prior to the performance of many of their movements. Differentiation of motor tics and psychogenic jerks can be particularly difficult as patients share the ability to suppress jerks, and because the jerk severity is influenced by psychological factors and distraction in both conditions. Moreover, there is considerable overlap in psychiatric co-morbidity in jerky movement disorders, such as anxiety disorders, obsessive-compulsive disorder and major depressive disorder. The differentiation between myoclonus and psychogenic jerks is often aided by electrophysiological testing for the Bereitschaftspotential (BP), a movement-related potential indicative of premotor cortex activation. The presence of a BP indicates that the movement is not myoclonus, and it is therefore used to establish a “laboratory-supported” diagnosis of psychogenic jerky movements. It is controversial whether motor tics are preceded by a short BP in some patients.

The main aim of this study is to assess the agreement among experienced clinicians on the diagnosis of psychogenic jerks. Second, to explore which clinical data and electrophysiological and a standardised psychiatric diagnostic assessment anchor accurate diagnosis and inter-rater agreement.
METHODS

PATIENTS

A case series of sixty patients with hyperkinetic movement disorders participated in the study. We included participants from 3 different sources: 1) Between 2007 and 2010, all consecutive patients seen with jerks in the outpatient clinic of the department of Neurology of the Academic Medical Center in Amsterdam, the Netherlands, who were clinically considered to have either myoclonus, psychogenic jerks, or motor tics by the hyperkinetic movement disorder specialists were included in this study. 2) In addition, we searched our movement disorders registry of patients seen between 2000 and 2007 with myoclonus, psychogenic jerks, or motor tics. 3) Medication-free patients referred from the Tourette Syndrome Patients’ Association with primarily motor tics.

EXPERTS

Experts were selected based on scientific ranking on ISI Web of Knowledge (http://apps.webofknowledge.com, October 2010). We searched experts with the key words ‘tics’, ‘Tourette syndrome’, ‘myoclonus’, ‘conversion disorder’ and ‘psychogenic movement disorder’. We ranked the number of articles published by the expert on each subject. For each of the three topics (myoclonus, PMD, tics) a list was created and we invited the top 20 experts within each specific area of expertise and sent them an invitation for this study via email. Moreover, we asked the experts identified in this way who they considered to be experts themselves and invited additional experts with multiple recommendations. All participating experts gave information on their academic background including self-rated area of expertise (myoclonus, psychogenic, tic) and specialization.

RANDOMISATION OF PATIENTS

We divided the patients in four equal groups (15 patients per group). About half of the included patients had clinical features to suggest psychogenicity (revised Fahn and Williams criteria).(4,6) The remaining patients (tic and myoclonus) were equally distributed by the organising team. Patients were further divided into groups of truncal, arm or leg jerks and these groups were randomly assigned to the 4 datasets.

FIXED ORDER ONLINE SURVEY

All patients underwent a standardised clinical work-up. In the structured online survey, we serially presented 1) a short video to provide a first visual impression of the patient, 2) the patient’s medical history, 3) neurological examination (including distractibility, suppression and release of jerks) on video, 4) electrophysiological data (Bereitschaftspotential), and finally 5) the results of the diagnostic interview for psychopathology (Mini-International Neuropsychiatric Interview-Plus (MINI-Plus)) by a neuropsychologist, blinded for subject status.(13) At each step the experts were asked about their most likely diagnosis (‘Diagnosis at step 1-4’ Figure 1). Experts could either choose myoclonus, psychogenic jerk or tic. At the end of each single case, after the provision of the psychiatric assessment, experts were
asked to choose their final diagnosis from the same list of options. After each diagnostic step, experts were also asked about the degree of certainty of their diagnosis on a 5 point Likert scale, ranging from 1 = very uncertain to 5 = absolutely certain. Participants were unable to return to previous answers and had to finish the review of an entire case at once, but could log-out between cases. Invitations were sent in November 2010, and the video study was accessible online from December 1st 2010 until March 1th 2011. Details of the survey can be found in the Supplementary Methods section.

Figure 1 Overview of the study design.
Chapter 2

**STATISTICS**
It should be noted that the clinical diagnoses of our patients as made by the organising team were not included in the statistic calculations and results of this study. Agreement between the clinical experts was expressed as the proportion of observed agreement and chance-corrected agreement using Fleiss’ Kappa (κ). Fleiss' Kappa is a statistical measure specifically designed for assessing the agreement among multiple raters and multiple categories.(14) Fleiss’ Kappa values can be interpreted as; κ <0 indicating no agreement, 0–.20 slight, .21–.40 fair, .41–.60 moderate, .61–.80 substantial, and .81–1 as excellent agreement. Between group statistics were performed with non-parametric Mann-Whitney U and Kruskal–Wallis tests. Difference in the degree of diagnostic certainty in relation to additional diagnostic information was analysed using the Freidman test and the Wilcoxon signed rank test when appropriate. Statistical tests were conducted using the Statistical Package for the Social Sciences (SPSS) software 16.0 and reported with a significance threshold of p< 0.05.

**APPROVAL AND ETHICS**
The study was approved by the medical ethics committee of the Academic Medical Center in Amsterdam, the Netherlands. All patients gave written informed consent to accessibility of their (blinded) medical data, including video recordings, on the hospital server via a secure internet connection. All patients were fully informed about the background and aim of the study. All clinical experts signed a non-disclosure agreement.
RESULTS

EXPERTS

Out of the scientifically ranked experts (n=60), a total of 35 experts initially agreed to take part in the study. These participating experts suggested 92 additional experts, of whom 27 experts were suggested more than once. Of this latter group, 20 experts had already been included in the original list of experts. The remaining 7 experts were invited and agreed to take part. Thirty-nine of the 42 experts who signed the consent form finished the review process. The primary reason given by experts to discontinue the review process was lack of time.

The final 39 experts included 37 neurologists (including two neurologists trained in Paediatric Neurology) and two psychiatrists (including one neuropsychiatrist). Twenty three out of 39 participants annotated membership of the international professional society for Movement disorders (MDS). Fifteen participants were annotated by the organizing team and themselves as experts on myoclonus, 10 on psychogenic movement disorders (all MDS members) and 14 on tic disorders. We assigned experts randomly to one of the four review sets, each containing 15 cases to review, and took care that experts working at the same institute or with the same amount of clinical experience would get different sets to review. Thus, each dataset of 15 patients was rated by 9 to 11 experts.

OBSERVED AGREEMENT ON DIAGNOSIS

An absolute 100% observed agreement between all experts was reached for 6 cases (10%) after the first impression and for 12 cases (20%) at the final diagnosis. Agreement between 75% of experts was reached in 18 cases after the first impression and in 43 cases (72% of cases) at the final diagnosis. A greater agreement than 50% of experts was reached in 56 cases after the first impression and in almost all (58; 97%) of the cases at the final diagnosis (see Table 1). Agreement increased with incremental supply of diagnostic information (from first impression to final diagnosis). A post-hoc analysis demonstrated no significant influence of personal expertise on the frequency of specific diagnosis (p = 0.70).
Table 1 displays number of cases (total n=60) with agreement of experts over diagnoses. Observed agreement is expressed as absolute (100%), >75% and >50% of the experts raters.

**Inter-rater agreement**

Chance-corrected agreement amongst clinical experts after the first impression was fair and final agreement was moderate (see Figure 2A) (in detail: first impression $\kappa = 0.28 \pm 0.14$ (SD); medical history $\kappa = 0.51 \pm 0.07$; neurological exam $\kappa = 0.54 \pm 0.11$; electrophysiology $\kappa = 0.56 \pm 0.09$, and final diagnosis after psychiatric evaluation $\kappa = 0.56 \pm 0.1$). The agreement increased most markedly from the first impression towards the medical history.

At first impression agreement on myoclonus was fair ($\kappa = 0.29 \pm 0.13$) and at final diagnosis after psychiatric evaluation the Fleiss’ Kappa was moderate ($\kappa = 0.46 \pm 0.17$) (details: medical history $\kappa = 0.40 \pm 0.11$; neurological exam $\kappa = 0.43 \pm 0.18$; and electrophysiology $\kappa = 0.46 \pm 0.19$). Psychogenic jerks had a fair inter-rater agreement ($\kappa = 0.22 \pm 0.16$) at first impression and moderate ($\kappa = 0.54 \pm 0.09$) at final diagnosis (medical history $\kappa = 0.47 \pm 0.12$; neurological exam $\kappa = 0.5 \pm 0.12$; $\kappa = 0.54 \pm 0.08$). Finally, for tics the Fleiss’ Kappa was fair ($\kappa = 0.32 \pm 0.18$) after the first impression and substantial ($\kappa = 0.68 \pm 0.11$) at the final diagnosis (medical history $\kappa = 0.62 \pm 0.08$; neurological exam $\kappa = 0.68 \pm 0.10$; and electrophysiology $\kappa = 0.68 \pm 0.08$). Overall, when calculated per diagnosis, the inter-rater agreement of tics was substantially higher than that of psychogenic jerks. The inter-rater agreement of myoclonus was the lowest.
DEGREE OF DIAGNOSTIC CERTAINTY
Incremental information increased the degree of certainty of the experts, which was relatively low at 1.9 (± 0.97) at first impression and 3.5 (± 1.22) at final diagnosis (p < 0.001) (medical history 2.5 ± 1.15; neurological exam 3.2 ± 1.2; and electrophysiology 3.4 ± 1.2). The addition of medical history, neurological exam and electrophysiological data did alter the degree of certainty (all p < 0.001), but the addition of psychiatric data (prior to final diagnosis) did not (p = 0.323).

Figure 2 Inter-rater agreement

\[ \text{Kappa} \]

Figure 2A: Mean (SD) values for κ as a function of the incremental provision of diagnostic medical information. Agreement after the first agreement was fair (κ=0.28±0.14) and final agreement was moderate (κ=0.56 ±0.1). Figure 2B: Inter-rater agreement, expressed as mean (SD) κ per diagnosis as a function of the incremental provision of diagnostic medical information.
Figure 3 Diagnostic certainty of expert panel

Figure 3 displays the mean (SD) degree of diagnostic certainty in relation to the incremental diagnostic information provided.
DISCUSSION
The current study demonstrates that the inter-rater agreement on diagnosis by experts after the first video evaluation combined with BP measurements and psychiatric evaluation was moderate ($\kappa = 0.56$) with >75% observed agreement in 72% of cases. The addition of the patient’s history after the first impression resulted in the greatest increase in inter-rater agreement. The diagnostic degree of certainty improved further by the subsequent addition of the medical history, neurological examination and electrophysiology. Interestingly, psychiatric evaluation did not increase the inter-rater agreement nor did it improve the degree of certainty. The inter-rater agreement and diagnostic certainty concomitantly increase with the addition of diagnostic information.

The established diagnostic criteria for conversion disorder rely on the assumption that conversion disorders can be distinguished from neurological disorders through clinical assessment. Although the inter-rater agreement between experts on jerky movement disorders was only moderate, particularly in the diagnosis of myoclonus, the agreement increases as additional information becomes available and would possibly markedly increase if the experts had the opportunity to personally evaluate the patients rather than rely on the phenomenology captured on the video. It is possible that the diagnosis of myoclonus is challenging as most subtypes of myoclonus are rare. Moreover, there are different classifications of myoclonus and the standard terminology may not apply to psychogenic myoclonus.\((15,16)\) For instance, spontaneous axial flexion jerks have been termed ‘propriospinal myoclonus’, suggesting spinal generation of the jerk. However, it has been established recently that a proportion of cases could be classified as psychogenic or idiopathic spinal myoclonus.\((17,18)\) The aforementioned example illustrates the lack of uniform classification for movement disorders in general and PMD in particular. This problem was illustrated in dystonia, where in various types of organic dystonia, agreement on diagnosis was even lower than the rates in our study of jerky movements.\((19)\) In certain subtypes of dystonia the agreement of the diagnosis was only moderate (blepharospasm $\kappa = 0.51$, and cervical dystonia $\kappa = 0.52$), fair (writer’s cramp $\kappa = 0.29$), and slight (oromandibular dystonia $\kappa = 0.20$). The relatively good inter-rater agreement on the diagnosis of tics in our study ($\kappa = 0.68$) might be explained by the overall high frequency of tics in the general population and the well-defined diagnostic criteria.\((1,8)\)

The moderate inter-rater agreement on psychogenic jerks in our study contrasts with a recent study of YouTube videos where seven neurologists evaluated the most frequently viewed YouTube videos of patients with a movement disorder and were asked to judge whether the movement disorder depicted appeared to be psychogenic or organic.\((20, 21)\) Of 29 videos showing people with various movement disorders, 66% (19 out of 29) were rated as psychogenic and 34% as organic in aetiology. Interestingly, an excellent inter-rater agreement ($\kappa = 0.89$) was found. The discrepancy between their high inter-rater agreement and our numbers are likely to be due to differences in selection and presentation of video
tapes. We used a case series encountered in daily practice, whereas the videotapes of patients from YouTube may have been more extreme cases in terms of presentation and phenomenology. In keeping with the findings of our study, another small sample video-based study of various subtypes of PMD and a video-EEG recordings of another frequent neurological conversion disorder, psychogenic non-epileptic seizures both found moderate inter-rater agreements. (22,23) Hence, in clinical cases encountered in daily practice, the diagnosis of both neurological or psychogenic movement disorders appears to be difficult even for experts in the field.

The addition of the history of the patient to the first impression resulted in the most pronounced increase in inter-rater agreement of the diagnosis. Based on the first impression, experts tended to diagnose myoclonus rather than psychogenic jerks, as one would expect without sufficient documentation. This initial inclination was reversed after all diagnostic information had been provided. The diagnosis of PMD is mainly based on the patients’ history, but the certainty of the diagnosis improved with both history and neurological examination. The provision of the Bereitschaftspotential did not result in much greater inter-rater agreement, but did provide a higher degree of diagnostic certainty. Literature on the diagnostic value currently limited, which might explain the low value attributed to presence or absence of the BP by experts in their clinical decision making. Recently, we have published the findings of the BP in this cohort. (12) Patients with psychogenic jerks significantly more frequent have a BP prior to their jerks and with a significantly earlier onset compared to GTS patients, although replication in a prospective cohort is warranted.

An alternative explanation for the low decisive value of the BP could be the particular setup of our study, in which the electrophysiological and psychiatric evaluation was provided at the final steps, and earlier disclosure of this information might have resulted in a higher diagnostic value. However, it has been shown previously in a survey of movement disorders specialists that psychiatric evaluation is not regarded as the decisive factor in the diagnosis of PMD. (7) In a recent study, it was demonstrated that the association with psychological issues is less evident than hypothesized. (3) In line with this notion, the proposed revised criteria for the DSM-5 have suggested to remove the prerequisite of an identifiable psychological aetiology. (24)

The moderate agreement among experts suggests that additional useful aspects of the usual consultation process, such as (non-) verbal communication, might be essential for a proper diagnosis. It may be essential for movement disorders, and especially PMD, to have a direct personal contact between the patient and the doctor. For example, the way patients present their case, the presence of suggestibility (improving or worsening the movements) or other features often enhance the certainty of a diagnosis of PMD. (25,26) It has been shown that interactional and linguistic examination of the communication between neurologists and patients during a first encounter can contribute to the differential
diagnosis of epilepsy or psychogenic non-epileptic seizures (PNES). (27) With regard to non-verbal communication, the ‘hands-on’ neurological examination provides the possibility for the observation of spontaneous movement and body language in real life clinical situations. Hence, it may be of interest to study the (non)-verbal communication and diagnostic agreement during personal evaluation of cases in further detail.

Inter-rater agreement studies are important to clarify discrepancies in the interpretation of clinical criteria that lead to diagnostic disagreement. This is especially important in the diagnosis of PMD as there is no ‘gold standard’. Expert opinions and practices related to diagnosis and management of PMD patients differ considerably among movement disorders neurologists, as shown previously. (7) The results of this study clearly reflect these discrepancies and call for the development of practice guidelines, possibly via consensus meetings. For example, the establishment of diagnostic criteria, such as the use of panels of raters, have improved diagnostic certainty in the field of epilepsy. (28) Although the importance of clinical guidelines seems obvious, future studies need to assess to what extent they will add to inter-rater agreement.

LIMITATIONS OF THE STUDY
In this study, international experts were invited to participate. As a consequence some experts experienced time constraints, reflected in the response rate and the inability to finish the full review process. We have specifically chosen not to include an ‘uncertain’ category to the differential diagnosis, which may have influenced the agreement on the other diagnoses as in case of uncertainty, experts were obliged to choose a diagnostic category. Interestingly, although the experts disagreed, most of them were certain (mean degree of certainty 3.5). Their rather high level of certainty at their final diagnosis is an argument against a subcategory of ‘uncertain’ in the diagnostic differential diagnosis. Moreover, we chose a standardised instead of a randomized order of information presentation. The latter would test the additional value of the psychiatric and electrophysiological information; however, this would not reflect usual order in the daily clinical practice, which was the primary aim of this study. A fundamental limitation is the lack of a ‘gold standard’ for the diagnosis of psychogenic jerks, and therefore we could only assess concordance of opinion amongst expert raters. The organising team specifically chose not to designate their own diagnosis as a ‘gold standard’, though we did use our diagnostic categories to balance the sample. Some additional information was available to the organising team, for instance that some patients demonstrated full resolution of their movements, either spontaneously or with suggestion. However, remissions and fluctuations also occur in case of tics. Therefore, the treatment and follow-up time was not considered sufficient to label the entire cohort with a ‘gold standard’ final diagnosis made by the organising team for the time being. For the electrophysiological testing, we provided information on the BP only and not on burst durations of EMG recordings. This might have limited the certainty of the diagnosis of myoclonus. Finally, it is possible that key psychiatric
issues or subtleties were omitted by the use of standardized psychiatric evaluation (MINI-Plus) for the assessment of psychiatric comorbidity, and that assessment of more specific psychosocial predisposing, triggering or maintaining factors would be more useful in the diagnostic process.

In conclusion, we demonstrate that the inter-rater agreement of experts in jerky movement disorders is moderate when based on a review of videos and descriptive clinical and electrophysiological information. This implies the need for caution in the diagnosis of psychogenic jerking. It needs to be acknowledged that some cases remain a considerable diagnostic challenge.

ACKNOWLEDGEMENTS


We thank Sofyan Alamudi and Lieke Wiggers for help with the online web survey.
Chapter 2

In conclusion, we demonstrate that the inter-rater agreement of experts in jerky movement disorders is moderate when based on a review of videos and descriptive clinical and electrophysiological information. This implies the need for caution in the diagnosis of psychogenic jerking. It needs to be acknowledged that some cases remain a considerable diagnostic challenge.

ACKNOWLEDGEMENTS


We thank Sofyan Alamudi and Lieke Wiggers for help with the online web survey.

REFERENCES

Chapter 2

SUPPLEMENTARY METHODS

DETAILS OF PROTOCOL OF STUDY

The clinical data were presented starting with a brief first video impression of the jerky movements (20 second video clip showing at least 2 jerks). This was followed by a standardised medical history, including tic and psychogenic features and full social and family history in all cases. A compulsory video clip (about 3 minutes) with a summary of the full neurological examination was shown and had to be viewed in full in order to proceed with the questionnaire. All patients were videotaped during rest, posture, attempting to suppress their jerks and the release thereafter. In addition, all patients were videotaped without the examiner in the room. Additional videos with details of the neurological exam were provided on demand. Electrophysiological results (presence, onset and amplitude of the Bereitschaftspotential) were provided next. In this study, the BP was defined as present if the following criteria were met. Firstly, at t0 the amplitude was at least 5 μV. Second, movement artefact should not interfere with the BP. This was defined as absence of a slow rising potential over the occipital electrodes (O1, O2). If the same deflection was present at O1 or O2, the deflection at Cz was considered to be due to artefact and not labelled as a BP. Finally, before the ultimate diagnoses were established, the results of the psychiatric evaluation were shown (MINI-PLUS).(13)

ABBREVIATED DESCRIPTION ILLUSTRATIVE CASE

A 49 year old woman presented with a 5 year history of jerks of predominantly the right hand. After a car accident, she had been in coma and one year thereafter the jerks had started subacutely. Jerks have not progressed over time and are only present for about 3 days on a regular basis solely during her menses. Jerks are most pronounced on the right hand, increases in intensity during the day and are present during sleep. She felt not able to suppress jerks, feels no premonitory sensation or relief. She had no relevant family or psychiatric history. The supplementary video demonstrates the first impression of the patient and a compilation of the neurological examination and extra video material that was available upon demand. Please note that upon request of the patient, her face was covered by a black bar to further assure patient anonymity upon publication in the Journal of Neurology, Neurosurgery and Psychiatry. However, the experts participating in the study were shown the complete video without a black bar in front of the patient’s face.

During electrophysiological investigations, a BP analysis could not be conducted as the jerks were too frequent. The psychiatric evaluation demonstrated no psychiatric co-morbidity.

We diagnosed this patient as myoclonus and 55% of the expert raters concurred. The experts raters that diagnosed the patient’s jerks as myoclonus were significantly more certain than the experts that differed in diagnosis (p:0.39; Spearman’s rho correlation coefficient: 0.628 )