Published in final edited form as:

Mov Disord. 2012 April ; 27(4): 548–552. doi:10.1002/mds.24903.

Diagnostic Agreement in Patients with Psychogenic Movement Disorders

Francesca Morgante, PhD¹, Mark J. Edwards, PhD², Alberto J. Espay, PhD³, Alfonso Fasano, PhD⁴, Pablo Mir, PhD^{5,6}, and Davide Martino, PhD^{7,8,*} on behalf of the DISMOV-SIN study group on psychogenic movement disorders^{8,†}

¹Dipartimento di Neuroscienze, Scienze Psichiatriche ed Anestesiologiche, Università di Messina, Messina, Italy

²Sobell Department of Motor Neuroscience and Movement Disorders, UCL Institute of Neurology, Queen Square, London, United Kingdom

³University of Cincinnati Neuroscience Institute, Department of Neurology, Gardner Center for Parkinson's Disease and Movement Disorders, University of Cincinnati, Cincinnati, Ohio, USA

⁴Istituto Neuromed, Pozzilli, Italy

⁵Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain

⁶Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas, Barcelona, Spain

⁷Dipartimento di Scienze Neurologiche e Psichiatriche, Università di Bari, Bari, Italy

⁸Neuroscience and Trauma Center, Barts and The London School of Medicine and Dentistry, London, United Kingdom

Abstract

Background—The reliability and applicability of published diagnostic criteria for psychogenic movement disorders (PMDs) have never been examined.

Methods—Eight movement disorder and six general neurologists rated 14 patients diagnosed with PMD and 14 patients diagnosed with organic movement disorders. Raters provided a dichotomous judgment (i.e., psychogenic or organic) upon review of video-based movement phenomenology and a category of diagnostic certainty based on the Fahn-Williams and Shill-Gerber criteria after accessing standardized clinical information. We measured interobserver agreement on the diagnosis and clinical certainty judgment of PMD.

Results—In both groups of raters, agreements were “fair” on the video-based dichotomous judgment, but improved to “substantial” after access to standardized clinical information. “Slight”

© 2012 Movement Disorder Society

*Correspondence to: Davide Martino, Dipartimento di Scienze Neurologiche e Psichiatriche, Università di Bari, Piazza Giulio Cesare 11, I-70124 Bari, Italy; davidemartino@virgilio.it.

†See Appendix for list of collaborators.

Relevant conflicts of interest/financial disclosures: Nothing to report. Full financial disclosures and author roles may be found in the online version of this article.

Supporting Information may be found in the online version of this article.

to “poor” agreement was reached for the “probable” and “possible” categories of diagnostic certainty corresponding to both diagnostic criteria.

Conclusions—Diagnosis according to clinical available criteria for PMD yields poor diagnostic agreement.

Keywords

psychogenic movement disorders; diagnostic criteria; interobserver agreement; epidemiology; phenomenology

The diagnosis of psychogenic movement disorders (PMDs) may be challenging, given the lack of accurate and reliable tests.¹ Two sets of clinical criteria, the Fahn-Williams (FW)² and Shill-Gerber (SG)³ criteria, have been developed to assist in the diagnostic process and avoid the diagnosis-by-exclusion approach to PMD. Reliability and applicability of these criteria are essential because misdiagnosis may delay the initiation of appropriate treatment, leading to poor outcomes.⁴

We measured interobserver diagnostic agreement and clinical certainty judgment (according to FW and SG criteria) of PMD among movement disorder specialists (MDS) and general neurologists (non-MDS). We also examined whether diagnostic agreement based exclusively on motor phenomenology was satisfactory.

Patients and Methods

Patients were consecutively enrolled from eight Italian tertiary referral centers for movement disorders. Fifteen patients had a diagnosis of PMD and 14 patients had a diagnosis of organic movement disorder by their treating specialist. Inclusion criteria were a clinical diagnosis of PMD according to FW criteria or a diagnosis of organic movement disorder according to clinical criteria.^{5–7} Patients were matched by predominant type of movement disorder and main body part involved. After enrolment, patients underwent a standardized clinical interview and videotaped clinical examination (Supporting Material 1).

Six neurologists with expertise in movement disorders (the study investigators) reviewed clinical files and videotaped clinical examinations of 29 patients. Panel members expressed a dichotomous (i.e., psychogenic versus organic) diagnostic judgment. Consensus was reached if agreement was obtained in at least five of the six panel members. In the absence of a “gold standard” diagnosis in PMD, we used this consensus approach as a benchmark of diagnostic agreement to be measured among MDS and non-MDS neurologists. Consensus was reached on 28 patients (14 with PMD and 14 with organic movement disorders), unanimously on 21 of these. A single patient upon whom consensus could not be reached was excluded from the study. Clinical details are shown in Table 1. The study was approved by the local ethics committees and all subjects gave their written informed consent, according to the Declaration of Helsinki.

Eight MDS and six non-MDS raters were recruited. One subgroup of raters (five MDS and three non-MDS) provided a dichotomous judgment and a clinical certainty judgment, based on FW and SG criteria, after having reviewed both the videorecordings and the standardized clinical files. The remaining raters provided two ratings: (1) a dichotomous judgment on the videorecordings and (2) a dichotomous judgment and clinical certainty judgment, based on FW and SG criteria, after viewing both videorecordings and standardized clinical files.

The Mann-Whitney U test was used to assess differences in clinical and demographic features between PMD and organic movement disorders. Interrater reliability of the

dichotomous judgment was assessed using the κ statistic. Reliability of clinical certainty judgments using FW and SG criteria was measured using two different approaches. Our primary approach was to consider levels of clinical certainty in the two sets of criteria as separate categories (i.e., the categorical approach). Here, we used weighted κ statistic to measure the agreement reached on each of the different levels of clinical certainty independently. In a secondary analysis, we considered the level of clinical certainty as a continuous variable (i.e., the dimensional approach). Here, we calculated intraclass correlation coefficients (ICCs) and related standard errors (SEs) and 95% confidence intervals (CIs) using a large one-way analysis of variance. A standard statistical package (*STATA 11*; StataCorp LP, College Station, TX) was used to estimate κ values. SEs and two-sided 95% CI for each value of κ were computed⁸; the differences between κ estimates from different groups were considered statistically significant if their respective 95% CI did not overlap. We interpreted κ values on the basis of the Landis-Koch classification: $\kappa < 0.0$, poor; $\kappa = 0$ to 0.2, slight; $\kappa = 0.21$ to 0.4, fair; $\kappa = 0.41$ to 0.6, moderate; $\kappa = 0.61$ to 0.8, substantial; and $\kappa = 0.81$ to 1.0, almost perfect. With respect to reliability on dichotomous judgment, the study sample size yielded a study power of 80% to detect a κ value of 0.5 and of 90% to detect a κ value of 0.6, assuming a 0.5 proportion of positive ratings while performing a two-tailed test of the null hypothesis $\kappa = 0$.

Results

Table 2 shows the level of agreement for each variable tested. The overall chance-corrected level of agreement on the dichotomous judgment, based upon videorecordings examination alone, was equally “fair” in both groups of raters. Agreement improved when raters reviewed standardized clinical files and video recordings, reaching the level of “substantial” in both. The level of agreement on the two subgroups of patients where consensus was unanimous and not unanimous did not significantly differ among MDS and non-MDS raters. The overall level of agreement on the FW-based categories of clinical certainty was “moderate” in both groups of raters. When divided into each of the four categories, agreement was substantial in both groups for “not psychogenic”, and moderate or substantial in the MDS and non-MDS groups, respectively, for the “clinically definite” judgment. The “probable” judgment yielded a fair level of agreement, whereas the “possible” judgment yielded only “poor” to “slight” agreement. Overall agreement for SG criteria was moderate for both groups of raters. Agreement was substantial for not psychogenic in both groups and moderate or fair for the clinically definite judgment in the MDS and non-MDS groups, respectively, whereas there was only slight agreement for the probable and possible categories. ICC analysis is presented in Supporting Material 2.

MDS raters reached lower agreement on patients with “jerks”; non-MDS raters reached higher agreement when rating women and patients with a median age older than 53 years. Regarding type of movement disorder, non-MDS raters reached higher agreement on patients with speech difficulties, but lower agreement on patients with parkinsonism, jerks, and gait difficulties. Demographic and clinical features influencing interrater agreement are shown in Supporting Material 3.

Discussion

We demonstrated a low level of diagnostic agreement for PMD, regardless of experience, when the diagnostic judgment was exclusively based on phenomenology of the movement disorder documented with a standardized videorecording. When asked to express, in a dichotomous fashion, whether a patient has a psychogenic or an organic movement disorder, both MDS and non-MDS experts reach substantial agreement only when provided with a detailed clinical history. Current criteria perform poorly when rating possible and probable

PMD, and only the FW criteria reached substantial agreement in cases of clinically definite PMD.

The application of a limited number of diagnostic certainty levels within a set of criteria (i.e., definite, probable, and possible) is important when there is paucity of diagnostic tests or pathognomonic features, as in PMD. This subcategorization is meant to optimize diagnostic agreement for cases falling in a “gray” zone. FW and SG criteria share some principles, including the criterion that movements should be inconsistent and incongruent with those of recognized organic movement disorders. However, psychological features are given disproportionately greater relevance than phenomenology in SG criteria, allowing a diagnosis of PMD to be made with little regard for the phenomenology.⁹ In the present study, both sets of criteria did not seem sufficiently reliable when applied to those cases where there was a degree of clinical uncertainty in the diagnosis, thus failing to serve their main purpose. This finding supports the outcome of a recent survey of 519 MDS in which the diagnosis was more commonly made after excluding a wide range of organic illnesses (i.e., inappropriate exclusionary diagnosis), rather than by applying the FW or SG diagnostic criteria (i.e., desirable inclusionary diagnosis).¹⁰

Our results further suggest that the applicability of clinical criteria is compromised by their insufficient interrater reliability. Only the clinically definite category is sufficiently reliable and clinically useful in applying an inclusionary (i.e., FW-based), rather than exclusionary, diagnosis. Recently, the weaknesses of the probable and possible categories have been highlighted by Gupta and Lang,¹¹ who proposed to reclassify probable PMD patients into the possible category and to eliminate the possible category, as described in the FW criteria, because it likely includes patients with organic movement disorders with superimposed psychiatric disorders and psychological embellishment.^{11,12}

We acknowledge some limitations of our study. The classification of patients as psychogenic or organic before the rating procedure could be based only on consensus within a group of experts. This is because there is no diagnostic gold standard for PMD. This is an inherent limitation to any study that attempts to study patients with PMD, and we accept the possibility that some patients could have been misclassified by this process. Both groups of raters were guided by a standardized grid, which, unlike most instances in routine clinical practice, was particularly comprehensive with regard to historical features relevant to PMD. We have probed only two sets of available diagnostic criteria. Brown and Thompson¹⁶ and, more recently, Gupta and Lang¹¹ have proposed a refinement of FW criteria, taking into account findings from electrophysiological studies^{17,18} to support the diagnosis of PMDs. We could not evaluate these criteria, because an electrophysiology laboratory was not available in all recruiting centers, in keeping with the survey finding that 24% of MDS neurologists do not have routine access to neurophysiological assessment.¹⁰

Conclusion

In conclusion, our data suggest that the diagnostic agreement on PMD is poor, based on review of phenomenology by video alone, and strongly relies upon clinical interpretation of historical features and diagnostic workup. The limited usefulness of currently available clinical criteria for PMD in reliably classifying uncertain cases suggests the need for new diagnostic criteria. In our opinion, these should focus mainly on improving reliability within the major area of diagnostic weakness identified by our study, that is, the recognition of objective phenomenological features specific for each type of PMD (e.g., tremor, dystonia, and myoclonus). The misdiagnosis rate and the time to diagnosis could both potentially be reduced by minimizing emphasis on psychological factors.¹⁹

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Crimlik HL, Bhatia K, Cope H, David A, Marsden CD, Ron MA. Slater revisited: 6 year follow up study of patients with medically unexplained motor symptoms. *BMJ*. 1998; 316:582–586. [PubMed: 9518908]
2. Williams DT, Ford B, Fahn S. Phenomenology and psychopathology related to psychogenic movement disorders. *Adv Neurol*. 1995; 65:231–257. [PubMed: 7872143]
3. Shill H, Gerber P. Evaluation of clinical diagnostic criteria for psychogenic movement disorders. *Mov Disord*. 2006; 21:1163–1168. [PubMed: 16685685]
4. Feinstein A, Stergiopoulos V, Fine J, Lang AE. Psychiatric outcome in patients with a psychogenic movement disorder: a prospective study. *Neuropsychiatry Neuropsychol Behav Neurol*. 2001; 14:169–176. [PubMed: 11513100]
5. Fahn S, Eldridge R. Definition of dystonia and classification of the dystonic states. *Adv Neurol*. 1976; 14:1–5. [PubMed: 941763]
6. Deuschl G, Bain P, Brin M. Consensus statement of the Movement Disorder Society on Tremor. *Ad Hoc Scientific Committee Mov Disord*. 1998; 13(Suppl 3):2–23.
7. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992; 55:181–184. [PubMed: 1564476]
8. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005; 85:257–268. [PubMed: 15733050]
9. Voon V, Lang AE, Hallett M. Diagnosing psychogenic movement disorders—which criteria should be used in clinical practice? *Nat Clin Pract Neurol*. 2007; 3:134–135. [PubMed: 17262077]
10. Espay AJ, Goldenhar LM, Voon V, Schrag A, Burton N, Lang AE. Opinions and clinical practices related to diagnosing and managing patients with psychogenic movement disorders: an international survey of movement disorder society members. *Mov Disord*. 2009; 24:1366–1374. [PubMed: 19425106]
11. Gupta A, Lang AE. Psychogenic movement disorders. *Curr Opin Neurol*. 2009; 22:430–436. [PubMed: 19542886]
12. Schrag A, Lang AE. Psychogenic movement disorders. *Curr Opin Neurol*. 2005; 18:399–404. [PubMed: 16003115]
13. Baik JS, Lang AE. Gait abnormalities in psychogenic movement disorders. *Mov Disord*. 2007; 22:395–399. [PubMed: 17216648]
14. Keane JR. Hysterical gait disorders: 60 cases. *Neurology*. 1989; 39:586–589. [PubMed: 2927686]
15. Hayes MW, Graham S, Heldorf P, de Moore G, Morris JG. A video review of the diagnosis of psychogenic gait: appendix and commentary. *Mov Disord*. 1999; 14:914–921. [PubMed: 10584664]
16. Brown P, Thompson PD. Electrophysiological aids to the diagnosis of psychogenic jerks, spasms, and tremor. *Mov Disord*. 2001; 16:595–599. [PubMed: 11481681]
17. Zeuner KE, Shoge RO, Goldstein SR, Dambrosia JM, Hallett M. Accelerometry to distinguish psychogenic from essential or parkinsonian tremor. *Neurology*. 2003; 61:548–550. [PubMed: 12939436]
18. Esposito M, Edwards MJ, Bhatia KP, Brown P, Cordivari C. Idiopathic spinal myoclonus: a clinical and neurophysiological assessment of a movement disorder of uncertain origin. *Mov Disord*. 2009; 24:2344–2349. [PubMed: 19908306]
19. Edwards MJ, Stone J. How “psychogenic” are psychogenic movement disorders? *Mov Disord*. 2011; 26:1787–1788. [PubMed: 21761457]

Table 1

Demographic and Clinical Features of Patients With Psychogenic and Organic MD

	Age/Gender	Education (yr)	Dis-D (yr)	Onset/Course	Phenomenology	Level of Agreement According to FW
Psychogenic MD						
1	27/M	15	1	Gradual/paroxysmal with remission	Dystonia/myoclonus/tremor/gait disorder	Probable
2	53/F	7	3	Abrupt/paroxysmal	Tremor/dystonia	Definite
3	39/F	13	6	Abrupt/stable with paroxysmal symptoms	Dystonia	Definite
4	50/F	8	9	Abrupt/stable with paroxysmal symptoms	Dystonia	Definite
5	32/M	8	8	Gradual/stable with paroxysmal symptoms	Dystonia	Definite
6	21/F	13	1	Abrupt/stable	Dystonia and gaze palsy	Probable
7	21/M	13	2	Abrupt/slowly progressive with remissions	Gait disorder	Probable
8	31/F	12	4	Abrupt/paroxysmal with remission	Tremor/parkinsonism	Definite
9	45/M	11	15	Gradual/slowly progressive	Myoclonus	Definite
10	58/M	16	3	Abrupt/stable	Parkinsonism	Probable
11	34/F	13	10	Abrupt/paroxysmal with remission	Tremor/parkinsonism	Probable
12	70/F	8	11	Abrupt/slowly progressive	Tremor/gait disorders	Definite
13	43/M	6	1	Abrupt/paroxysmal with remission	Tremor	Probable
14	58/F	8	3	Abrupt/slowly progressive	Tremor	Definite
Organic MD						
15	63/F	5	9	Gradual/slowly progressive	Dystonia/stereotypies/tics/gait disorder	Possible
16	69/F	13	7	Gradual/rapidly progressive	Dystonia	Possible
17	59/M	13	2	Gradual/stable	Dystonia	N/A
18	52/F	18	2	Gradual/slowly progressive	Dystonia	Possible
19	40/F	17	18	Gradual/slowly progressive	Dystonia	N/A
20	35/M	8	6	Gradual/slowly progressive	Gait disorder	N/A
21	58/M	18	1	Gradual/slowly progressive	Tremor	N/A
22	65/M	13	10	Gradual/slowly progressive	Tremor	Possible
23	53/F	13	11	Gradual/slowly progressive	Tremor	N/A
24	50/F	16	30	Gradual/slowly progressive	Tremor	N/A
25	73/M	13	3	Gradual/stable	Tremor	N/A
26	66/M	8	5	Abrupt/slowly progressive	Parkinsonism	N/A
27	64/M	13	3	Gradual/slowly progressive	Parkinsonism	Possible

Age/Gender	Education (yr)	Dis-D (yr)	Onset/Course	Phenomenology	Level of Agreement According to FW
28 70/M	18	9	Gradual/slowly progressive	Tremor/parkinsonism	N/A

Compared to organic movement disorders (58.4 ± 11.2 years), PMD patients were younger (41.6 ± 14.9 ; $P = 0.004$). No differences were found between the two groups for disease duration (5.5 ± 4.4 versus 8.3 ± 7.8 years; $P = 0.3$), number of school years (10.8 ± 3.2 versus 13.3 ± 4.0 ; $P = 0.06$), and gender ($P = 0.3$, by Fisher's exact test).

Abbreviations: F, female; M, male; Dis-D, disease duration; MD, movement disorder; N/A, not applicable.

Table 2

Level of Agreement Based on FW and SG Categories of Clinical Certainty

Dichotomous Judgment	MDS Experts κ (SE, 95% CI)	Non-MDS Experts κ (SE, 95% CI)
Video	0.40 (0.06, 0.34–0.46)	0.34 (0.08, 0.27–0.42)
Video+historical details	0.81 (0.04, 0.78–0.85)	0.74 (0.03, 0.71–0.7)
Degree of Clinical Certainty	MDS Experts κ (SE, 95% CI)	Non-MDS Experts κ (SE, 95% CI)
FW criteria		
Weighted	0.57 (0.021, 0.543–0.597)	0.57 (0.033, 0.537–0.603)
Not psychogenic	0.73 (0.036, 0.699–0.77)	0.72 (0.048, 0.672–0.768)
Possibly psychogenic	0.04 (0.036, 0.004–0.071)	0.07 (0.051, 0.019–0.121)
Probably psychogenic	0.22 (0.036, 0.186–0.258)	0.26 (0.048, 0.217–0.313)
Definitely psychogenic	0.59 (0.036, 0.556–0.628)	0.69 (0.048, 0.642–0.738)
Panel unanimous	0.67 (0.05, 0.62–0.73)	0.98 (0.01, 0.97–0.99)
Panel not unanimous	0.53 (0.04, 0.49–0.58)	0.48 (0.02, 0.46–0.50)
SG criteria		
Weighted	0.53 (0.023, 0.508–0.554)	0.43 (0.031, 0.399–0.461)
Not psychogenic	0.83 (0.032, 0.79–0.862)	0.74 (0.048, 0.692–0.788)
Possibly psychogenic	0.14 (0.032, 0.101–0.173)	0.22 (0.048, 0.172–0.268)
Probably psychogenic	0.1 (0.032, 0.066–0.138)	0.12 (0.048, 0.072–0.168)
Definitely psychogenic	0.56 (0.032, 0.522–0.594)	0.31 (0.048, 0.262–0.358)
Panel unanimous	0.74 (0.05, 0.69–0.79)	0.51 (0.04, 0.47–0.55)
Panel not unanimous	0.43 (0.043, 0.382–0.476)	0.12 (0.05, 0.07–0.17)