Review Article

Consistency of inclusion criteria for functional movement disorder clinical research studies: A systematic review

Brian Kirkwood[a] and Victor W. Mark[a,b,c,*]

[a]*Department of Physical Medicine and Rehabilitation, University of Alabama at Birmingham, Birmingham, AL, USA*
[b]*Department of Neurology, University of Alabama at Birmingham, Birmingham, AL, USA*
[c]*Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, USA*

Abstract.

BACKGROUND: Functional movement disorders (FMDs) are a common cause of disability. With an increasing research interest in FMD, including the emergence of intervention trials, it is crucial that research methodology be examined, and standardized protocols be developed.

OBJECTIVE: To characterize the current inclusion criteria used to select patients for FMD research studies and review the consistency and appropriateness of these criteria.

METHODS: We identified studies of potential biomarkers for FMD that were published over the last two decades and performed a qualitative analysis on the finally included studies.

RESULTS: We identified 79 articles and found inconsistent inclusion criteria. The Fahn-Williams and DSM-IV criteria were the most commonly applied, but neither accounted for the majority (Fahn-Williams 46%, DSM-IV 32% of the total). The selection of the inclusion criteria depended in part on the phenotype of FMD under investigation. We also identified inclusion methodologies that were not appropriate, such as the inclusion of low-certainty diagnoses and diagnosing by excluding specific biomarkers rather than including patients based on clinical characteristics that commonly are thought to suggest FMD.

CONCLUSIONS: Significant variability exists with the inclusion criteria for FMD research studies. This variability could limit reproducibility and the appropriate aggregation of data for meta-analysis. Advancing FMD rehabilitation research will need standardized inclusion criteria. We make some suggestions.

Keywords: Functional neurological disorder, rehabilitation, MRI

1. Background

Functional movement disorders (FMDs) are common and disabling conditions (Gelauff & Stone,

2016; Stone, Carson, et al., 2010). These have historically been identified by diverse alternate terms, including "hysteria," "conversion disorder," "psychogenic disorder," and "dissociative disorder." The term "functional" in recent years has become the accepted term among clinicians who specialize in care for persons with FMD, because it is judged not to be pejorative, but rather it conveys causative

---

*Address for correspondence: Victor Mark, 619 19th Street South, SRC 190, Birmingham 35249, AL, USA. Tel.: +1 205 934 3499; Fax: +1 205 975 4524. E-mail: vwmark@uabmc.edu.

neutrality and may increase patient understanding and acceptance (Ding & Kanaan, 2017). Adopting the term "FMD" over prior terms has been viewed as a step toward advancing the scientific evaluation and care for this disorder.

FMDs are thought to involve the impaired voluntary control of movement over parts of the body in the presence of normal intent to move and intact neuromuscular capacity for movement (Espay et al., 2018). Interest in FMD has increased over the last several decades, with a surge of intervention trials in recent years (Pick et al., 2020). The consistency for diagnosing FMD has become all the more important because the National Institutes of Health of the United States recently announced requests for funding applications to validate functional neurological disorder diagnosis through biomarker assessment, thus to generally improve assessment for this illness (National Institute of Neurological Diseases and Stroke, 2021a, 2021b). However, the consistency for diagnosing FMD across studies has not so far been determined. Consistent diagnostic methods would allow accumulating data for meta-analyses of rehabilitation and support replication in the clinical setting. We therefore systematically reviewed a subset of such studies for their consistency of methods of FMD diagnosis.

2. Methods

For this systematic review, we elected to evaluate only studies that investigated laboratory techniques to identify potential FMD biomarkers. We focused on biomarker studies because by their using largely objective measures (e.g., structural or physiological brain imaging, electromyography, electroencephalography), this could eventually support identifying reliable study inclusion criteria that could be used for later clinical trials. We assumed that the diagnostic methods for FMD in biomarker studies would represent the broader body of FMD research.

We modeled our literature search from that of Thomsen et.al. (Thomsen et al., 2020), who evaluated the sensitivity of specific biomarker assessments for diagnosing FMD. We restricted the dates of publication to between 1 January 2001 and 1 January 2021 to represent current diagnostic methods. The inclusion criteria were further modified to select original data studies that investigated biological correlates of abnormal movement, included an FMD experimental group, and publication in English. Studies were excluded based on sample size less than four.

A PubMed search conducted on 28 February 2021 identified 921 records that met the search criteria (Fig. 1) Twelve additional articles were included while reviewing the citations of included articles. Seven hundred ninety-three were excluded from inspecting the titles and abstracts, resulting in 140 full text articles. Seventy-nine articles were finally found to satisfy all criteria. The web-only supplementary table summarizes all data extracted from the studies.

We then subdivided the studies according to the following major categories of inclusion criteria, based on our preliminary observation of the most frequently used inclusion criteria.

(1) The *Fahn-Williams criteria* and variants (FW). The original FW criteria were published in 1988 with respect to functional dystonia (Fahn & Williams, 1988). The criteria were updated in 1995 to include tremor, parkinsonism, gait disturbances, and myoclonus (Williams et al., 1995). Notably, this update did not include functional paresis. Fahn and Williams established demonstrating *inconsistency* and *incongruence* as the diagnostic pillars. *Incongruence* has been described as "Movements [that] do not present or progress according to the wide phenotypic range of known organic movement disorders (Espay & Lang, 2015)," while *inconsistency* can refer to symptoms that vary over time.

It is important to note that these criteria do not require psychological disturbance, but rely substantially on historical, emotional, and non-movement related findings. Fahn and Williams created categories of grades of diagnostic certainty. The highest level of certainty, *clinically-definite,* was created in the 1995 update by combining the *documented* and *clinically-established* categories originally described in 1988 (Williams et al., 1995). Patients were considered to have *documented* FMD if the signs were observed to remit when the patients were unobserved or with efficacious therapy. In the *clinically-established* category, patients had movements that were *inconsistent* or *incongruent* but also had to have other "false" signs (equated to "psychogenic" signs, though not specified), multiple somatizations, or obvious psychiatric disturbance. Lesser levels of certainty included *probable*, where patients could be diagnosed based solely on "false signs" (without *inconsistent* or *incongruent* movements), and *possible*, which simply required emotional disturbance. The categories of *probable* and *possible* were subsequently shown to have poor interrater reliability and are of no clinical use (Morgante et al., 2012).
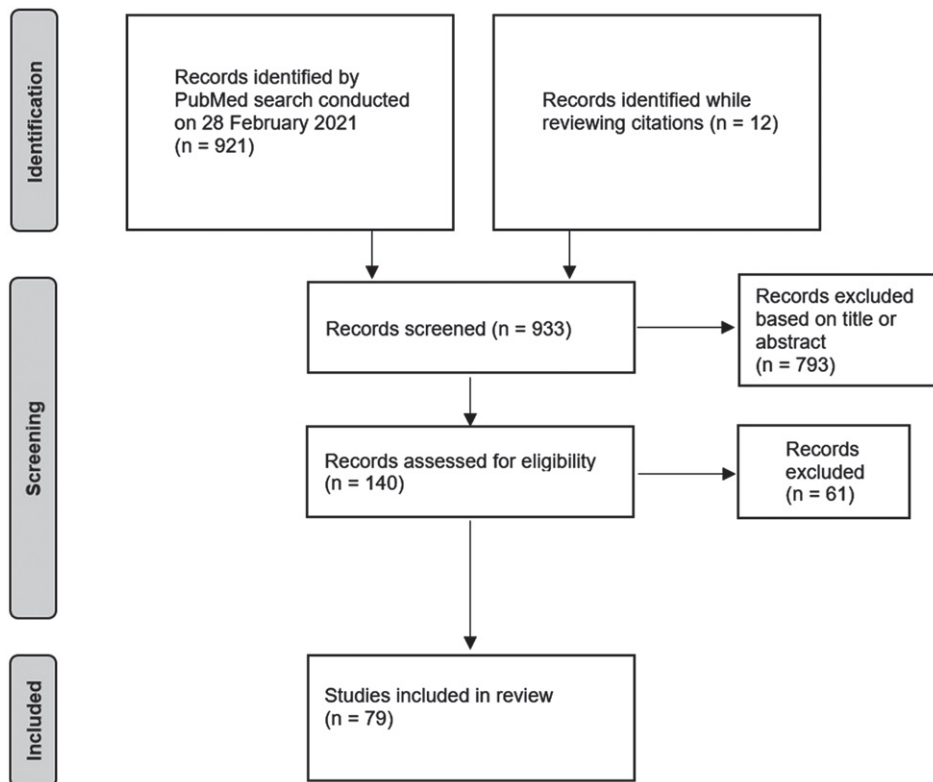
Fig. 1. PRISMA diagram for literature review.

Subsequently, the Shill-Gerber (SG) criteria were intended to expand on the FW by creating an avenue for diagnosis that did not require *inconsistent* or *incongruent* movements, although incongruence could still be considered among subsidiary criteria (Shill & Gerber, 2006). "Clinically proven" FMD was diagnosed provide that it either (1) remits with psychotherapy, (2) remits when the patient feels to be unobserved, or (3) has a demonstrated premovement *Bereitschaftspotential* on electroencephalography (EEG) in cases of myoclonus. [The *Bereitschaftspotential*, also known as the "readiness potential" or "motor-related cortical potential," is the slowly rising negative electrocortical activity on EEG that precedes experimentally-verified intended movement (Di Russo et al., 2017).] In contrast, suspected FMD that was not considered to fall under these criteria was classified as "clinically definite" FMD by creating "primary" vs. "secondary" criteria. Under primary criteria, the disease had to either be incongruent with so-called organic disease (including inconsistent symptoms presentations), involve excessive pain or fatigue, or have previous exposure to a disease model or potential for secondary gain.

In contrast, the secondary criteria involved "multiple somatizations (other than pain and fatigue) and/or obvious psychiatric disturbance." To qualify as "clinically definite" FMD, it had to meet "at least three primary criteria and at least one secondary criterion." The category of "clinically probable" FMD entails two primary criteria and two secondary criteria. Finally, "clinically possible" FMD involves one primary criterion and two secondary, or two primary and one secondary criterion. A major pitfall of this system is that any patient with pain, exposure to a disease model, potential for secondary gain, and multiple somatizations could incorrectly be identified as "clinically definite" FMD.

The still later Gupta-Lang (GL) criteria were intended to address potential problems with FW and to include advances in laboratory-supported evidence of FMD in diagnosis (Gupta & Lang, 2009). To address potential problems, Gupta and Lang removed the *probable* and *possible* classifications from their criteria due to their lack of reliability, and rejected the inclusion of patients with *consistent/congruent* movements. Additionally, they created a *laboratory-supported definite* category

to facilitate diagnosing FMD in cases of tremor and myoclonus where electrophysiological data were available. Specifically, these were entrainment of tremor as reflected by surface electromyography and back-averaged *Bereitschaftspotentials* during electroencephalography-electromyography of myoclonus.

Subsequently, Espay and Lang expanded on GL by creating a phenotype-specific diagnostic process for FMD (Espay & Lang, 2015). This rejected the requirement of historic and emotional data and updated laboratory evidence (flurorodopa positron emission tomography, or PET; single photon emission computed tomography, or SPECT). To create a phenotype-specific diagnosis process, they identified core features that were individually specific to functional tremor, dystonia, myoclonus, tics, parkinsonism, and gait impairment. This set of criteria requires that all of these core features be present to make a *clinically-definite* diagnosis. Except for functional dystonia, where rapid onset is considered to be a diagnostic feature, all core features are movement-related phenomena rather than historical data. The core features are specific and generally lack the subjective language of FW, SG, and GL such as "bizarre" and "unusual." The authors argued against the use of historical and emotional data, citing their prevalence as well in structurally-defined neurological disease (e.g., stroke, brain tumor) and the lack of sensitivity and specificity required to provide clinical usefulness.

(2) The *Diagnostic and Statistical Manual* (DSM)-IV laid out criteria for diagnosing "conversion disorder" (CD), a term based on the Freudian hypothesis that psychological stress is converted into somatic symptoms (American Psychiatric Association, 1994). Because the earliest DSM editions were published before FW had become standard, the DSM-IV was relatively uninfluenced by the norms established in FW. The diagnosis was considered to be psychological and primarily required the diagnostician to demonstrate that psychological factors were associated with the symptom. This created logistical barriers that impeded accurate and efficient diagnosis.

Many of these limitations were addressed in the DSM-V, which encouraged the clinician to look for signs that are diagnostic for FMD and removed the criteria for psychological association and excluding feigning (American Psychiatric Association, 2000).

(3) The category *Other* was applied to the studies that did not rely on either the FW variants or the DSM.
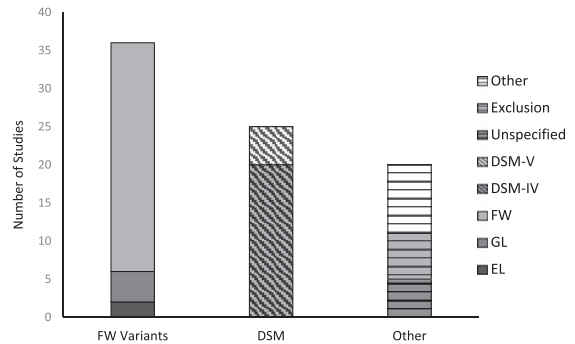

Fig. 2. Inclusion criteria utilization in FMD biomarker studies. FW: Fahn & Williams (1995); GL: Gupta & Lang (2009); EL: Espay & Lang (2015); DSM: Diagnostic and Statistical Manual.

3. Results

3.1. Criteria utilization

As seen in Fig. 2, the FW criteria were the most prevalent diagnostic criteria used in FMD research trials. FW was used in 30 of the 79 investigations. Six additional studies employed a variant of FW—four studies using GL and two using EL. SG were not found to be used as inclusion criteria in any investigation identified by this review. The DSM-IV was the second most common method, involving 25 studies. Two studies used both DSM and FW criteria. Finally, 'Other' encompassed 20 studies. Five studies did not specify how subjects were diagnosed. This group included studies that simply used the referring physician's diagnosis (van der Stouwe et al., 2016), without specifying the physicians' criteria. Ten studies created their own diagnostic criteria or used some other set of criteria—for example, the ICD-10 criteria for "dissociative disorder" (Liepert et al., 2011). Five studies diagnosed FMD based primarily on excluding structural disease. These studies were mostly investigations of suspected stroke patients who were considered to have a functional etiology after repeated structural imaging ruled out cerebrovascular accident (Benussi et al., 2020; Premi et al., 2017).

3.2. Criteria by studied phenotype

As seen in Fig. 3, the selection of inclusion criteria highly depends on the phenotype under investigation. DSM variants are significantly represented in studies that investigated functional paresis. Functional paresis was not included by Fahn and Williams in their
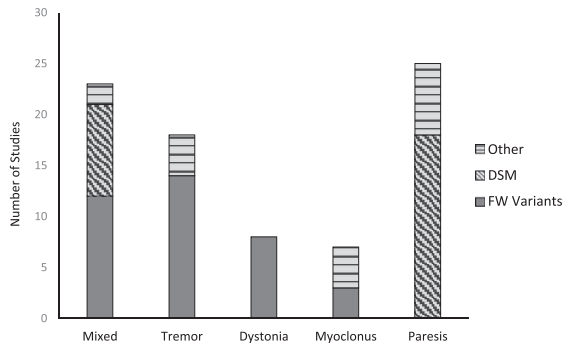
Fig. 3. Inclusion criteria utilization according to studied phenotype. FW variants: Fahn & Williams (1995), Gupta & Lang (2009), or Espay & Lang (2015); DSM: Diagnostic and Statistical Manual 4th and 5th edition; Other: studies not using a FW Variant or the DSM.

Fig. 4. Inclusion of 'probable' in studies using the Fahn-Williams Criteria (Williams et al., 1995).

original criteria and has not been added to the updates or variants thereafter (Espay & Lang, 2015). There were no instances of studies investigating tremor, dystonia, or myoclonus that employed DSM. Those investigations most often used the FW variants. Of the 23 mixed studies—that is, those investigating more than one phenotype of FMD—FW variants were used in 12 instances, DSM variants in nine, and two publications that used neither. There was considerably more variance compared to the studies of single phenotypes.

3.3. Inclusion of probable and possible

The categories of *probable* and *possible* are only found within FW and SG. Eight of the 30 studies using FW allowed the inclusion of *probable*. Given the increasing awareness of the problems with the *probable* designation over the period investigated, we looked for changes in the allowance of *probable* over time. As shown in Fig. 4, the inclusion of subjects diagnosed with *probable* FMD remains pervasive. Between 2016 and 2020, nine studies were identified that employed FW as their primary inclusion method, with three of the nine allowing the inclusion of *probable*. Consequently, although the majority of studies of biomarkers of FMD that used the FW inclusion criteria did not apply the *probable* category, nonetheless there remains a continuing minority of studies that use the *probable* category, despite the published indication that the *probable* and *possible* categories are unreliable (Morgante et al., 2012). None of the studies had included subjects who were designated to have *possible* FMD.
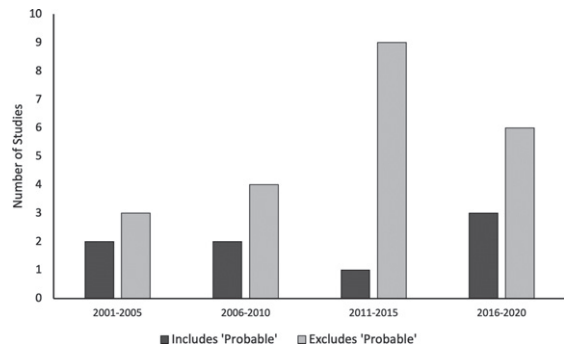
4. Discussion

4.1. Criteria application in research

This systematic review revealed considerable discrepancies in the choice of diagnostic criteria used to include subjects in FMD research. There is relative consistency *within* studies that investigate the same phenotype, but the choice of criteria varies depending on what specific phenotype is being studied. For example, it was seen that all of the investigations of functional dystonia employed FW, while no study of paresis employed FW. This variability is likely related to criteria sets (FW, SG, GL, and EL) that did not include methods for diagnosing functional paresis. Because functional paresis investigations cannot use FW variants criteria, these researchers rely on the DSM-IV disproportionately to other groups. This review also found significant variability in the criteria used in mixed studies (those that evaluated more than one phenotype). These studies may or may not include functional paresis (Nahab et al., 2017; Wegrzyk et al., 2018). There is a clear need for a criterion set that may be used to include all patients with symptoms of FMD.

Twenty of the identified studies did not use a variant of FW or DSM. These are not strictly of low quality but do reveal some of the poorer methods that are common in FMD research. One of these is including FMD subjects based purely on exclusion of "organic" disease, without specifying the bases of exclusion. It was also common for researchers to provide unreproducible methods, such as providing no information as to the inclusion criteria or reporting that the diagnosis is based on the expertise of the investigators (Huys et al., 2020).

4.2. Current criteria

4.2.1. Fahn-Williams criteria and variants

The FW criteria have been accepted by many researchers and clinicians for diagnosing FMDs or including patients in treatment trials. In our literature review, the FW inclusion criteria were the leading measures for FMD studies. However, several limitations pertain to the elements of the criteria, which raise questions on the comparability of study groups and treatment outcomes from different study sites when they have applied these criteria.

Williams et al. (1995) identified the central challenge of diagnosing FMD to be "establishing that the abnormal movements derive in part or fully from an underlying psychological disturbance." Since that time, the field has come to question the necessity of underlying psychological pathology, has moved away from terms that imply a psychological origin like "psychogenic" or "conversion disorder," and proposed alternative pathologic mechanisms (Edwards & Bhatia, 2012; Stone, LaFrance, et al., 2010).

One of the pillars of the FW criteria for functional dystonia is its "incongruity" with legacy (or "organic") dystonia. However, such criterion begs precise definition of legacy dystonia in the first place, given that dystonia is a heterogeneous disorder (Conte et al., 2020) and that the range of self-administered "sensory tricks" or "*gestes antagonistes*" (maneuvers that can briefly interrupt legacy dystonia) can include widely varying remedies such as bending forward, singing, humming, piano playing, or running in a counterclockwise direction (Ramos et al., 2014). Therefore, establishing the boundary between functional dystonia and legacy dystonia is unclear, to say the least.

Another problem inherent to the criteria for diagnosing functional dystonia can include suspected either factitious disorder (as reflected by self-injury) or malingering (reflected by "false weakness"). Making such judgments by the clinician requires extended evaluation and review of medical records and reports from observers for these behaviors (Bass & Wade, 2019). It is most likely impractical to apply such criteria.

Yet another weakness within the FW design is the observation of beneficial outcomes from adjuvant behavioral therapies, including tricyclic antidepressants and extensive physical therapy. Although tricyclic antidepressants may be initiated to treat depression, however, this class also can treat neuropathic pain. Considering that "pain out of proportion

to exam" is a supportive feature of the inclusion criteria, initiating an antidepressant may involve a confounding variable. This could similarly by applied to the initiation of intensive physical therapy. In particular, the outcome of physical therapy may depend in part on the specific content of the treatment or qualifications of the therapist. Hence, absence of benefit from physical therapy should not imply absence of FMD.

As alluded to above, most patients who were included in the creation of FW criteria were diagnosed based on documenting symptom *remission* rather than signs that are present during a specific clinical evaluation. This is distinctly different from the current application of FW, where the observation of concurrent signs is regarded as the hallmark of the Fahn-Williams criteria (Bhatia et al., 2018; Espay & Lang, 2015). Remission with psychotherapy was the goal of the combined diagnostic and therapeutic process laid out by the publications. In fact, only four of the 21 patients included by Fahn and Williams (1988) were considered to be *clinically-established*. The stated reasoning for arriving at a *clinically-established* diagnosis rather than pursuing an *established* diagnosis was the patient being unwilling to complete the diagnostic process. This is not to say that Fahn and Williams did not regard clinical signs as important. They emphasized the importance of clinical signs as clues for a functional origin but did not consider the recognition of clinical signs as sufficient for a *definite* diagnosis. Clinical signs, especially those that have been validated, may be diagnostic, but that was not the intention of FW.

The GL and EL criteria addressed many of the limitations with FW but remain flawed. GL removed the criterion regarding intentionality, freeing the clinician from this undue burden. However, GL prematurely places much emphasis on laboratory-supported criteria where clear laboratory markers have not been validated. Additionally, GL does not provide diagnostic guidelines that are phenotype-specific. EL went further to identify criteria that were phenotype-specific. A major problem with EL is that it has strict inclusion criteria based on sensitivities established using FW as a gold standard. Since no gold standard has been established for FMD, strict inclusion requirements are premature. However, EL include phenomenological descriptions of the abnormal movements that provide clinicians the most guidance of any criteria set that was identified in this review.

4.3. DSM-IV

The DSM-IV was the second most common method of including patients in FMD treatment trials that we identified. Many authors have offered salient critiques of the DSM-IV's diagnostic criteria for CD, especially in the years leading up to the publication of the DSM-V (Spiegel et al., 2011; Stone et al., 2011). However, researchers continued to turn to the DSM-IV, even after the DSM-V had become available (Blakemore et al., 2015; Hassa et al., 2016). This is especially true of studies that included patients with functional paresis, perhaps because the FW criteria did not include functional paresis among the possible diagnoses. Additionally, large differences exist between the FW criteria and the DSM-IV, which calls into question whether these inclusion methods would enroll comparable groups.

Whereas FW considered factitious and malingering disorders to be FMD and finding evidence of these conditions was considered diagnostic (Williams et al., 1995), the DSM-IV definition of CD requires that "The symptom or deficit is not intentionally produced or feigned" (Association, 2000). This not only demonstrates that the two criteria sets would select different patients, but that the application of the DSM-IV criteria is burdensome and of uncertain reliability. Demonstrating that a patient is not intentionally producing symptoms can be an impossible task, because a person's intent may not be apparent even with intensive investigation (Stone et al., 2011).

Another problem with the DSM-IV criteria is the requirement for the association of psychological factors. This is both clinically and theoretically problematic. This criterion is derived from research that demonstrates an association between stressful life events and the development of FMD (Binzer et al., 1997). However, it has never been shown that these events cause the functional presentation; in fact, many patients with known FMDs do not have identifiable psychological stressors associated with the symptom (Stone et al., 2009). Clinically, this criterion is problematic because many patients may be reluctant to discuss such matters and they may not be apparent from the available history.

The final major problem with diagnosis by the DSM-IV is that these criteria do not regard the phenomenology of movement in the diagnosis. Abnormal movements or the variable lack of movement are among the presenting symptoms in patients with FMDs (Espay et al., 2018). These criteria do not instruct the clinician in the recognition of diagnostic signs of FMD within these abnormal movements. Without this criterion, the DSM-IV does little to diagnose FMD besides attempt to exclude various other disease states. Researchers who employ FW criteria primarily make the diagnosis based on clinical signs that characterize FMD, whereas researchers using the DSM-IV may not even consider them.

4.4. General comments on discrepancies within FMD inclusion criteria

The published criteria for diagnosing FMD contains significant discrepancies that could result in variability within the FMD populations being studied by various investigators. These discrepancies include reliance on historical and emotional data, level of exclusionary diagnosis, and allowance for low certainty diagnoses.

4.4.1. Supportive features

The degree to which sets of criteria rely on historical and emotional features varies significantly. These supportive features include multiple somatizations, psychiatric disturbance, childhood trauma, exposure to disease model, and potential for secondary gain. Supportive features make poor criteria because they have been shown to be neither sufficient nor necessary for the diagnosis (Espay et al., 2018), based on sensitivity and specificity. The low sensitivity of these factors is related to their prevalence within legacy neurological diseases, i.e., those not commonly regarded as functional disorders (LaFaver et al., 2020). The low specificity is related to the poor generalizability of signs found to be specific to a signal phenotype. For example, the sudden onset has been stated to have good specificity for functional dystonia (Frucht et al., 2020), but such information may be vulnerable to recall bias in the patient (Espay & Lang, 2015). The DSM-IV diagnoses CD almost solely based on these supportive features, without regard for movement phenomenology. The DSM-IV considers CD a purely psychological diagnosis and views the role of the neurological investigation as a component of excluding medical diagnosis. This is partially true of SG, where movement phenomenology can be disregarded with sufficient "secondary features." Both the DSM-IV and SG allow the diagnosis to be made on criteria that are not sufficient for the diagnosis. The FW rejects this idea and requires clinical signs of FMD. However, the criteria require as well historical features. This is problem-

atic because these tests lack the sensitivity to be required and are therefore unnecessary. The EL criteria neither require nor rely on these supportive features.

4.4.2. Movement-related features

The criteria vary both on their requirement for movement phenomenology and in the interpretability of the criteria regarding movement-related features. As stated above, some criteria either disregard movement-related features, such as the DSM-IV, or do not require these features, such as the SG. Movement-related features should be core for diagnosing FMD. However, even for the criteria that require movement phenomena, the degree of specificity used to describe movement phenomenology varies widely. Fahn and Williams, and the subsequent variants, described functional movements as being incongruent and inconsistent, which we review next.

4.4.2.1. Incongruence.
One of the pillars of diagnosing FMD, established by Fahn and Williams, is that the movement phenomena should be incongruent with legacy neurological diseases. This is problematic because it requires a diagnostician to recognize the nuanced presentations of legacy diseases that are needed to differentiate them from functional presentations (Espay et al., 2018; Gasca-Salas & Lang, 2016; Gupta & Lang, 2009). The lack of clarity has led to criteria including non-scientific, imprecise descriptions such as "bizarre," "unusual," and "abnormal" (Espay & Lang, 2015; Fahn & Williams, 1988; Williams et al., 1995). This is further complicated by "incongruent" signs that are not generalizable to all phenotypes. For example, suppressibility of a functional tic is not incongruent with legacy tic disorders. Rather than designating signs as being incongruent with legacy disease, clinical signs that are to be used to rule-in FMD should be phenomenologically specific and not require simply to exclude legacy disease.

4.4.2.2. Inconsistency.
Inconsistency applies to the fluctuating patterns of movement within the individual patient, which may be better described as variability. This variability may be demonstrated in movements that change over time, that are suppressed with complex tasks, or where the disability is disproportionate to exam findings (Espay & Lang, 2015). These features suggest that the symptom can vary with the patient's self-attention. This thus

suggests possible specific pathophysiologic mechanisms and potential for remission with behavioral therapy. Identifying patients whose symptoms are likely to remit with behavioral therapy is one of the most important features of FMD diagnostic criteria. However, the frequency of symptoms can vary considerably also in legacy movement disorders (Lieberman, 2006; Pareés et al., 2012; Stone et al., 2005). Consequently, it will be important to revise the diagnostic criteria of FMD by acknowledging that symptoms of legacy neurological disorders can also fluctuate.

4.5. Diagnostic certainty

When applying FW or one of its variants, we suggest that only a *clinically definite* diagnosis should be considered for inclusion in FMD research, and only with respect to inconsistency of movement within the individual, and not with respect to incongruence with other neurological disorders. FW and SG both categorized diagnostic certainty with *probable* and *possible* FMD diagnoses included. A significant problem with these categories is that the diagnosis can be made in the absence of movement pathology. Additionally, these designations have been shown to have poor interrater reliability (Morgante et al., 2012). SG did not allow for these low-certainty diagnoses under their criteria, with the EL following suit. This review identified eight studies that allowed for the inclusion of patients diagnosed as *probable* FMD. This represents 22% of the 37 studies that employed the FW criteria. These studies should be considered poor evidence and caution should be taken when including their results in metanalyses.

4.6. Limitations

This study intentionally restricted its review to studies that related the diagnosis of FMD to objective biomarker evaluations (thus, MRI, EEG, and so forth). We restricted our literature review so as to model our method from that of Thomsen et al (Thomsen et al., 2020), who evaluated the psychometric validation of various biomarker evaluation techniques for diagnosing FMD. We had a simpler objective: to evaluate the consistency of diagnosing FMD among the studies that had used objective biomarkers for making this determination. We intentionally selected studies that had related biomarkers to the diagnosis accuracy of FMD, because such approaches would be regarded as scientifically rig-

orous. In contrast, we did not evaluate the broader literature for diagnosing FMD regardless of the inclusion of biomarkers, because we felt that our narrowing our literature review would be sufficiently informative with respect to the present status with respect to the consistency of diagnosing FMD. As a result, we found marked inconsistency for diagnosing FMD. We doubt that our evaluating the broader literature on FMD would have changed our conclusions.

5. Conclusion

This systematic review identified variability in the choice of diagnostic criteria used to include subjects in FMD research studies. The criteria choice was found to vary with the phenotype being investigated. This has implications in mixed studies in which more than one phenotype is investigated but only a single set of diagnostic criteria may be used. Mixed studies were found to have more variable inclusion criteria than studies of a single phenotype. Variability in criteria selection is problematic because significant differences exist between the sets of published diagnostic criteria, as reviewed above. When considering the creation of new criteria, we propose that the criteria should be inclusive of all FMDs, based on demonstrating symptom variability in response to self-attention to the symptoms during clinical examination. A unified set of inclusion criteria for FMD would consequently support future studies for their response to psychological therapy. Furthermore, such criteria should avoid subjective, unscientific language such are "bizarre" and "unusual," and exclusionary relics such as designating a symptom's being *incongruent* with other neurological diseases, given at present unsettled diagnostic criteria for many such diseases.

Conflict of interest

The authors declare that they have no conflicts of interest.

Supplementary material

The supplementary material is available from https://dx.doi.org/10.3233/NRE-228002.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders (4th ed.)* Washington, USA, American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)* Washington, USA, American Psychiatric Association.

Bass, C., & Wade, D. T. (2019). Malingering and factitious disorder. *Practical Neurology*, *19*, 96-105.

Benussi, A., Premi, E., Cantoni, V., Compostella, S., Magni, E., Gilberti, N., Vergani, V., Delrio, I., Gamba, M., Spezi, R., Costa, A., Tinazzi, M., Padovani, A., Borroni, B., & Magoni, M. (2020). Cortical inhibitory imbalance in functional paralysis. *Frontiers in Human Neuroscience*, *14*, 153.

Bhatia, K. P., Bain, P., Bajaj, N., Elble, R. J., Hallett, M., Louis, E. D., Raethjen, J., Stamelou, M., Testa, C. M., & Deuschl, G. (2018). Consensus statement on the classification of tremors. From the task force on tremor of the International Parkinson and Movement Disorder Society. *Movement Disorders*, *33*, 75-87.

Binzer, M., Andrsen, P. M., & Kullgren, G. (1997). Clinical characteristics of patients with motor disability due to conversion disorder: a prospective control group study. *Journal of Neurology, Neurosurgery, and Psychiatry*, *63*, 83-88.

Blakemore, R. L., Hyland, B. I., Hammond-Tooke, G. D., & Anson, J. G. (2015). Deficit in late-stage contingent negative variation provides evidence for disrupted movement preparation in patients with conversion paresis. *Biological Psychology*, *109*, 73-85.

Conte, A., Defazio, G., Mascia, M., Belvisi, D., Pantano, P., & Berardelli, A. (2020). Advances in the pathophysiology of adult-onset focal dystonias: recent neurophysiological and neuroimaging evidence. *F1000Research*, *9*, F1000.

Di Russo, F., Berchicci, M., Bozzacchi, C., Perri, R. L., Pitzalis, S., & Spinelli, D. (2017). Beyond the "Bereitschaftspotential'': action preparation behind cognitive functions. *Neuroscience & Biobehavioral Reviews*, *78*, 57-81.

Ding, J. M., & Kanaan, R. A. A. (2017). Conversion disorder: a systematic review of current terminology. *General Hospital Psychiatry*, *45*, 51-55.

Edwards, M. J., & Bhatia, K. P. (2012). Functional (psychogenic) movement disorders: merging mind and brain. *Lancet Neurology*, *11*, 250-260.

Espay, A. J., Aybek, S., Carson, A., Edwards, M. J., Goldstein, L. H., Hallett, M., LaFaver, K., LaFrance, W. C., Lang, A. E., Nicholson, T., Nielsen, G., Reuber, M., Voon, V., Stone, J., & Morgante, F. (2018). Current concepts in diagnosis and treatment of functional neurological disorders [review]. *JAMA Neurology*, *75*, 1132-1141.

Espay, A. J., & Lang, A. E. (2015). Phenotype-specific diagnosis of functional (psychogenic) movement disorders. *Current Neurology and Neuroscience Reports*, *15*, 32.

Fahn, S., & Williams, D. T. (1988). Psychogenic dystonia. *Advances in Neurology*, *50*, 431-455.

Frucht, L., Perez, D. L., Callahan, J., MacLean, J., Song, P. C., Sharma, N., & Stephen, C. D. (2020). Functional dystonia: differentiation from primary dystonia and multidisciplinary treatments. *Frontiers in Neurology*, *11*, 605262.

Gasca-Salas, C., & Lang, A. E. (2016). Neurologic diagnostic criteria for functional neurologic disorders. *Handbook of Clinical Neurology*, *139*, 193-212.

Gelauff, J., & Stone, J. (2016). Prognosis of functional neurologic disorders. *Handbook of Clinical Neurology*, *139*, 523-541.

Gupta, A., & Lang, A. E. (2009). Psychogenic movement disorders. *Current Opinion in Neurology*, *22*, 430-436.

Hassa, T., de Jel, E., Tuescher, O., Schmidt, R., & Schoenfeld, M. A. (2016). Functional networks of motor inhibition in conversion disorder patients and feigning subjects. *Neuroimage: Clinical*, *11*, 719-727.

Huys, A. M. L., Bhatia, K. P., Edwards, M. J., & Haggard, P. (2020). The flip side of distractibility-executive dysfunction in functional movement disorders. *Frontiers in Neurology*, *11*, 969.

LaFaver, K., Lang, A. E., Stone, J., Morgante, F., Edwards, M., Lidstone, S., Maurer, C. W., Hallett, M., Dwivedi, A. K., & Espay, A. J. (2020). Opinions and clinical practices related to diagnosing and managing functional (psychogenic) movement disorders: changes in the last decade. *European Journal of Neurology*, *27*, 975-984.

Lieberman, A. (2006). Are freezing of gait (FOG) and panic related? *Journal of the Neurological Sciences*, *248*, 219-222.

Liepert, J., Hassa, T., Tüscher, O., & Schmidt, R. (2011). Motor excitability during movement imagination and movement observation in psychogenic lower limb paresis. *Journal of Psychosomatic Research*, *70*, 59-65.

Morgante, F., Edwards, M. J., Espay, A. J., Fasano, A., Mir, P., & Martino, D. (2012). Diagnostic agreement in patients with psychogenic movement disorders. *Movement Disorders*, *27*, 548-552.

Nahab, F. B., Kundu, P., Maurer, C., Shen, Q., & Hallett, M. (2017). Impaired sense of agency in functional movement disorders: an fMRI study. *PLoS One*, *12*, e0172502.

National Institute of Neurological Diseases and Stroke. (2021a). *Notice of Special Interest: Biomarker Discover and Validation in Functional Neurological Disorders, NOT-NS-22-010*. Bethesda, USA, National Institutes of Health. https://grants.nih.gov/grants/guide/notice-files/NOT-NS-22-010.html

National Institute of Neurological Diseases and Stroke. (2021b). *Clinical Trial Readiness for Functional Neurological Disorders, PAR-22-053*. Bethesda, USA, National Institutes of Health. https://grants.nih.gov/grants/guide/pa-files/PAR-22-053.html

Pareés, I., Saifee, T. A., Kassavetis, P., Kojovic, M., Rubio-Agusti, I., Rothwell, J. C., Bhatia, K. P., & Edwards, M. J. (2012). Believing is perceiving: mismatch between self-report and actigraphy in psychogenic tremor. *Brain*, *135*, 117-123.

Pick, S., Anderson, D. G., Asadi-Pooya, A. A., Aybek, S., Baslet, G., Bloem, B. R., Bradley-Westguard, A., Brown, R. J., Carson, A. J., Chalder, T., Damianova, M., David, A. S., Edwards, M. J., Epstein, S. A., Espay, A. J., Garcin, B., Goldstein, L. H., Hallett, M., Jankovic, J.,... Nicholson, T. R. (2020). Outcome measurement in functional neurological disorder: a systematic review and recommendations. *Journal of Neurology, Neurosurgery and Psychiatry*, *91*, 638-649.

Premi, E., Benussi, A., Compostella, S., Gilberti, N., Vergani, V., Delrio, I., Spezi, R., Gamba, M., Costa, A., Gasparotti, R., Magoni, M., Padovani, A., & Borroni, B. (2017). Multimodal brain analysis of functional neurological disorders: a functional stroke mimic case series [letter]. *Psychotherapy and Psychosomatics*, *86*, 317-319.

Ramos, V. F., Karp, B. I., & Hallett, M. (2014). Tricks in dystonia: ordering the complexity. *Journal of Neurology, Neurosurgery and Psychiatry*, *85*, 987-993.

Shill, H., & Gerber, P. (2006). Evaluation of clinical diagnostic criteria for psychogenic movement disorders. *Movement Disorders*, *21*, 1163-1168.

Spiegel, D., Loewenstein, R. J., Lewis-Fernández, R., Sar, V., Simeon, D., Vermetten, E., Cardeña, E., & Dell, P. F. (2011). Dissociative disorders in DSM-5. *Depression and Anxiety*, *28*, 824-852.

Stone, J., Carson, A., Aditya, H., Prescott, R., Zaubi, M., Warlow, C., & Sharpe, M. (2009). The role of physical injury in motor and sensory conversion symptoms: a systematic and narrative review. *Journal of Psychosomatic Research*, *66*, 383-390.

Stone, J., Carson, A., Duncan, R., Roberts, R., Warlow, C., Hibberd, C., Coleman, R., Cull, R., Murray, G., Pelosi, A., Cavanagh, J., Matthews, K., Goldbeck, R., Smyth, R., Walker, J., & Sharpe, M. (2010). Who is referred to neurology clinics?—the diagnoses made in 3781 new patients. *Clinical Neurology and Neurosurgery*, *112*, 747-751.

Stone, J., Carson, A., & Sharpe, M. (2005). Functional symptoms and signs in neurology: assessment and diagnosis. *Journal of Neurology, Neurosurgery and Psychiatry*, *76 (Supplement 1)*, i2-12.

Stone, J., LaFrance, W. C., Brown, R., Spiegel, D., Levenson, J. L., & Sharpe, M. (2011). Conversion disorder: current problems and potential solutions for DSM-5. *Journal of Psychosomatic Research*, *71*, 369-376.

Stone, J., LaFrance, W. C., Levenson, J. L., & Sharpe, M. (2010). Issues for DSM-5: Conversion disorder. *American Journal of Psychiatry*, *167*, 626-627.

Thomsen, B. L. C., Teodoro, T., & Edwards, M. J. (2020). Biomarkers in functional movement disorders: a systematic review. *Journal of Neurology, Neurosurgery and Psychiatry*, *91*, 1261-1269.

van der Stouwe, A. M., Elting, J. W., van der Hoeven, J. H., van Laar, T., Leenders, K. L., Maurits, N. M., & Tijssen, M. A. (2016). How typical are 'typical' tremor characteristics? Sensitivity and specificity of five tremor phenomena. *Parkinsonism and Related Disorders*, *30*, 23-28.

Wegrzyk, J., Kebets, V., Richiardi, J., Galli, S., de Ville, D. V., & Aybek, S. (2018). Identifying motor functional neurological disorder using resting-state functional connectivity. *Neuroimage: Clinical*, *17*, 163-168.

Williams, D. T., Ford, B., & Fahn, S. (1995). Phenomenology and psychopathology related to psychogenic movement disorders. *Advances in Neurology*, *65*, 231-257.