Assessment of Fatigue in Multiple Sclerosis

Introduction

Fatigue, with a prevalence of at least 65% [1, 2], is one of the most common symptoms of multiple sclerosis (MS) and has far-reaching effects on the quality of life and employment, even statistically taking into account the degree of motor impairment [3]. Fatigue is a subjective sensation of lack of energy and exhaustion which can only be measured directly in the form of verbal reports. Due to its primarily subjective character, the concept of fatigue includes various aspects of human experience and behavior. Therefore, its measurement is mainly based on questionnaires and clinical interviews. It should be kept in mind that the structure of these 2 data collection methods depends on the internal model the researcher employs. On the other hand, focus was on the increase in loss of power and fatigue under continuous stress, with a measurement for determining prognosis, as there are indications for a relationship to relapse frequency, a change from a clinically isolated syndrome (CIS) to an MS diagnosis with an increase of brain atrophy. All in all, it can be said, that despite the growing convergence of diagnostic criteria, the available studies of the objective as well as the subjective assessment of fatigue are still not sufficient and further research is needed.

The following overview will discuss different possibilities to assess fatigue with the help of questionnaires, clinical interviews or the objective measurement of cognitive performance. Furthermore, a structured guideline for the assessment of fatigue will be proposed. Clinical criteria for MS related fatigue include the main symptoms, their everyday relevance, the possible causal relation with the cause of disease (the underlying MS), as well as an exclusion of other possible somatic or psychological reasons. It is recommended to use the “Würzburger Erschöpfungs-Inventar bei Multipler Sklerose (WEIMUS)” (English: Würzburg Fatigue Inventory for Multiple Sclerosis) and especially the “Fatigue Scale for Motor and Cognitive Functions” (FSMC), as these questionnaires distinguish between motor and cognitive fatigue and due to the larger number of research studies using the FSMC. Moreover, the presence of depression, quality of sleep and daytime sleepiness should be assessed. In addition, general cognitive performance as well as performance during monotonous stimulation (vigilance) should be assessed. This may be relevant for the evaluation of a patient’s capacity to work, and for determining prognosis, as there are indications for a relation to relapse frequency, a change from a clinically isolated syndrome (CIS) to an MS diagnosis with an increase of brain atrophy. All in all, it can be said, that despite the growing convergence of diagnostic criteria, the available studies of the objective as well as the subjective assessment of fatigue are still not sufficient and further research is needed.

Specific questions that categorize the current state [4]. They do not record the experienced change in fatigue in certain stressed or un-stressed situations. Similarly, an attempt can be made to classify objective behavior. In the past, it was attempted to measure fatigue directly through the application of tests, comparing patients with and without fatigue while assuming different performance in the tests employed. On the other hand, focus was on the increase in loss of power and fatigue under continuous stress, with a measurement of either motor or cognitive effort. The following provides an overview of the results obtained through these methods. The focus is, of course, on the questionnaires for the assessment of fatigue, since these continue to be the core of the fatigue diagnostics.

Criteria for the Diagnosis of “Fatigue”

There are currently no uniform criteria for the clinical diagnosis of MS-related fatigue, but rather various recommendations for the definition of fatigue and the possibility of measurement using questionnaires. Therefore Kluger et al., relying on an earlier review article on fatigue [4], suggested criteria for determining fatigue in
patients with Parkinson’s disease [5] which has been adapted to MS (▶Table 1). Item A should provide differentiation from normal physiological tiredness (which would result in a value of 3), in that at least 4 of the possible 9 points should be confirmed for a diagnosis. Item B should include the impact on social and professional life. Item C should confirm the relationship with the primary disease, and Item D should rule out other possible causes of fatigue.

The definition of fatigue criteria in ▶Table 1 appears to be very successful; direct application to clinical practice is also recommended in MS patients (including Item D, which is not always assumed with the necessary specificity).

**Questionnaires for Diagnosing Fatigue**

**Diagnosis of clinically-relevant fatigue**

In recent years the functionality, validity and reliability of available German-language questionnaires for the measurement of fatigue have been analyzed and summarized in various reviews [6–14]. We shall refer to these reviews in order to briefly recap their results (presented in ▶Table 2) and shall then discuss additional aspects. Further English-language methods can be found in Khan, Amatya and Galea [13].

The functionality of questionnaires depends on their purpose. The focus is on 2 diametrically opposed issues. On the one hand, using questionnaires can be used to distinguish patients who show clinically-relevant fatigue from those who show no clinically-relevant fatigue. Questionnaires devoted to this purpose should have appropriate cut-off values that allow such discrimination. They should take into account all relevant aspects of fatigue and, if possible, provide different related cut-off values. In our view, the WEIMUS, FSMC as well as the MFIS would be suitable (▶Table 2). On the other hand, measurement of change is a focal point, that is, the extent to which change caused by an intervention affects a specific aspect of the experience of fatigue. In general such questionnaires are unidimensional and should have a high degree of reliability so that even small test–retest deviations can be detected as interpretable changes. As multidimensional questionnaires, both the WEIMUS and FSMC meet these criteria (▶Table 2).

With respect to construct validity, not all questionnaires fulfill the criterion of differentiated recognition of different aspects of fatigue. Thus, the Fatigue Severity Scale, which is certainly the most frequently used fatigue scale, does not distinguish between motor and cognitive fatigue, whereas e.g., the MFIS, WEIMUS and the FSMC provide such a distinction and also offer differentiated cut-off values. Furthermore, the MFIS questionnaire provides the dimension of socially-interactive fatigue. However, to date, there is no evidence of whether this dimension is truly an independent form of fatigue.

On the other hand, the difference between motor and cognitive fatigue is plausible based on outer appearances, and it has now also been shown that this aspect may have a prognostic significance. Our recent study [15] demonstrated that the indication of cognitive fatigue at time t1 predicts the development of cerebral atrophy and subclinical demyelination in the area of the corpus callosum as well as the relapse rate for the following 17 months. On the other hand, the presence of motor fatigue at time point t1 did not correlate with the rate of relapse or cerebral atrophy or subclinical demyelination in the region of the corpus callosum [15].
The varying dimensionality of the questionnaires could also be a reason that the correlation between different questionnaires is usually moderate but not necessarily high. Flachenecker et al. showed, for example, that the correlation between MFIS and FSS is in the range of $r = 0.56$, which corresponds to a shared variance of only 31 %, which also means that substantial proportions of the variance detected are differently elucidated [16]. According to our data based on 168 patients with MS who were interviewed in the context of different research projects, FSS and FSMC total scores correlate with an $r$ of 0.76, which corresponds to a shared variance of approximately 60 %.

Overall, however, it must be noted that comparative studies among the questionnaires and the external validity of the fatigue questionnaires are scarce, and even more rarely those which are concerned with the question of whether the grading of mild, moderate and severe fatigue is comparable and meaningful. However, the establishment of clinically-relevant fatigue would require precisely such external validation.

In the past, we carried out a series of studies that included, among other things, occupational activity and a whole series of questionnaires on fatigue, depression, sleep quality and daytime tiredness. We relied on the FSS and FSMC for the assessment of fatigue. The FSMC is the only fatigue questionnaire listed in ▶ Table 2 providing cut-off scores for grading the severity of fatigue. Breaking down our data of the 168 MS patients (see above) by the status of occupational activity: employed full-time ($n = 36$), working half-time ($n = 35$), not employed (generally women with children) ($n = 21$) and retired ($n = 76$) only the total value of the FSMC results in a significant difference among these groups, after statistical control. Retired people indicate significantly higher fatigue values than those in the employed groups. On the other hand, the results for the other questionnaires (as well as the FSS) are far from showing significant differences.

A corresponding validation of the fatigue scales would also be useful with regard to the clinical relevance of the proposed cut-off values, but the authors are unaware of such studies. In our database according to the FSS, 44.4 % of MS patients with no fatigue (FSS $< 4$) were employed fulltime and 38.8 % had at least moderate fatigue (FSS $> 5$). Among the retired, the ratio was 27.6 to 44.7 %. For the FSMC total, the ratio for full-time employees was 16.7 to 44.4 %, and for pensioners 3.9 %–80.3 %. According to the FSMC, 92 % of the pensioners suffer at least moderate fatigue (44.7 % according to the FMS). These numbers show that the external validity of the severity classification of the FSMC significantly exceeds that of the FSS. However, 44.4 % of full-time workers rate their fatigue as severe using the FSMC, thus documenting that the mere extent of reported fatigue should not be sufficient for deciding social-medical issues.

Fortunately, however, the reliability of most of the fatigue questionnaires in the reviews we have mentioned is considered to be adequate, as was shown by Flachenecker’s study (▶ Table 2) [6].

Regarding the practical aspect of the different amounts of time required to complete the questionnaires, in our view they all fall within a reasonably tolerable range (approx. 5–10 min)

**Assessment of clinically-relevant changes in fatigue experience**

Most fatigue questionnaires are not designed to assess short-term changes in the experience of fatigue (e.g., immediately after a relaxation exercise or 1–2 h after administration of L-dopa) because they presuppose internal averaging of the experience over a period of time (refer to the relevant user instructions). Because of this, researchers usually rely on so-called visual analog scales. However, this change to another mode of representation is less useful, since different studies show that such analog scales correlate weakly to moderately to more comprehensive fatigue scales.

---

**Table 2** Overview of the German-language fatigue questionnaires.

<table>
<thead>
<tr>
<th>Name</th>
<th>Number of items</th>
<th>Forms of fatigue</th>
<th>Construct validity</th>
<th>Retest reliability</th>
<th>Scope of application</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSS [9]</td>
<td>9</td>
<td>General fatigue</td>
<td>Excellent [7, 10]</td>
<td>Acceptable (more than 6 months) [7, 6]</td>
<td>– Screening for fatigue</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Particularly for measuring motor fatigue [7]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Follow-up [7]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Classification of the severity of fatigue [13]</td>
</tr>
<tr>
<td>MFIS [14]</td>
<td>21</td>
<td>Cognitive, motor and psychosocial fatigue</td>
<td>High (particularly with respect to motor aspects of fatigue) [7]</td>
<td>Acceptable (more than 6 months) [6, 7]</td>
<td>– Influence of fatigue</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– In particular measurement of fatigue on motor function [7]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Follow-up [7]</td>
</tr>
</tbody>
</table>

FSS: Fatigue Severity Scale, WEIMUS: Würzburger Erschöpfungs-Inventar bei Multipler Sklerose, (Würzburger Fatigue Inventory for Multiple Sclerosis), FSMC: Fatigue Scale for Motor and Cognitive Functions, MFIS: Modified Fatigue Impact Scale
Conventional fatigue scales can also be used for change measurement if assessment of a longer-term intervention with a gap of at least 2 weeks is needed. However, little or nothing is known about their sensitivity and specificity in the detection of changes in fatigue experience. Based on the literature, Learmonth et al. concluded that the FSS is suitable for measuring changes in the fatigue experience, although no corresponding information is available for most of the other scales [7]. In our non-interventional follow-up study [15], in which MS patients as well as healthy controls were examined at 2 time points at an interval of 17 months with respect to existing fatigue symptoms as well as changes in imaging, a relatively high correlation was found for the subscales of the FSMC, whereas the FSS did not demonstrate high time stability (Fig. 1).

Likewise, the study by Johansson et al. showed a marked fluctuation of fatigue experienced over time when assessed with the FSS [17]. Given the absence of an external criterion and the lack of targeted observation studies, it is impossible to indicate whether this variation in FSS values is an expression of higher change sensitivity or inaccurate measurement. In any case, our results show that a change in fatigue measurable with the FSMC would have a high significance, as the values in this questionnaire obviously fluctuate little over time.

A study by Learmonth et al. to determine the psychometric properties of FSS and MFIS over 6 months also measured clinically-relevant change. For the FSS, 1.9 points (38% of the total score) indicate a clinically-relevant change. For the MFIS, 20.2 points are necessary to achieve a clinically-relevant change in the score [7].

Svenningsson, Falk, Celius, et al. [18] investigated the effect of natalizumab on the perception of fatigue. The FSMC was used to assess subjectively perceived fatigue, and a change of 9 points in the total score was considered clinically-relevant.

Lacking further studies, the definition of clinically-relevant changes in fatigue questionnaires is largely subjective. If classification is based on the degree of severity, the score between two degrees of severity could be considered as clinically relevant, pragmatically. In the case of the FSMC this would be 6 points per scale (that is, for the cognitive and motor scale). If a change of ½ SD would be considered clinically relevant for the FSMC, then according to our data, this would represent a change of 4 points. This corresponds with the classification of Svenningsson, Falk, Celius et al., since the calculated 4 points per subscale necessary for a clinical change coincide almost with the 9 points necessary for the overall scale [18].

For the FSS a change by one average point would be relevant (i.e., between light fatigue, FFS value of 4, and moderate fatigue with an FFS value of 5).

Measurement of the responsive experience of fatigue

Distinguishing between patients with and without fatigue is an issue in everyday clinical practice. A measure of the subjective increase in fatigue under standardized conditions would be of great importance for the assessment of socio-medical questions as well as for the relationship between fatigue and tiredness. In principle this issue would relate to psychophysics. Which standardized change in the objective burden or relief of MS patients corresponds to what difference in fatigue?

In fact, there have been very few attempts at such a psychophysical measurement of fatigue; for the most part fatigue and the manifestation of fatigue at time of onset are purely additive. Thus Moller et al. [19] demonstrated that with subjectively the same motor effort, healthy controls and MS patients respond with the same degree of additional fatigue over time. The induced fatigue after motor stress was as high in non-patients as in MS patients and was therefore additive to the fatigue at the time of onset. Similar results were also obtained by Lehmann et al. in a cross-over experiment in which MS patients and healthy controls spent over 10 min on a working memory task, then watched a stimulating or relaxing video sequence before resuming the working memory activity [20]. The group which was first allowed to relax, then received stimulation, while the other group was allowed to relax. The result showed that the extent of induced fatigue was the same in all groups (the MS group was divided into those with and without fatigue) and also did not depend on the interim conditions. In the past we systematically analyzed the existing literature on this topic [21]. Overall, there is so far no significant evidence that MS patients and healthy...
controls differ in the increase in fatigue experience when they are well as Claros-Salinas, Dittmer, Neumann et al. [29]). Thereby it can equally burdened. Those MS patients starting at a higher initial level not be assured that the decline of the reaction time can be attrib- do not differ in relative increase.

**Measurement of Fatigue using Neuropsychological Test Procedures**

At its core, fatigue is a vegetative-cognitive syndrome [4], corre- spondingly it is presumed that fatigue is best measured using neu-ropsychological tests, thus enabling identification of chronic traits value, so that the effect of mild depression cannot be fully ruled out [4]. However, as 2 reviews have shown, measuring fatigue using (see [26, 29]). In addition, the correlation with the fatigue experi- neuropsychological testing is highly complex [21, 22]. To do this, ence did not relate to the characteristic value of phasic alertness, 2 strategies must be distinguished. On the one hand, cognitive dif- ferences between patients with and without fatigue must be compensated by a previous acoustic cue [26].

In tests that vary very finely in time, this correlation can cause pa- thological effects. Inspired by animal experiments, theoretically motivat- ed replication would be desirable, e.g., between contextually in- strategy in the measurement of fatigue analyzes the effect on the fluenced memory performance and those without the requirement test process, the so-called time-on task effect, and this could for consideration of context-related information [25]. Neverthe- actually represent a central variable [33]. Fot the PASAT several re- less, until such studies have been established, it is not possible to search groups have shown (although not always reproducible) that distinguish MS patients with fatigue from those who do not suffer from fatigue groups at the end of the test [34]. For the group of vigilance tests it could also be demonstrated (with a nota- bly higher replication rate) that MS patients with fatigue show a holding attention over a period of time, or vigilance, i.e. maintaining slowed response time from 20 min onwards and, in some cases, an attention in the face of monotony, seem to be an exception in this increased rate of disregard of critical stimuli. Prior studies on alert-context. Testing procedures to investigate attention could include ness testing indicated that their sensitivity is high if testing is repeat- the Alertness and Vigilance Test of the TAP as well as the Symbol Digit ed after a phase of intensive load. In this respect, the duration of a ed Test. Especially regarding the Alertness Test there are a load with a more monotonous character seems to be a decisive fac-number of studies suggesting a valid variable selectivity among MS tor in determining whether fatigue is reflected in the measurement of vigilance. For the Alertness Test and the Symbol Digit Modalities Test it could also be demonstrated (with a nota- bly higher replication rate) that MS patients with fatigue show a holding attention over a period of time, or vigilance, i.e. maintaining slowed response time from 20 min onwards and, in some cases, an attention in the face of monotony, seem to be an exception in this increased rate of disregard of critical stimuli. Prior studies on alert-context. Testing procedures to investigate attention could include ness testing indicated that their sensitivity is high if testing is repeat-

The studies by Weinges-Evers, Brandt, Bock et al. [27] are an ex-

These remarks concern only the direct application of relatively showing a relation between subjective fatigue and cognitive per-
formance. Inspired by animal experiments, theoretically motivat-
ed replication would be desirable, e.g., between contextually in-
strategy in the measurement of fatigue analyzes the effect on the fluenced memory performance and those without the requirement test process, the so-called time-on task effect, and this could for consideration of context-related information [25]. Neverthe-

On the other hand, tests, which focus on sustained attention, i.e.

The Alertness and Vigilance Test of the TAP as well as the Symbol Digit ed after a phase of intensive load. In this respect, the duration of a ed Test. Especially regarding the Alertness Test there are a load with a more monotonous character seems to be a decisive fac-num-ber of studies suggesting a valid variable selectivity among MS tor in determining whether fatigue is reflected in the measurement of vigilance. For the Alertness Test and the Symbol Digit Modalities Test it could also be demonstrated (with a nota-

Summary

The correlative relationship between fatigue experience and About 30 years after Krupp et al. designed the Fatigue Severity extent of neurological impairment has been repeatedly replicated. Scale [9], adequate measurement of fatigue remains a central prob-
In tests that vary very finely in time, this correlation can cause pa-
lem for the clinical practice. Similar to other symptoms (such as de-
tients with fatigue to perform worse because their fine motor per-
ession) which recur due to purely internal sensations, diagnosis formance is often lower. This applies to the Alertness Test as well remains based on clinical plausibility assumptions.

as the Symbol Digit Modalities Test, in which symbols or numbers This short survey article shows that especially when it comes to have to be written in small boxes, and the number of the correct to social-medical considerations, the mere use of questionnaires boxes within a time span is the critical variable. A further problem is insufficient. Clinical judgment based each entry in ▶ Table 1 is occurs if patients with cognitive fatigue are compared with healthy recommended for diagnosis. This presumes targeted questioning con-
trasts as well as Claros-Salinas, Dittmer, Neumann et al. [29] as Using the FFS exclusively appears insufficient to get an impression

of the experience of fatigue. Due to more extensive evaluation, the WEIMUS is recommended as well as the FSMC for the German-speaking region. Both differentiate between motor and cognitive fatigue. In addition, the assessment of depression, sleep quality and daytime sleepiness is supported by questionnaires such as the Beck Depression Inventory [35] (excluding the somatic items), the Hospital Anxiety and Depression Scale (HADS) [36], the Pittsburgh Sleep Quality Index (PSQI) [37] and the Epworth Sleepiness Scale (ESS) [38]. In terms of research, external validation of the degrees of severity of the fatigue scales, for example, decisions relating to shortening working hours, represents a simple to accomplish – but urgently necessary – task.

In addition to surveying MS patients, objectification of the fatigue experience should be attempted through psychological testing. This is the case because the experience of fatigue as such is not equivalent to the inability to work, which must be subject to a separate consideration; conversely, cognitive deficits can also be present without the presence of fatigue. In the context of neuropsychological research, in addition to the determination of cognitive performance, the assessment of performance under monotony is a focal point. Vigilance tests requiring maintenance of continuous performance, the assessment of performance under monotony is a focal point. Vigilance tests requiring maintenance of continuous performance under monotony is required. Vigilance tests requiring maintenance of continuous performance under monotony are recommended as well as the FSMC for the German-speaking region. Both differentiate between motor and cognitive fatigue. In addition, the assessment of depression, sleep quality and daytime sleepiness is supported by questionnaires such as the Beck Depression Inventory [35] (excluding the somatic items), the Hospital Anxiety and Depression Scale (HADS) [36], the Pittsburgh Sleep Quality Index (PSQI) [37] and the Epworth Sleepiness Scale (ESS) [38]. In terms of research, external validation of the degrees of severity of the fatigue scales, for example, decisions relating to shortening working hours, represents a simple to accomplish – but urgently necessary – task.

Objectification of the fatigue experience plays a role beyond social-medical decisions. Recent studies show that the experience of fatigue could also play a prognostic role in whether MS patients will experience a relapse in the coming months [15], will change from CIS status toward a diagnosis of MS [40] or suffer increased cerebral atrophy [15, 41, 42]. Furthermore, there are very limited indications that the choice of immunomodulatory medication may offer a more comprehensive assessment of fatigue in MS? Multiple Scler 2005; 11: 198–202


Conflict of interest

No conflict of interest has been declared by the authors.

References


