The Mental Disability Military Assessment Tool: A Reliable Tool for Determining Disability in Veterans with Post-traumatic Stress Disorder

Andrea S. Fokkens · Johan W. Groothoff · Jac J. L. van der Klink · Roel Popping · Roy E. Stewart · Lex van de Ven · Sandra Brouwer · Jolanda Tuinstra

Abstract  Purpose An assessment tool was developed to assess disability in veterans who suffer from post-traumatic stress disorder (PTSD) due to a military mission. The objective of this study was to determine the reliability, intra-rater and inter-rater variation of the Mental Disability Military (MDM) assessment tool. Methods Twenty-four assessment interviews of veterans with an insurance physician were videotaped. Each videotaped interview was assessed by a group of five independent raters on limitations of the veterans using the MDM assessment tool. After 2 months the raters repeated this procedure. Next the intra-rater and inter-rater variation was assessed with an adjusted version of AG09 computing weighted percentage agreement. Results The results of this study showed that both the intra-rater variation and inter-rater variation on the ten subcategories of the MDM assessment tool were small, with an agreement of 84–100 % within raters and 93–100 % between raters. Conclusions The MDM assessment tool proves to be a reliable instrument to measure PTSD limitations in functioning in Dutch military veterans who apply for disability compensation. Further research is needed to assess the validity of this instrument.

Keywords Veterans disability claims · Reliability · Insurance disability · Instruments · Post-traumatic stress disorder

Introduction

Soldiers in combat are exposed to extreme conditions. These conditions can be experienced as a traumatic event, which may cause psychological and psychiatric disabilities. The most common psychiatric disorder resulting from war-related trauma is post-traumatic stress disorder (PTSD) [1]. PTSD is characterized by increased tension, reliving the traumatic event and avoidance of stimuli related to the traumatic event [2]. PTSD can be a chronic, severely disabling condition causing sustained loss of functionality [3–5].

The prevalence of combat-related PTSD among military personnel and veterans is variable due to heterogeneity of methods used across studies [6]. Reported point prevalence rates showed that combat-related PTSD in US military veterans since the Vietnam war ranges from 2 to 17 %, in US Iraq war veterans between 4 and 17 % and in UK Iraq war veterans between 3 and 6 % [6]. In The Netherlands a screening study showed a current prevalence between 1 and 17 % in veterans from various Dutch missions [7].

Diagnosing a person with PTSD is difficult, due to the multiplicity of symptoms and the changing disease criteria over the years [8–10]. Moreover, it is difficult to determine if the PTSD is combat-related. PTSD as a single disorder following trauma is relatively uncommon: about 80 % of
cases of PTSD arise in the context of comorbidity, particularly with depression, anxiety and substance misuse [9, 11–13]. Furthermore, the length of time elapsed since the trauma occurred, may lead to the question as to whether the diagnosis might be influenced by other factors. Mental and/or physical disability is a frequent consequence of military service and some veterans are fully or partially disabled [14–16].

To compensate veterans for illness or injuries incurred during military service, military-specific disability compensation programs exist. In contrast to other disability compensation programs, the military program is an indemnity program providing financial payments to veterans who are partially or totally disabled by injuries or illnesses that were incurred or aggravated during military service [14]. This compensation is independent from employment and is intended to compensate veterans for illness or injuries incurred during military service [14, 17].

To determine the severity and disabilities due to combat-related PTSD, an assessment is made by the insurance physician (IP), consisting of psychological health status evaluation, and the use of medical and military service records. To measure the severity, several instruments have been developed to assess (combat-related) PTSD symptomatology [6, 18–24]. All these instruments have been developed to measure symptom severity or functional impairments, rather than characterizing a person’s overall functioning.

In the disability compensation assessment, not only the symptomatology, but the disability status is also influenced by functional limitations. Therefore, measuring functional limitations is an important part of the evaluation for the disability pension and have great individual, financial and social consequences. Several instruments are available to provide information about functional limitations in the context of the disability compensation [25], however, they have not been developed to measure severity of PTSD and the functional limitations among veterans.

In the Netherlands, disability evaluations of combat veterans are performed by IPs of a specialized care office of military service and war victims. In 2008, an assessment tool was developed by the Dutch ministry of defense to assess PTSD severity and limitations in functioning in veterans who suffer from trauma or stress-related disorders due to duty. The Mental Disability Military (MDM) assessment tool (in Dutch Psychische Invaliditeit Militairen) consists of a scoring list, which is filled in by the IP after an interview and health status examination [26]. Although this instrument has been used for a couple of years in daily practice, its reliability has not been studied.

In the present study, the aim was to determine the reliability, intra-rater and inter-rater variation of the MDM assessment tool for determining severity of PTSD and functional limitations in military veterans.

Methods

Design

In this reliability study the intra- and inter-rater variation of the MDM assessment tool was determined. Twenty-four videotaped assessment interviews were performed by IPs rated twice by five different raters.

This study adhered to the ethical standards of the Declaration of Helsinki and guidelines of the association of universities in the Netherlands [27]. Ethical clearance was provided by the Medical Ethics Committee of the University Medical Center Groningen.

Sample

The veterans included in this study were former Dutch soldiers, post-active for longer than 66 months, who went on a military operation and claimed military pension because of trauma-or stress related disorders due to duty. For inclusion into the study the following criteria had to be met:

1. Initial MDM assessment only.
2. Veterans between the ages of 18 and 64 years old.
3. The interview took place in the insurance physicians’ office.
4. During the videotaped assessment interview at least six out of ten subcategories were discussed.

Recordings of reassessments have a risk of missing relevant information addressed in an earlier interview. Veterans who, for whatever reason (physical reasons, detention), were unable to reach the insurance office, were excluded for this study. Included veterans for this study were informed by their case manager. If the veteran gave permission for filming the assessment interview an informed consent was signed. The interview could then be videotaped.

Of the 44 veterans that were approached for this study, 28 (64 %) gave permission for filming the interview, 12 refused and 4 did not respond. Reasons for not giving permission were: feeling discomfort with camera (n = 10) and emotional too stressful (n = 2). Of the 28 veterans, four were excluded, because the interview was held on another location (n = 2), it was a re-assessment (n = 1) and mainly somatic disorder (n = 1). In total 24 (55 %) interviews with veterans were videotaped.
Videotaped Assessment Interview

Five IPs consented in videotaping their interviews of which one IP was a reserve in case an IP was unable to attend. Videocconferencing was used, so there was no extra person in the room that could influence the assessment. On film the veteran is hearable and visible; the IP can only be heard. The interview was held according the protocol. The IP had the possibility to indicate if there were any peculiarities. They also filled in a short form with background information of the veteran for the raters. The IP could stop the filming at any time if necessary, i.e. on request of the veteran.

Next, all videotaped interviews were assessed by two external experts (JH Wijers, SK Knepper) who were involved in the development of the MDM. Videotapes in which at least six out of ten subcategories of the MDM were discussed during the interview were included. All 24 videotaped interviews met this criterion.

Raters

In total 14 independent IP raters participated, all familiar with work disability evaluations, but not in military context. They were trained in using the MDM assessment tool, by the same trainers that trained the IPs working with the assessment tool.

The Mental Disability Military Assessment Tool (MDM)

The MDM assessment tool consists of a scoring list, which is filled in after an interview and patient examination by the IP. The IP determines also causality, whether the limitations are due to duty. The question of causality is not part of the present reliability study.

With the information from the interview the IP fills in the MDM and determines the percentage disability. The MDM consists of four categories “Activities of daily living”, “Social functioning”, “Concentration, perseverance and speed” and “Adaption to stressful situation”. These four categories are composed from ten subcategories, see Box 1. Disability severity is classified per subcategory in rating classes from “0” to “5”, “0” is normal, “5” is extreme limited. Every rating class for every subcategory is described in detail in the assessment tool. The scores on the subcategories are converted towards a disability percentage. All four categories, independent of the number of subcategories, weight equally in the calculation. The ratings of all subcategories, of the same category, are added and subsequently divided by the number of subcategories. These scores of the four categories are added and divided by four. This quotient, a number between 0 and 5 represents the percentage between 0 and 100 %. As on every subcategory rating class “1” is rated, the percentage is 20 %, and as on every subcategory rating class “5” is rated, the percentage is 100 %.

Procedure

The 24 videotaped assessment interviews were rated, using the MDM assessment tool. Each interview was assessed by five raters for the inter-rater variation, and 2 months later for the second time by the same group of five raters for the intra-rater variation. For every videotaped interview a unique combination of five raters out of the 14 raters was composed. Every rating day started plenary with a repetition of the tool instructions. Background information of the veterans was available for the raters. Each rater watched and rated the videotaped interview separately, consultation was not possible.

The raters filled in the form with the ten subcategories, marked rating class “0” to “5”. The researchers converted these scores afterwards into the disability percentage. Besides the rating class “0” to “5” an extra rating class was added in the form, “not discussed”, in case the rater could not mark a rating class because he/she did not had enough information from the videotaped assessments.

Analysis

First, the distribution on the five rating classes was assessed of every subcategory, including the additional rating class “not discussed”. Analyses were performed on 24 assessments rated by five raters, which makes in total 120 scores: 24 multiplied by 5.

Next the intra-rater variation and inter-rater variation was assessed. The computation of an agreement index correcting for expected agreement like Cohen’s kappa or Scott’s Pi was not possible [28, 29], due to skewed distributions of the rating classes. This will cause the amount of expected agreement is overestimated. Therefore the final index will get a value that is too low. A requirement for the use of this index is that the rating classes are used with more or less the same frequency [30, 31]. The statistical software package AG09 was used to compute a weighted percentage agreement [32]. This package calculates an average squared weighted percentage of observed agreement between all pairs of raters based on an algorithm proposed by Cicchetti [33]. The weighting will take the interval character of the rating classes into account. Generally, a percentage agreement of 70 % or more is considered “good” and more than 90 % is considered “excellent” [34].

Differences between the outcome measure, i.e. the disability percentages on the first and second rating within the raters is presented in Fig. 1. Also the range between the disability percentages between the five raters is presented in Fig. 2. In all comparisons the rating class “not
discussed” is recoded as rating class “0”, normal. The effect of this on the results was tested by making also a comparison in percentage agreement when the rating class “not discussed” is recoded as missing.

**Results**

**Characteristics**

The characteristics of the veterans, the IP, the raters and the videotaped interviews are presented in Table 1. The veterans were mostly male with a mean age of 43.3 years.

**Distribution of the Rating Classes**

Table 2 shows the distribution of the rating classes of the different subcategories. The distribution was skewed, rating class “0 “ normal was often scored, class “5” extreme limitations was never scored. On the subcategory 3 “Sleep” the highest rating classes were scored, followed by subcategory 10 “coping with stressful events”, subcategory 7 “communicative (emotional) skills”, and subcategory 6 “communicative social functioning”. For most subcategories “not discussed” was scored rarely (3.3 %). Only for

**Table 1** Characteristics of veterans, raters, IP, and filmed interviews

<table>
<thead>
<tr>
<th>Veterans</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>24</td>
<td>Age, mean (SD)</td>
<td>43.3 (6.3)</td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>23 (96 %)</td>
<td>Mission*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Libanon</td>
<td>9 (37.5 %)</td>
<td>Yugoslavia</td>
<td>15 (62.5 %)</td>
<td></td>
</tr>
<tr>
<td>Iraq</td>
<td>1 (4.2 %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raters</td>
<td></td>
<td>Age, mean (SD)</td>
<td>50.5 (6.2)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>14</td>
<td>Male (N)</td>
<td>9 (64.3 %)</td>
<td></td>
</tr>
<tr>
<td>Years of work experience, mean (SD)</td>
<td>16.3 (7.3)</td>
<td>Rated films per rater, mean (SD)</td>
<td>8.6 (2.8)</td>
<td></td>
</tr>
<tr>
<td>Insurance physicians</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>Maleb (N)</td>
<td>4 (80 %)</td>
<td></td>
</tr>
<tr>
<td>Filmed interviews per IPc, mean (SD)</td>
<td>5.8 (1.0)</td>
<td>Film duration (seen by raters), mean (SD)</td>
<td>1:52 h (26 min)</td>
<td></td>
</tr>
</tbody>
</table>

---

* More than one mission per person possible
b One interview was taken by a female rater, because of sickness of one rater who participated in the study
c IP = insurance physician

---

**Fig. 1** The difference of disability percentage between the first and second rating per rater (number of cases N = 120: N = 24 films* N = 5 raters)

---

**Fig. 2** Range of outcome measure, i.e. percentage disability between the 5 raters, per film N = 24
the subcategory 4 "sexual functioning" "not discussed" was scored more often (21.7 %).

Intra-Rater Variation

Table 3 presents the percentages of quadratic weighted agreement between the first and second rating (2 months later) within the raters. All subcategories showed a percentage of agreement of at least 84 %. For subcategory 1 "personal hygiene and self-care" every rater had a 100 % agreement. With a range from 84 to 100 % subcategory 4 "sexual functioning" had most variation in scored rating classes between the first and the second rating. When the rating class "not discussed" was analyzed as missing instead of "0" the percentage agreement of nine subcategories remained the same, only for the subcategory "sexual functioning" the range changed to 78–100 % (was 84–100 %, see Table 3).

Figure 1 presents the differences between the first and the second rating per rater in disability percentage. In 98 of the 120 cases (81.2 %) the first and second rating differed 5 % or less. Differences were normally divided, the second rating was not consequently higher or lower than the first rating.

Inter-Rater Variation

The percentage agreement between raters for the different subcategories is presented in Table 4. The range of inter-rater variation over the subcategories was 93–100 %. When the rating class "not discussed" was analyzed as missing instead of “0” the percentage agreement of nine subcategories remained the same, only for the subcategory "sexual functioning" the percentage agreement changed to 78 % (was 94 %, see Table 4).

Figure 2 presents the range of percentage disability between the five raters per film. In 17 films (70.8 %) the range between the lowest scoring rater and highest scoring rater per group of raters and per videotaped interview was <10 %. One film showed a large range with 30 %. On this

<table>
<thead>
<tr>
<th>Subcategory</th>
<th>Percentage agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Personal hygiene and self-care</td>
<td>100</td>
</tr>
<tr>
<td>2 Mobility</td>
<td>92–100</td>
</tr>
<tr>
<td>3 Sleep</td>
<td>95–99</td>
</tr>
<tr>
<td>4 Sexual functioning</td>
<td>84–100</td>
</tr>
<tr>
<td>5 Basic communication</td>
<td>98–100</td>
</tr>
<tr>
<td>6 Communicative social functioning</td>
<td>93–99</td>
</tr>
<tr>
<td>7 Communicative (emotional) skills</td>
<td>86–98</td>
</tr>
<tr>
<td>8 Structure</td>
<td>95–100</td>
</tr>
<tr>
<td>9 Household activities</td>
<td>94–100</td>
</tr>
<tr>
<td>10 Coping with stressful events</td>
<td>96–100</td>
</tr>
</tbody>
</table>

All filmed interviews together, per rater, the number of rated interview varies per rater from 4 to 13. "not discussed" is included as normal in this calculation

a Quadratic weighted percentage agreement, weighting used is fixed number of classes

Table 2 Distribution of classes of MDM subcategories, first rating (for each subcategory total N = 120)

<table>
<thead>
<tr>
<th>Subcategory</th>
<th>Normal 0</th>
<th>Slight limitations 1</th>
<th>Mild limitations 2</th>
<th>Moderate limitations 3</th>
<th>Severe limitations 4</th>
<th>Extreme limitations 5</th>
<th>Not discussed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Personal hygiene and self-care</td>
<td>117 (97.5)</td>
<td>1 (0.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 (1.7)</td>
</tr>
<tr>
<td>2 Mobility</td>
<td>92 (76.7)</td>
<td>25 (20.8)</td>
<td>2 (1.7)</td>
<td></td>
<td></td>
<td></td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>3 Sleep</td>
<td>17 (14.2)</td>
<td>30 (25)</td>
<td>55 (45.8)</td>
<td>12 (10)</td>
<td>5 (4.2)</td>
<td></td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>4 Sexual functioning</td>
<td>82 (68.3)</td>
<td>4 (3.3)</td>
<td>5 (4.2)</td>
<td>2 (1.7)</td>
<td>1 (0.8)</td>
<td>26 (21.7)</td>
<td></td>
</tr>
<tr>
<td>5 Basic communication</td>
<td>117 (97.5)</td>
<td>2 (1.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>6 Communicative social functioning</td>
<td>43 (35.8)</td>
<td>50 (41.6)</td>
<td>16 (13.3)</td>
<td>5 (4.2)</td>
<td>6 (5.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Communicative (emotional) skills</td>
<td>43 (35.8)</td>
<td>42 (35)</td>
<td>28 (23.3)</td>
<td>5 (4.2)</td>
<td>2 (1.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Structure</td>
<td>80 (66.7)</td>
<td>29 (24.2)</td>
<td>8 (6.7)</td>
<td></td>
<td></td>
<td></td>
<td>3 (2.5)</td>
</tr>
<tr>
<td>9 Household activities</td>
<td>93 (77.5)</td>
<td>17 (14.2)</td>
<td>8 (6.7)</td>
<td></td>
<td></td>
<td></td>
<td>2 (1.7)</td>
</tr>
<tr>
<td>10 Coping with stressful events</td>
<td>28 (23.3)</td>
<td>63 (52.5)</td>
<td>21 (17.5)</td>
<td>3 (2.5)</td>
<td>1 (0.8)</td>
<td>4 (3.3)</td>
<td></td>
</tr>
</tbody>
</table>
film the rater with the highest outcome of percentage disability scored relatively high on the subcategory 10 “Coping with stressful events” with rating class “4” compared to class “2” of the other four raters. The raters with the lowest outcome of percentage disability on this film scored relatively low on subcategory 6 “Communicative social functioning” with rating class “1” compared to rating class “4” of three other raters and “3” of one other rater.

Discussion

Given the great individual, financial and social consequences of disability assessments, it is important to use assessment tools which provide reliable and valid information to enable appropriate decisions about the work disability pension. In this study, the reliability of the MDM has been examined, by evaluating both intra- and inter-rater variability. The results of this study showed that both the intra-rater variation and inter-rater variation on the ten subcategories of the MDM assessment tool were small, with an agreement of 84–100 % within physicians and 93–100 % between physicians. Reliability is good to excellent.

A limited number of studies are available that evaluated the reliability of assessment tools used in the disability compensation context. Spanjer et al. [25] developed the Disability Assessment Structured Interview (DASI) and found reasonable to good inter-rater reliability between IPs. Slebus et al. [35] developed a checklist for work ability assessment in prolonged depressive illness as part of the Dutch disability evaluation by IPs of the Social Security Institute. In their vignettes study they found that the variation between IPs did not diminish by using the checklist. Due to variation in research design (number of raters, time interval), methods (vignettes, videos, real-life assessments), outcome (percentage agreement, Cohen’s kappa, Scott’s pi, intra-class correlations) and context (employees vs. veterans) we cannot compare our results with those studies. Nonetheless, taking into account these differences, the percentages of agreement, found in our study are high, in the context of other published results.

Two subcategories of the MDM showed a ceiling effect and a skewed distribution of the scores: “Personal hygiene and self-care” and “Basic communication”. Of all scores (N = 120) of these two subcategories 97.5 % was scored as normal (rating class 0). This could indicate that the MDM is not performing optimally for these two subscales and should be further examined in a validity study. More distribution in scores was found for the subcategories “Sleep”, “Communicative social functioning”, “Communicative (emotional) skills” and “Coping with stressful events”, which may indicate that these subcategories are more related to PTSD-related complaints.

In general in about 3 % of the assessments reports the raters marked “not discussed” in the report form. It could not be scored because that specific topic was not discussed during the interview, or that the rater missed the information during video-watching. Remarkably, “not discussed” was found in 21 % for the subcategory “Sexual functioning”, which may indicate that talking about this topic may be difficult for the veteran as well as the IP and therefore not covered. Further evaluation study is necessary to determine the reason behind this finding and whether it influences the outcome of the MDM results.

The results suggest a bit higher between physician agreement than within physician agreement, however the percentage agreements are not comparable. The inter-rater reliability, between physicians, is presented as the mean agreement between five physicians (Table 4). The intra-rater reliability, within physicians, is presented as the range of five individual agreement percentages between the first and second rating (Table 3).

Strengths and Limitations

An important strength of the study is that videotapes are used which enables multiple rater assessments. Moreover, the 2-month interval between the ratings for analyzing the intra-rater reliability was considered sufficient to prevent raters from recollecting their first ratings. There is no golden standard for interval periods as it varies in the literature from 2 days till 6 months [36–39]. In the evaluation of the data-collection of this study the IPs stated they had

<table>
<thead>
<tr>
<th>Subcategory</th>
<th>Percentage agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Personal hygiene and self-care</td>
<td>100</td>
</tr>
<tr>
<td>2 Mobility</td>
<td>99</td>
</tr>
<tr>
<td>3 Sleep</td>
<td>97</td>
</tr>
<tr>
<td>4 Sexual functioning</td>
<td>94</td>
</tr>
<tr>
<td>5 Basic communication</td>
<td>100</td>
</tr>
<tr>
<td>6 Communicative social functioning</td>
<td>93</td>
</tr>
<tr>
<td>7 Communicative (emotional) skills</td>
<td>93</td>
</tr>
<tr>
<td>8 Structure</td>
<td>98</td>
</tr>
<tr>
<td>9 Household activities</td>
<td>98</td>
</tr>
<tr>
<td>10 Coping with stressful events</td>
<td>97</td>
</tr>
</tbody>
</table>

N = 24 filmed interviews and 5 raters per film

Table 4 Inter-rater reliability, paired quadratic weighted percentage agreement between raters per subcategory

Paired quadratic weighted agreement, weighting used is fixed number of classes “not discussed” is included as normal in this calculation

Springer
no recollection of their previous ratings. Another strength is the use of the statistical package AG09 that computes weighted agreement, taking into account the interval character of the rating scales. There are few computer programs that can perform these analyses. AG09 computes the average weighted agreement over all pairs of rates, while other programs compute unweighted agreement per pair of raters and generally have no possibility to use weights.

A limitation of this study is the items that are filled in as “not discussed”. These items were coded in the analyses as normal, i.e. “normal, or no limitation”. One might ask whether no discussion implies no limitations. However, it is not very likely that in case of severe problems the subcategory was not discussed. Additional analyses were performed, whereas omitting the rating class “not discussed” out of the data file as missing cases, remaining the same results and conclusions.

Another relevant remark in the present study is the possibility of a learning effect of the raters. During the assessments the raters become more experienced in using the assessment tool, which theoretical can lead towards systematic patterns in results: (1) systematic higher or lower scores of the second rating compared to the first rating, and (2) systematic higher agreement during the data collection. Both types of learning effects have been tested and are not found. In addition, sub analyses showed that no differences exist in agreement percentages during the data collection.

A study limitation might be a selection bias of the study sample as we found that extreme limitations were never rated, most of the ratings were mild in severity. However, we have no indications that our finding differs from practice and the results of a study of De Raad et al. [40] also suggest a rather low average disability pension. In addition, the group who gave permission for filming their interview assessment was rather high, i.e. 64 %.

A final remark concerns possible selection by including only veterans that where more than 66-months post-active and only first assessments. Delayed PTSD develops after an asymptotic period of only 6 months [14–16]. By a longer onset the chance of comorbidity increases and the diagnosis might be influenced by other personal- and social factors.

Implications

In the present study, a first important step has been conducted of testing the psychometric properties of the MDM. The next step should be to examine the construct validity of the MDM assessment tool by a comparison with another assessment tool aiming to determine disability in people with severe mental disorders. Although no golden standard is present in the assessment of disability for mental disorders, the assessment tool in the mental and behavioral disorder chapter of the AMA guides of the evaluation of permanent impairment 6th edition may be used to examine the construct validity of the MDM assessment tool. However, it should be noted that the MDM tool has been specifically developed to measure in the military context, the AMA instead is a generic tool [41].

Conclusion

The MDM assessment tool appears to be a reliable instrument to measure PTSD-related and functional limitations in Dutch military veterans who apply for disability compensation. Further research is needed to assess the validity of this instrument.

Conflict of interest

Andrea S. Fokkens, Johan W. Groothoff, Jac J. L. van der Klink, Roel Popping, Roy E. Stewart, Lex van de Ven, Sandra Brouwer and Jolanda Tuinstra declare that they have no conflict of interest.

References

41. Rondinelli RD, Genovese E, Brigham CR. Guides to the evaluation of permanent impairment. Chicago: American Medical Association; 2008;