Validation and Shortening Questionnaires Used in Surveys and Trials on Populations with Substance use Disorders: A Method using the Fatigue Severity Scale

Jørn Henrik Vold (jom.vold@uib.no)
Haukeland Universitetssjukehus

Rolf Gjestad
Haukeland Universitetssjukehus

Christer Aas
Universitetet i Bergen

Eivind Meland
Universitetet i Bergen

Kjell Arne Johansson
Universitetet i Bergen

Lars Thore Fadnes
Universitetet i Bergen

Research

Keywords: Substance-related disorders, fatigue, quality of life, validation study, fatigue severity scale, visual analog fatigue scale, activities of daily living

DOI: https://doi.org/10.21203/rs.3.rs-33806/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Precise and valid measurements of health outcomes and exposures among people with substance use disorders (SUDs) is essential to improve health services and health within this group. Unfortunately, many of the standardized questionnaires are validated on other populations and are often too comprehensive, and insufficiently adapted to the group. This may introduce limitations on several aspects that can be assessed but also biases due to research participation fatigue. New methods are needed to validate shortened and adapted questionnaires for this population. This study aims to present a method on how to shorten validated questionnaires and assure a construct validity when applied to SUD populations.

Methods

We used the data from a nested cohort with responses on a nine-item Fatigue Severity Scale (FSS-9), and Visual Analog Fatigue Scale (VAFS) collected from 655 people with SUD in Bergen and Stavanger, Norway, in the period 2016-2020. A total of 225 people filled out FSS-9 and VAFS at least twice. We defined baseline as the first measurement of FSS and VAFS when the measurements were sorted chronologically per participant. A three-step method was used for validation and shortening: Checking of internal consistency; longitudinal confirmatory factor analysis; linear mixed model analysis.

Results

The internal consistency of FSS-9 was excellent with a Cronbach's coefficient $\alpha$ on 0.94 at baseline and 0.92 at the second measurement. When shortening FSS-9 to a three-item FSS (FSS-3) (items 5-7), the Cronbach's coefficient $\alpha$ was 0.87 at baseline and 0.85 at the second measurement. The internal consistency was not affected when VAFS was added to FSS-9 and FSS-3, respectively. The longitudinal confirmatory factor analysis model between baseline and second measurement showed a well-fitted model for FSS-3 ($\chi^2 = 13.33$, degree of freedom $= 8$, $P = 0.101$) with a correlation on $r = 0.52$, $P < 0.001$ between the measurements. The linear mixed model analysis showed equal linear changes at individual level for FSS-3 (slope: 0.00, $P > 0.05$) and FSS-9 (slope: 0.01, $P > 0.05$) between the measurements.

Conclusion

The ten items could be shortened to a three-item version with excellent validity and reliability. This method could be useful for validating and shortening other questionnaires among patients with SUD and other populations.

1. Background

Substance use disorder (SUD) is a chronic relapsing disorder affecting two percent of the population worldwide [1]. Extensive medical and psychosocial health challenges, including temporary living situations, injecting drug use, unemployment, comorbid psychiatric and physical diseases, typically cause reduced life expectancy, great suffering, and a high burden of diseases in this population [1, 2]. These challenges may raise subjective health complaints impairing quality of life [1–4]. Therefore, high-quality research on these marginalized people is essential to get more precise evidence of SUD epidemiology and its treatment effects. Unfortunately, the vast majority of the standardized questionnaires are validated on other populations. Additionally, the questionnaires are often too comprehensive, particularly when combined with other questionnaires assessing other health facets, causing discontinuations and biased responses from people [5]. This may reduce the quality of studies conducted on populations with SUD. To get validated and shorter questionnaires for this population, standardized and systematic methods are needed to ensure well-fitted questionnaires and facilitate reliable results from people with SUD.

Fatigue is a common subjective health complaint involving a persistent and overwhelming feeling of exhaustion and loss of energy and is frequently reported in people with chronic diseases [6–10]. Different questionnaires to measure fatigue are developed with high validity and reliability. The nine-item Fatigue Severity Scale (FSS-9) and visual analog fatigue scale (VAFS) are two of the most commonly used questionnaires [11, 12]. The FSS is recognized by the items that are graded as a Likert scale from one to seven points for quantifying the level of fatigue [12]. Initially, the FSS was developed to evaluate fatigue in people with multiple sclerosis and systemic lupus erythematosus [11]. Later, the FSS-9 was showed to have excellent reliability and validity also for
people with other chronic diseases such as hepatitis C virus infection [13], Parkinson's disease [14], stroke [15], and myasthenia gravis [16]. For these chronic diseases, the scores were homogeneous across items when validated FSS-9, and the level of fatigue did not change substantially from one item to another through the questionnaire. Furthermore, VAFS is a 10 cm horizontal line ranged from zero to 10 that estimates the level of fatigue the last week, and was frequently correlated to the level of fatigue reported by FSS [15, 17, 18].

The FSS-9 has recently been a subject of being shortened in validation studies due to homogenous internal consistency across different study populations [19, 20]. For instance, a seven-item FSS was validated among people with human immunodeficiency virus (HIV) and stroke, showing better psychometric properties of fatigue compared to FSS-9 [19, 21]. To obtain high completion rates of questionnaires and ensure high construct validity and collaboration among people with SUD, the FSS is applicable for validation and shortening for this population.

Thus, this cohort study aims to present a method for shortening validated questionnaires and assess the longitudinal construct validity of people in need of shortened versions of measurement scales. In this case, a systematic three-step method including internal consistency, complementary factor analysis and linear mixed model analysis for shortening and validating the nine-item Fatigue Severity Scale (FSS-9) and the Visual Analog Fatigue Scale (VAFS) is presented by using data from people with SUD.

2. Methods

2.1. Data source

We drew data from a nested cohort from the INTRO-HCV trial collecting data on people with SUDs in Bergen and Stavanger, Norway [22]. We initiated data collection in May 2017 and recruited people receiving on opioid agonist treatment in Bergen and Stavanger or people with SUDs receiving healthcare from the primary health clinics in the municipality of Bergen. This study included all people that had filled out VAFS or FSS in the cohort in the study period from May 2017 to January 2020.

2.2. Data collections

All included people were assessed yearly with a questionnaire that included fatigue symptoms and the sociodemographic situation. We collected all data in the health register by using an electronic data collection software (Checkware®) under supervision from research nurses.

2.3. Study population

We included 917 measurements of FSS-9 or VAFS from 655 people in the study period. We defined measurement as when at least one of the items in the questionnaires were filled out in the annual health assessment. Of the measurements, 916 FSS-9 and VAFS were completely filled out, while one participant only conducted VAFS. Of 655 people included, 188 and 37 people filled out FSS-9 and VAFS twice and thrice, respectively. Of 37 people having three measurements, we included the second or third measurement that was closest to the date for one-year after the baseline date for creating a substantially similar time interval between measurements to each participant when calculating internal consistency and confirmatory factor analysis (mean: 12 months (Standard deviation: 4 months), see Additional file 1).

2.4. Measuring fatigue

We used the nine-item FSS and VAFS to measure the level of fatigue in the annual health assessment. The FSS measured the extent of fatigue the last week, including mental and physical functioning, motivation, carrying out duties, and interfering with work, family, or social life. VAFS measured the people's general experience of fatigue. VAFS is correlated to FSS-9 among various chronic diseases [15], including people with chronic hepatitis C virus infection [18]. VAFS is simpler to fill out compared to FSS-9 and may be used as a supplement to FSS-9 if high reliability and validity. A high score on items of FSS or VAFS indicated high levels of fatigue, respectively. The data collection software only allowed valid responses to each question and prompted empty questions before submission of the questionnaire to minimize missing data. People who only filled out FSS or VAFS were included. In the past, the FSS has been translated and back-translated from the US-English version into Norwegian by a qualified native Norwegian-speaking translator (Additional Table 2) [23].
Table 2
Descriptive information of single items in Fatigue Severity Scale and Visual Analog Fatigue Scale.

| Item                                                                 | Mean (SD)  |
|----------------------------------------------------------------------|------------|
| FSS                                                                  |            |
| I1: My motivation is lower when I am fatigued                       | 5.38 (2.04)|
| I2: Exercise brings on my fatigue                                    | 4.73 (2.12)|
| I3: I am easily fatigued                                            | 4.54 (2.15)|
| I4: Fatigue interferes with my physical functioning                 | 4.95 (2.07)|
| I5: Fatigue causes frequent problems for me                         | 4.43 (2.22)|
| I6: My fatigue prevents sustained physical functioning              | 4.62 (2.19)|
| I7: Fatigue interferes with carrying out certain duties and         | 5.00 (2.14)|
| responsibilities                                                      |            |
| I8: Fatigue is among my three most disabling symptoms                | 4.64 (2.34)|
| I9: Fatigue interferes with my work, family, or social life         | 4.85 (2.17)|
| VAFS                                                                 | 5.19 (2.65)|

Legends: Table 2 displays descriptive information in the nine-item Fatigue Severity Scale (FSS-9) and the Visual Analog Fatigue Scale (VAFS) at baseline. The FSS items were answered on a Likert scale (1 (no fatigue) – 7 (worst fatigue)), while VAFS were ranged from 0 (no fatigue) to 10 (worst fatigue). The mean results of FSS items showed some variations from 4.43 to 5.38. The mean value of VAFS showed 5.19. FSS-3 items are in bold.

2.5. Statistical analysis

We used Stata 16.0 for descriptive analysis (mean, standard deviation, and frequency analysis), and IBM SPSS version 24.0 and Mplus version 8.4 for reliability analysis (Cronbach’s $\alpha$ and $\alpha$ if item deleted), confirmatory factor analysis, and linear mixed model analysis (Mplus: TwoLevel analysis). The threshold for statistical significance was set to $P < 0.05$ for all analyses unless stated otherwise. The shortening and validation analyses of FSS and VAFS are conducted in three steps.

2.5.1. Step 1: Internal consistency of FSS-9 and a shortened FSS-9 version with added VAFS

We calculated internal consistency for all items in the FSS-9 and VAFS. A shortened version of FSS-9 was constructed by removing items from FSS-9 based on clinical judgment and internal consistency. The purpose of including clinical judgment was to emphasize simplicity and intelligibility and to remove difficult readable and abstract items that could be difficult to comprehend for people with SUD. Internal consistency was calculated by removing one by one item from the scale by using Cronbach’s $\alpha$. A very good internal consistency was defined if Cronbach’s $\alpha$ exceeds an accepted cut-off of 0.80 [24].

2.5.2. Step 2: Longitudinal confirmatory factor analysis for evaluating the fit of FSS-9 and the shortened FSS-9 version with added VAFS.

We used confirmatory factor analysis to test the structure of the items in FSS-9 to evaluate relationships between the items and their underlying latent factors. In addition, the relationships between FSS-9 factors and VAFS were evaluated regarding construct overlap. The confirmatory factor analysis models were based on FSS-9 and the shortened version of FSS-9 alone at baseline and second measurement. In addition, the FSS factor models were combined with the covariance with VAFS to explore how VAFS affected the model. Further, we conducted four longitudinal confirmatory factor analysis models for the shortened version of FSS-9 by using longitudinal data at baseline and second measurement. First, we estimated a free longitudinal confirmatory factor analysis with all unique parameter values. We then tested for constraints in the model by setting the factor loadings within each item equal at baseline and second measurement. Third, we constructed a model by setting the residuals at baseline and second measurement equal within each indicator. The last model constrained the intercept values in the same indicators at baseline and
second measurement. We used the Wald test to compare model restrictions. All confirmatory factor analysis models were evaluated with standard fit measures: $\chi^2$, degrees of freedom, $p$-values, Comparative fit index, Tucker Lewis Index, Root Mean Square Error of Approximation with 90% confidence interval, and the probability of close fit. A well-fitted model should be a statistically non-significant $\chi^2$, values of Comparative fit index and Tucker Lewis Index above 0.95, and Root Mean Square Error of Approximation, preferably below 0.05 (close fit) [25]. Root Mean Square Error of Approximation above 0.10 is considered as a poor fitted model. We used the modification index to explore model improvements if the goodness of fit measures indicated a poor fitted model. This is a chi-square difference if a constrained parameter is estimated. Thus, the model is improved at the cost of one degree of freedom. We analyzed all variables as continuous variables due to the relatively high number of categories (FSS items ranged from one to seven, VAFS ranged from zero to 10) in the ordinal variables.

2.5.3. Step 3: Linear mixed model analysis for evaluating changes in FSS-9 and a shortened version of FSS-9 with added VAFS

We used a linear mixed model analysis (Mplus multilevel modeling: TwoLevel) of all 917 measurement occasions to evaluate linear changes in FSS-9 and the shortened version of FSS-9’s sum scores, the nine separate FSS items, and VAFS. First, we estimated a full random intercept random slope model, giving mean and individual variance in level and change together with the relationship between level and change [26]. If we found the covariance between the intercept and the slope variance to be statistically non-significant, these were constrained, and the model re-estimated as a random intercept fixed slope model. We used the Mplus Maximum Likelihood Robust estimator, correcting standard errors for potential deviation from normality [27]. In addition, interclass correlations were estimated. We used full information maximum likelihood in order to use all available measurements. Full information maximum likelihood assumes missing at random [28].

2.6. Ethics approval and consent to participate

The study is reviewed and approved by the Regional Ethical Committee for Health Research (REC) West, Norway (reference number: 2017/51/REK Vest, dated 29.03.2017/20.04.2017). Each patient provided written informed consent prior to enrolling in the study.

3. Results

3.1. Participant characteristics at baseline

Seventy percent were male, and the mean age was 43 years (Table 1). Half had primary school as the highest level of education. Eighty-two percent received opioid agonist treatment, while 42 percent had injected drugs in the past 30 days. The mean values of FSS-9 items varied from 4.43 to 5.38 at baseline (Table 2). For VAFS, the mean value was 5.19 at baseline. Variables were slightly left-skewed (skewness ranged from -1.14 to -0.31) and tended to a flattened distribution (kurtosis ranged from -1.24 to -0.09).
Table 1
Descriptive characteristics in population at baseline presenting numbers (n) and percentages.

| Participant characteristics | N = 655 |
|-----------------------------|---------|
| Age (years), n (%)          |         |
| 18–29                       | 81 (12) |
| 30–39                       | 185 (28) |
| 40–49                       | 207 (32) |
| 50–59                       | 147 (22) |
| ≥ 60                        | 35 (5)  |
| Mean (SD)                   | 43 (11) |
| Gender, n (%)               |         |
| Male                        | 462 (71) |
| Female                      | 193 (29) |
| Highest education level, n (%) |       |
| Not finished primary school | 40 (6)  |
| Primary school (9 years)    | 286 (44) |
| High school (12 years)      | 260 (40) |
| ≤ Bachelor degree (≤ 15 years) | 57 (9)  |
| > Bachelor degree (> 15 years) | 12 (2)  |
| Receiving opioid agonist treatment (%) | 542 (83) |
| Of those (%)                |         |
| Methadone                   | 212 (39) |
| Buprenorphine*              | 323 (60) |
| Long-acting morphine sulfate| 7 (1)   |
| Living conditions past 30 days, n |         |
| Owned or rented housing     | 518 (79) |
| Temporary living (homeless shelter, with family or friends) | 156 (24) |
| Living on street            | 27 (4)  |
| Prison                      | 7 (1)   |
| Other                       | 27 (4)  |
| Ever injected drugs, n (%)  |         |
| Unknown                     | 573 (94) |
| Injected drugs past 30 days |         |
| Unknown                     | 257 (42) |
| * Buprenorphine includes buprenorphine, buprenorphine-naloxone, and buprenorphine depot. |
3.2.1 Step 1: Internal consistency of FSS-9 and the shortened version of FSS-9 with added VAFS

FSS-9's Cronbach's coefficient $\alpha$ was 0.94 at baseline and 0.92 at the second measurement. For the shortened version of FSS-9, retaining the items 5–7 from FSS-9 (FSS-3), the Cronbach's coefficient $\alpha$ was 0.87 at baseline and 0.85 at the second measurement, respectively. No items had Cronbach's coefficient $\alpha$ below 0.85. The internal consistency was not affected when we added VAFS to FSS-9 and FSS-3, respectively.

3.2.2 Step 2: Longitudinal confirmatory factor analysis for evaluating the fit of FSS-9 and FSS-3 with added VAFS

The results from the confirmatory factor analyses comparing the fit of items in FSS-9 and FSS-3 with added VAFS at baseline and second measurement are shown in Table 3. At baseline, the one-dimensional model with unique factor loadings, residuals, and intercept values resulted in a borderline fitted model, with Comparative Fit Index and Tucker Lewis Index below requested levels and the Root Mean Square Error of Approximation point estimate just at the level of a poor fit for FSS-9, respectively. The modification index showed that the estimation of the residual covariance between items two and three improved the model with the best result ($\Delta \chi^2 = 80.1$, degree of freedom (df) = 1, $P < .001$). The FSS-3 showed a well-fitted model. The FSS-9 and FSS-3 were highly correlated with $r = 0.95$, $P < 0.001$, giving 90% explained variance after estimating factor scores. VAFS, together with the FSS-9 and FSS-3, respectively, gave identical results. The correlations with VAFS were $r = 0.70$, $P < .001$ (FSS-9) and $r = 0.68$, $P < .001$ (FSS-3).

We revealed relatively similar results for the factor models at the second measurement occasion, both the FSS-9 and FSS-3 without and with the VAFS. The longitudinal confirmatory factor analysis model based on FSS-3 supported time-invariant equal factor loadings and equal residuals between the baseline and second measurement (Fig. 1). The correlation between the models at baseline and the second measurement was $r = 0.52$, $P < 0.001$. A small reduction in model fit was seen if intercept were constrained equal within each measured item between measurements, but this model was also well-fitted.
Table 3
Confirmatory factor analysis results at baseline and second measurement

|                  | $\chi^2$ | df  | $P$   | CFI  | TLI  | RMSEA | $\varepsilon_a$ | $\varepsilon_a$ 90%CI | $\varepsilon_a$ $p<.05$ |
|------------------|----------|-----|-------|------|------|-------|------------------|------------------------|------------------------|
| **Time 1 (baseline)** |          |     |       |      |      |       |                  |                        |                        |
| FSS-9            | 212.90   | 27  | <0.001| 0.93 | 0.90 | 0.10  | 0.09, 0.12      | 0.00                   |                        |
| FSS-9 ResCov     | 132.80   | 26  | <0.001| 0.96 | 0.94 | 0.08  | 0.07, 0.09      | 0.00                   |                        |
| FSS-9 ResCov + VAFS | 151.33  | 34  | <0.001| 0.96 | 0.95 | 0.07  | 0.06, 0.09      | 0.00                   |                        |
| FSS-3            | 0.48     | 2   | 0.788 | 1.00 | 1.00 | 0.00  | 0.00, 0.05      | 0.95                   |                        |
| FSS-3 + VAFS     | 4.42     | 4   | 0.352 | 1.00 | 1.00 | 0.01  | 0.00, 0.06      | 0.87                   |                        |
| **Time 2 (second measurement)** |          |     |       |      |      |       |                  |                        |                        |
| FSS-9            | 108.14   | 27  | <0.001| 0.90 | 0.86 | 0.12  | 0.10, 0.14      | 0.00                   |                        |
| FSS-9 ResCov     | 50.16    | 26  | 0.003 | 0.97 | 0.96 | 0.06  | 0.04, 0.10      | 0.18                   |                        |
| FSS-9 ResCov + VAFS | 65.44  | 34  | <0.001| 0.96 | 0.95 | 0.06  | 0.04, 0.09      | 0.15                   |                        |
| FSS-3            | 4.20     | 2   | 0.122 | 0.99 | 0.98 | 0.07  | 0.00, 0.17      | 0.27                   |                        |
| FSS-3 + VAFS     | 5.37     | 4   | 0.252 | 1.00 | 0.99 | 0.04  | 0.00, 0.11      | 0.50                   |                        |
| **Time 1–2**     |          |     |       |      |      |       |                  |                        |                        |

Legends: CFI: Comparative Fit Index; TLI: Tucker Lewis Index; VAFS: Visual Analog Fatigue Scale; FSS-9: Nine-item Fatigue Severity Scale; FSS-3: Three-item Fatigue Severity Scale; $\varepsilon_a$: Root Mean Square Error of Approximation (RMSEA); $\varepsilon_a$ $P<.05$: Probability of close fit (RMSEA = .05); ResCov: Residual covariance between item 2 and 3 estimated.

0 free factor loadings, intercepts, and residuals within and from baseline (time 1) to second measurement (time 2)

1 equal factor loadings from baseline (time 1) to second measurement (time 2): test of constraints: Wald = 0.77, $P = .682$

2 and equal residuals within each indicator from baseline (time 1) to second measurement (time 2): Wald test of constraints: 1.25, $P = .682$

3 and equal intercepts within each indicator from baseline (time 1) to second measurement (time 2): Wald test of constraints: 8.61, $P = .035$

The table displays the confirmatory factor analysis results presenting explained variance and model fit with different fatigue score versions including 9-item (FSS-9), 10-item (FSS-9 + VAFS), 3-item (FSS-3) and 4-item (FSS-3 + VAFS). Both are presented at baseline and a second measurement with a median interval of 11 months. In addition, longitudinal confirmatory factor analysis of FSS-3, including two measuring point (Time 1–2), baseline, and second measurement, was presented.
|        | $\chi^2$ | df | $P$  | CFI  | TLI  | RMSEA |
|--------|----------|----|------|------|------|-------|
| FSS-3 | 13.33    | 8  | 0.101| 0.99 | 0.99 | 0.03  |
| FSS-3  | 14.47    | 10 | 0.153| 1.00 | 0.99 | 0.03  |
| FSS-3  | 15.48    | 13 | 0.278| 1.00 | 1.00 | 0.02  |
| FSS-3  | 23.35    | 16 | 0.105| 0.99 | 0.99 | 0.03  |

Legends: CFI: Comparative Fit Index; TLI: Tucker Lewis Index; VAFS: Visual Analog Fatigue Scale; FSS-9: Nine-item Fatigue Severity Scale; FSS-3: Three-item Fatigue Severity Scale; $\varepsilon_a$: Root Mean Square Error of Approximation (RMSEA); $\varepsilon_a P < .05$: Probability of close fit (RMSEA = .05); ResCov: Residual covariance between item 2 and 3 estimated.

0 free factor loadings, intercepts, and residuals within and from baseline (time 1) to second measurement (time 2)

1 equal factor loadings from baseline (time 1) to second measurement (time 2): test of constraints: Wald = 0.77, $P = .682$

2 and equal residuals within each indicator from baseline (time 1) to second measurement (time 2): Wald test of constraints: 1.25, $P = .682$

3 and equal intercepts within each indicator from baseline (time 1) to second measurement (time 2): Wald test of constraints: 8.61, $P = .035$

The table displays the confirmatory factor analysis results presenting explained variance and model fit with different fatigue score versions including 9-item (FSS-9), 10-item (FSS-9 + VAFS), 3-item (FSS-3) and 4-item (FSS-3 + VAFS). Both are presented at baseline and a second measurement with a median interval of 11 months. In addition, longitudinal confirmatory factor analysis of FSS-3, including two measuring point (Time 1–2), baseline, and second measurement, was presented.

3.2.3 Step 3: Linear mixed model analysis for evaluating changes in FSS-9 and FSS-3 with added VAFS

The linear mixed model analysis showed considerable within individual clustering for the FSS-9 and FSS-3 with added VAFS, and separate items (Table 4). However, the intraclass correlation coefficient estimated variations from 0.18 to 0.52, which showed more change from baseline, and over the second, and third measurements in some variables than in others. The results showed the random slope and covariance between intercept and slope to be statistically non-significant for all models. This indicated an equal linear change between the measurement occasions at the individual level. These re-estimated linear random intercept and fixed slope models showed a small increase in the items one to four, while no similar change was found in the other items. A mean change was also found in the VAFS variable, but nor this variable showed individual differences in change.
Table 4
Linear mixed model analysis results of the level and linear change in Fatigue Severity scale (FSS-9, FSS-3, and single items) and Visual analog Fatigue Scale (VAFS).

| Fixed effects | Random effects |
|---------------|----------------|
|               | Baseline       | Change | Residual | Intercept | Slope | Cov_{i,s} | ICC |
|               | I   | P   | S   | P   | σ²_e | P   | σ²_i | P   | σ²_s | P   | σ²_{is} | P   |
| FSS -9 a      | 4.7 | *** | 0.0 | 1   | 1.4  | 6   | 1.6  | *** | 0.0  | -0.0 | 2     | .52 |
|               | 6   |      |     |     |      |     |      |      |      |      |       |      |
| FSS -3 b      | 4.6 | *** | 0.0 | 0   | 2.0  | 7   | 1.6  | *** | 0.0  | -0.0 | 1     | .44 |
|               | 4   |      |     |     |      |     |      |      |      |      |       |      |
| Item 1        | 5.3 | *** | 0.0 | 2   | 2.5  | 2   | 1.5  | *** | 0.0  | -0.0 | 4     | .38 |
|               | 5   |      |     |     |      |     |      |      |      |      |       |      |
| Item 2        | 4.7 | *** | 0.0 | 3   | 2.9  | 5   | 1.4  | *** | 0.0  | -0.0 | 3     | .32 |
|               | 1   |      |     |     |      |     |      |      |      |      |       |      |
| Item 3        | 4.5 | *** | 0.0 | 3   | 2.4  | 0   | 2.1  | *** | 0.0  | -0.0 | 2     | .47 |
|               | 1   |      |     |     |      |     |      |      |      |      |       |      |
| Item 4        | 4.9 | *** | 0.0 | 3   | 3.4  | 3   | 0.7  | *   | 0.0  | -0.0 | 4     | .18 |
|               | 0   |      |     |     |      |     |      |      |      |      |       |      |
| Item 5 sv     | 4.3 | *** | 0.0 | 2   | 2.9  | 9   | 1.8  | *** | 0.0  | -0.0 | 2     | .38 |
|               | 8   |      |     |     |      |     |      |      |      |      |       |      |
| Item 6 sv     | 4.5 | *** | 0.0 | 1   | 3.1  | 8   | 1.6  | *** | 0.0  | -0.0 | 1     | .34 |
|               | 7   |      |     |     |      |     |      |      |      |      |       |      |
| Item 7 sv     | 4.9 | *** | 0.0 | 0   | 2.9  | 2   | 1.6  | *** | 0.0  | -0.0 | 2     | .36 |
|               | 7   |      |     |     |      |     |      |      |      |      |       |      |
| Item 8        | 4.6 | *** | 0.0 | 2   | 2.9  | 2   | 2.5  | *** | 0.0  | -0.0 | 2     | .46 |
|               | 0   |      |     |     |      |     |      |      |      |      |       |      |
| Item 9        | 4.8 | *** | 0.0 | 1   | 2.9  | 0   | 1.9  | *** | 0.0  | -0.0 | 0     | .40 |
|               | 2   |      |     |     |      |     |      |      |      |      |       |      |
| VAFS           | 5.1 | *** | 0.0 | 2   | 4.7  | 3   | 2.2  | *** | 0.0  | -0.0 | 2     | .33 |
|               | 4   |      |     |     |      |     |      |      |      |      |       |      |

Cov_{i,s}, Covariance between intercept and slope; ICC: Intraclass Correlation Coefficient of between on total variance; I: Intercept (baseline); S: Slope (change); σ²: variance (residual, intercept and slope); σ²_{is}: Covariance between intercept and slope

Legends: * P < .05 ** P < .01 *** P < .001.

a Nine-item Fatigue Severity Scale (FSS-9);
b Three-item FSS (FSS-3);
sv Items included in the short FSS-3;

The table displays a linear mixed model analysis of FSS-3, FSS-9, VAFS, and single items, including fixed and random effects and ICC among people with substance use disorder. The random-effects present the slope and covariance between intercept and slope level and change.

4. Discussion
This study has shown how a three-step model consisting of internal consistency, longitudinal confirmatory factor analysis, and mixed model analysis may be useful for the shortening of validated questionnaires while retaining most of the included variance and validity when having longitudinal data. Questionnaires that easily comprehend with few items are essential to ensure high completion rates among people with SUDs. Our findings showed that the full-scaled FSS-9 had excellent internal reliability when tested empirically by using internal consistency. However, a three-item FSS did not reduce substantially the internal consistency compared to the FSS-9 with or without VAFS. The results from the longitudinal confirmatory factor analysis of FSS-3 showed a well-fitted unidimensional model with equal factor loadings and equal residuals between baseline and second measurement. The FSS-9 and FSS-3 were almost perfectly correlated. The mixed model analysis showed that even though the level of fatigue varied, the scoring structure was substantially stable and equal between the measurements.

The reliability analyses showed high internal consistency of FSS-9 and FSS-3, and the internal consistency dropped minimally from 0.94 to 0.87 when reducing the number of items from ten (FSS-9 + VAFS) to three (FSS-3). A homogenous and substantial equal internal consistency was also found in studies evaluating FSS-9 in other chronic diseases such as stroke [15], hepatitis C virus infection [13], and multiple sclerosis [29]. A shortened FSS on seven items has also been validated on patients with systemic lupus erythematosus, HIV, or stroke to have minimal change on reliability compared to the FSS-9 [19, 21, 30]. Unlike other populations, people with SUDs may have a broad specter of mental and physical diseases that could interfere with the people's experience of fatigue from measurement to another [31, 32]. The majority of people included in this study were marginalized, having more than one SUD, and more than 40% had injected drugs in the past 30 days. Consequently, people could be affected by centrally acting stimulants or sedatives taken immediately before the questionnaires were filled out, or some people could have substantial changes in their social conditions such as needing to move from housing to living on the street or changes in their income, which may impact the way they responded to fatigue. These psychosocial and medical conditions made the FSS-3 appropriate for ensuring credible results of items that were well-comprehended for people and considered whether the levels of functioning or self-perceived sense of fatigue were affected.

The confirmatory factor analyses demonstrated a unidimensional model at baseline and the second measurement for FSS-9 that was improved by adding residual covariance between items two and three. This pointed towards that the responses in these two items not adequately were captured by the one-factor model. The explanations could be that the Norwegian version had substantially similar wording of items two and three (Supplementary table 2), as well as the order of items in the questionnaire that affected people's perception of them. Unidimensional factor models are frequently reported when validated FSS on other populations, including the general populations [16, 23, 30]. However, previous studies have noted that multidimensional models in confirmatory factor analysis could be missed due to small study populations [16, 30]. Our findings on a large population on people with SUDs confirmed that one single factor for analyzing FSS-3 matched the observations adequately. Nevertheless, residual covariance between items two and three was of importance for maintaining a well-fitted unidimensional model in FSS-9.

The linear mixed model analysis showed that four of the items in FSS-9 and the VAFS changed differently between measurements compared with other items. This gives further arguments for better validity of the shortened version, which is assumingly less sensitive for fluctuations than the remaining items in FSS-9.

Nine-item FSS may be shortened to FSS-3 without VAFS when evaluating fatigue among people with SUDs. However, FSS-3 may increase the risk of common method bias compared to FSS-9 if the responses to items in a shorter questionnaire are more likely to be accessible in short-term memory and recalled when responding to other items [33]. Previous studies have validated a shortened FSS-9 on seven items across people with multiple sclerosis, stroke, and Human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS) showing high reliability, and however: present cross-sample differences in items three, five, six, and nine [19, 20]. Nevertheless, the items may have varying importance in different populations. Therefore, the construct validity of the FSS-3 may be limited to people with SUDs pending on validation studies that confirmed similar results on other populations.

**Strengths and limitations**

This study has several strengths. People who were included from the cohort are hard to reach for both research and health care. The majority of participants had an extensive history of injecting drug use with several drug addictions, and a substantial proportion had an ongoing injecting drug use. We plotted all collected information into the software directly for reducing the risk of
data entry errors and ensuring nearly complete data for each measuring point. The data within each interview were also mostly complete with a few missing items.

The study has some limitations. People recruited to this study filled out the FSS and VAFS questionnaires in different mental and physical health conditions. Changes in substance use, medical and psychosocial challenges may impact on significant variations within and between people. Further, many participants were recruited from people with opioid dependence who received opioid agonist treatment while others, to a more substantial degree, utilized street drugs. In addition, there was some variation in time between the baseline and the second and third measurements. To adjust for this, those who fulfilled FSS and VAFS thrice, only one of the two last measurements that were closest to date one year after the baseline date, were included in the reliability analysis and confirmatory factor analyses. Further, common method bias could affect the shortening from nine to three items in FSS. A shorter questionnaire is more likely to be accessible in short-term memory and recalled when responding to other items [33].

5. Conclusion

This three-step procedure, including internal consistency, longitudinal confirmatory factor analysis, and linear mixed model analysis, might be useful for shortening validated questionnaires while retaining the included variance and validity in settings with repeated measurements. We found that a three-item questionnaire for fatigue (FSS-3) was more consistent in the structure of changes in fatigue levels compared to a ten-item version (FSS-9 + VAFS) among people with SUDs. This method can be essential in the validation and shortening of homogenous questionnaires for use on people with SUD and other populations.

Abbreviations

AIDS
acquired immune deficiency syndrome
FSS
Fatigue Severity Scale
FSS-9
nine-item Fatigue Severity Scale
FSS-3
three-item Fatigue Severity Scale
HIV
Human Immunodeficiency virus
SUD
Substance use disorder
VAFS
Visual analog fatigue scale.

Declarations

Ethics approval and consent to participate

The study is reviewed and approved by the Regional Ethical Committee for Health Research (REC) West, Norway (reference number: 2017/51/REK Vest, dated 29.03.2017/20.04.2017). Each patient provided written informed consent prior to enrolling in the study.

Consent for publication

Participants have consented to publication

Availability of data and material

No additional data are available due to data protection requirements.

Competing interests
Funding

This work was supported by The Norwegian Research Council (BEHANDLING, contract no 269855); and the Western Norway Regional Health Authority («Åpen prosjektstøtte») with Department of Addiction Medicine, Haukeland University Hospital, Bergen, Norway as responsible institution. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Two of the authors are funded from the research grant BEHANDLING related to the project INTRO-HCV from the Norwegian Research Council. Their respective affiliations fund the other authors.

Authors’ contributions

Jørn Henrik Vold and Rolf Gjestad have led the study design, analysis, and writing the article preparation. Christer Aas, Eivind Meland, Kjell Arne Johansson, and Lars Thore Fadnes have contributed to the study design, analysis, and writing the article preparation. All authors have read and approved the final article.

Acknowledgements

We thank the devoted clinical staff for their enthusiasm during the planning of the project. We also thank Nina Elisabeth Eltvik and Christer Kleppe for valuable help and input during the planning and preparation phases.

Authors’ information

Jørn Henrik Vold, MD, Department of Addiction Medicine, Haukeland University Hospital, and Department of Public Health and Primary Care, University of Bergen. Mailing address: Department of Addiction Medicine, Haukeland University Hospital, Jonas Lies vei 65, N-5021 Bergen, Norway. E-mail: jorn.vold@uib.no.

References

1. World Drug Report 2019, Executive Summary. https://wdr.unodc.org/wdr2019/prelaunch/WDR19_Booklet_1_EXECUTIVE_SUMMARY.pdf: United Nations Office on Drugs and Crime 2019.

2. Health and social responses to drug problems - A European Guide. http://www.emcdda.europa.eu/system/files/publications/6343/TIPUBPDF_TD0117699ENN_PDFWEB_20171009153649.pdf: European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) 2017.

3. Desrosiers A, Blokhina E, Krupitsky E, et al. Psychiatric symptoms, quality of life, and HIV status among people using opioids in Saint Petersburg, Russia. Drug Alcohol Depend. 2017;172:60–5.

4. Levola J, Aalto M, Holopainen A, et al. Health-related quality of life in alcohol dependence: a systematic literature review with a specific focus on the role of depression and other psychopathology. Nord J Psychiatry. 2014;68(6):369–84.

5. Brorson HH, Ajo Arnevik E, Rand-Hendriksen K, et al. Drop-out from addiction treatment: a systematic review of risk factors. Clin Psychol Rev. 2013;33(8):1010–24.

6. Sarkar S, Jiang Z, Evon DM, et al. Fatigue before, during and after antiviral therapy of chronic hepatitis C: results from the Virahep-C study. Journal of hepatology. 2012;57(5):946–52.

7. Rottoli M, La Gioia S, Frigeni B, et al. Pathophysiology, assessment and management of multiple sclerosis fatigue: an update. Expert Rev Neurother. 2017;17(4):373–9.

8. Amaud L, Gavand PE, Voll R, et al. Predictors of fatigue and severe fatigue in a large international cohort of patients with systemic lupus erythematosus and a systematic review of the literature. Rheumatology. 2019;58(6):987–96.
9. Siciliano M, Trojano L, Santangelo G, et al. Fatigue in Parkinson's disease: A systematic review and meta-analysis. Movement disorders: official journal of the Movement Disorder Society. 2018;33(11):1712–23.

10. Bower JE. Cancer-related fatigue—mechanisms, risk factors, and treatments. Nature reviews Clinical oncology. 2014;11(10):597–609.

11. Krupp LB, LaRocca NG, Muir-Nash J, et al. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol. 1989;46(10):1121–3.

12. Lerdal A. Fatigue Severity Scale. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer Netherlands; 2014. pp. 2218–21.

13. Rosa K, Fu M, Gilles L, et al. Validation of the Fatigue Severity Scale in chronic hepatitis C. Health Quality of Life Outcomes. 2014;12:90.

14. Siciliano M, Chiorri C, De Micco R, et al. Fatigue in Parkinson's disease: Italian validation of the Parkinson Fatigue Scale and the Fatigue Severity Scale using a Rasch analysis approach. Parkinsonism Relat Disord. 2019;65:105–10.

15. Ozyemisci-Taskiran O, Batur EB, Yuksel S, et al. Validity and reliability of fatigue severity scale in stroke. Top. 2019;26(2):122–7.

16. Alekseeva TM, Gavrilov YV, Kreis OA, et al. Fatigue in patients with myasthenia gravis. J Neurol. 2018;265(10):2312–21.

17. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. Res Nurs Health. 1990;13(4):227–36.

18. Kleinman L, Zodet MW, Hakim Z, et al. Psychometric evaluation of the fatigue severity scale for use in chronic hepatitis C. Qual Life Res. 2000;9(5):499–508.

19. Lerdal A, Kottorp A, Gay C, et al. A 7-item version of the fatigue severity scale has better psychometric properties among HIV-infected adults: an application of a Rasch model. Quality of life research: an international journal of quality of life aspects of treatment care rehabilitation. 2011;20(9):1447–56.

20. Johansson S, Kottorp A, Lee KA, et al. Can the Fatigue Severity Scale 7-item version be used across different patient populations as a generic fatigue measure—a comparative study using a Rasch model approach. Health Qual Life Outcomes. 2014;12:24.

21. Lerdal A, Kottorp A. Psychometric properties of the Fatigue Severity Scale-Rasch analyses of individual responses in a Norwegian stroke cohort. Int J Nurs Stud. 2011;48(10):1258–65.

22. Fadnes LT, Aas CF, Vold JH, et al. Integrated treatment of hepatitis C virus infection among people who inject drugs: study protocol for a randomised controlled trial (INTRO-HCV). BMC Infect Dis. 2019;19(1):943.

23. Lerdal A, Wahl A, Rustoen T, et al. Fatigue in the general population: a translation and test of the psychometric properties of the Norwegian version of the fatigue severity scale. Scandinavian journal of public health. 2005;33(2):123–30.

24. Pedhazur EJ, Schmelkin LP. Measurement, design, and analysis: an integrated approach. Hillsdale: Lawrence Erlbaum Associates.; 1991.

25. Kline RB. Principles and Practice of Structural Equation Modeling. 4 ed. New York: The Guilford Press; 2016.

26. Hox JJ, Van de Schoot MM R. Multilevel Analysis. Techniques and Applications. New York, Routledge. Taylor & Francis Group; 2018.

27. Wang J, Wang X. Structural Equation Modeling: Applications Using Mplus. West Sussex, UK, Wiley, A John Wiley & Sons, Ltd., Publication; 2012.

28. Enders CK. Applied Missing Data Analysis. New York: The Guilford Press; 2010.

29. Mills RJ, Young CA, Nicholas RS, et al. Rasch analysis of the Fatigue Severity Scale in multiple sclerosis. Mult Scler. 2009;15(1):81–7.

30. Feng C, He Q, Wu Y, et al. Psychometric properties of fatigue severity scale in Chinese systemic lupus erythematosus patients. Health and quality of life outcomes 2019;17 (1) (no pagination)(71).

31. van Emmerik-van Oortmerssen K, van de Glind G, van den Brink W, et al. Prevalence of attention-deficit hyperactivity disorder in substance use disorder patients: A meta-analysis and meta-regression analysis. Drug Alcohol Depend. 2012;122(1–2):11–9.
32. van Emmerik-van Oortmerssen K, van de Glind G, Koeter MW, et al. Psychiatric comorbidity in treatment-seeking substance use disorder patients with and without attention deficit hyperactivity disorder: results of the IASP study. Addiction. 2014;109(2):262–72.

33. Podsakoff PM, MacKenzie SB, Lee JY, et al. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol. 2003;88(5):879–903.

**Figures**

![Figure 1](image)

**Figure 1**

Title: Longitudinal confirmatory factor analysis of Fatigue Severity Scale on people with substance use disorder Legends: Figure displays longitudinal factor analysis structure with upper circles presenting factor 1 and factor loadings between factor 1 and three separate question items at two different time points (baseline and second measurement) for Fatigue Severity Scale. The arrows at the bottom presents residuals, and alfa (α) the standardized mean values of each question item. The arrow between the circles presents the correlation (0.52) between the confirmatory factor analyses. The values 1.86 and 1.78 on the left side of the circles, respectively, present the variance of the latent variables.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- STROBEstatement.docx
- Additionalfile2.docx
- Additionalfile1.docx