



Development and validation of a speech pathology-specific questionnaire for persons with multiple sclerosis (SMS)

Sarah El-Wahsh, Robert Heard & Hans Bogaardt

To cite this article: Sarah El-Wahsh, Robert Heard & Hans Bogaardt (2018): Development and validation of a speech pathology-specific questionnaire for persons with multiple sclerosis (SMS), International Journal of Speech-Language Pathology

To link to this article: <https://doi.org/10.1080/17549507.2018.1499802>

 View supplementary material 

 Published online: 22 Oct 2018.

 Submit your article to this journal 

 View Crossmark data 

Development and validation of a speech pathology-specific questionnaire for persons with multiple sclerosis (SMS)

SARAH EL-WAHSH , ROBERT HEARD  & HANS BOGAARDT 

Department of Health Sciences, The University of Sydney, Lidcombe, New South Wales, Australia

Abstract

Purpose: The aim of this study was to develop and validate the *Speech pathology-specific questionnaire for persons with Multiple Sclerosis (SMS)*.

Method: Forty-one items were generated through a literature review. Items were submitted to a preliminary psychometric validation process consisting of principal component analysis, internal consistency, test-retest reliability, and floor and ceiling effects using data from 164 participants. Criterion validity was assessed by comparing the SMS with the 12-item Short Form Health Survey (SF-12). Participants were recruited internationally through online channels and questionnaires were completed online.

Result: The SMS contains 16 items describing three components: speech and voice, language, and swallowing. Internal consistency (Cronbach's alpha) of the three components was satisfactory ($\alpha = 0.89$ – $\alpha = 0.91$). Criterion validity was evaluated using Spearman's rank correlation coefficient (ρ). A statistically significant weak to moderate correlation between the SMS and the SF-12 was identified ($\rho = -0.004$ – $\rho = -0.359$). No floor or ceiling effects were present. The SMS demonstrated strong test-retest reliability. All items had an intra-class correlation coefficient ≥ 0.70 .

Conclusion: The SMS is a psychometrically robust patient-reported outcome measure to assess speech-language pathology symptoms in persons with MS.

Keywords: *patient-reported outcome measures; multiple sclerosis; surveys and questionnaires*

Introduction

Multiple Sclerosis (MS) is an autoimmune neurological disease in which inflammation and demyelination occur in the white matter of the central nervous system (CNS) (Tullman, 2013). The pathophysiology of MS is a complex process in which the body's immune system attacks the myelin sheaths surrounding nerve fibers, resulting in impaired transmission of neural signals in the CNS. The aetiology of MS is not completely understood, although several factors have been implicated in the disease process. These include genetic susceptibility and environmental triggers (Tullman, 2013). MS is associated with a myriad of clinical manifestations involving motor, sensory, and cognitive deficits, such as fatigue, pain, visual problems, bladder and bowel dysfunction, sexual dysfunction, depression, gait disorders, and swallowing and communication disturbances (Burks, Bigley, & Hill, 2009). MS affects ~30 per 100 000 people globally with an average age of onset of 20–40 years, and has an

established female preponderance with a female to male ratio of 3:1 (Cheak, 2011; Tullman, 2013). The incidence of MS is increasing; however, the reasons for this are not clear. Possible contributors include increased awareness of the disease, improved access to medical care, and enhanced diagnostic measures (Benito-León, 2011). At present, the primary foci in the assessment and treatment of persons with MS (PwMS) are medical and physical symptoms and pharmacological treatments (Burks et al., 2009). As such, health professionals tend to overlook speech-language pathology symptoms in PwMS (De Pauw, Dejaeger, D'hooghe, & Carton, 2002).

Speech-language pathology symptoms in MS

The prominent speech-language pathology functions affected in PwMS include speech, language, voice, and swallowing (Bauer et al., 2013; De Pauw et al., 2002; Guan, Wang, Huang, & Meng, 2015;

Hartelius, Runmarker, & Andersen, 2000; Konstantopoulos, Vikelis, Seikel, & Mitsikostas, 2010; Laakso, Brunnegard, Hartelius, & Ahlsen, 2000; Piacentini et al., 2014). Swallowing is a complex motoric function that involves the orchestration of respiratory, oropharyngeal, and smooth muscles of the oesophagus (Cheak, 2011). Dysphagia refers to the disturbance in these complex sensorimotor functions of swallowing and is frequently observed in PwMS. A recent systematic review found that the prevalence of dysphagia in MS is 36% ($n=4349$) (Guan et al., 2015). Similarly, speech movements require complex synchronisation between the speed, range, strength, timing, and accuracy of respiratory, phonatory, articulatory, and prosodic processes. MS can affect the muscular control of these motor-speech systems, causing dysarthria and voice problems (Bauer et al., 2013). There is also increasing evidence to suggest that PwMS experience impairments in language functions. The link between language difficulties and MS is an area of debate. The subcortical neuropathology affecting white matter in MS has led some to believe language should not be affected in this disease due to its presumed cortical control (Mackenzie & Green, 2009). However, several models on the possible role of subcortical structures in language processing have been described in the literature (Laakso et al., 2000; Nadeau & Crosson, 1997). In several studies, PwMS exhibited difficulties in semantic, vocabulary, and naming tasks in comparison to controls (Lethlean & Murdoch, 1994; Mackenzie & Green, 2009). However, limited test sensitivity and visual, oculomotor, and phonatory impairments have been proposed as an alternate explanation for the manifestation of language difficulties in PwMS (Mackenzie & Green, 2009).

Health-related quality of life (HRQoL)

Disturbances in speech-language pathology functions have been found to affect the HRQoL of PwMS (Cheak, 2011; Mackenzie & Green, 2009; Piacentini et al., 2014). HRQoL can be defined as a patient's self-perceived health status, including physical, mental, and social wellbeing and functioning, as affected by a disease (Karimi & Brazier, 2016). PwMS report reduced speech intelligibility, decreased speech rate, altered voice quality, impaired verbal fluency, coughing and choking prandial, and altered eating habits as common speech-language pathology symptoms affecting their HRQoL (De Pauw et al., 2002; Mackenzie & Green, 2009; Piacentini et al., 2014). These symptoms are associated with negative physical and psychosocial consequences, including problems with communication, frustration, low self-esteem, and restricted participation in everyday activities and the workforce (Klugman & Ross, 2002).

Patient-reported outcome measures (PROMs)

PROMs are increasingly used by allied health professionals to evaluate HRQoL and to guide evidence-based practice. PROMs are questionnaires completed by the patient without involvement of a third-party judgment or interpretation and pertain to the patient's health condition. Growing interest in these measures can be attributed to the valuable information a health professional can extract from the data. They provide clinicians with a structured avenue to assess disease impact from the patient's perspective. The use of this patient driven information can guide personalised healthcare management (Sansoni, 2016). These measures can be administered prior to, during, and after an episode of care in clinical trials and rehabilitation settings to aid the evaluation of treatment effectiveness and to identify whether the patient feels the benefit of intervention. This information can in turn ratify the appropriateness of treatment and whether revision of the intervention of choice is warranted (Piacentini et al., 2014). There is also growing recognition that PROMs should be integrated into routine assessment and measurement of health outcomes, as they allow for a more comprehensive indicator of patient need than instrumental diagnosis alone (Sansoni, 2016).

Broadly speaking, PROMs can be classified as generic or disease-specific instruments. Generic measures contain broad health-related constructs, such as bodily pain, emotional wellbeing, and physical functioning, and hence are applicable for a wide range of disease populations. Whilst beneficial for comparing disease populations, they are problematic in capturing the unique concerns of a specific disease. Furthermore, general health-related questions are not specific enough to detect change in domains targeted in speech-language pathology intervention. By contrast, disease-specific measures are sensitive enough to capture the specific effects of a particular disease on a patient's wellbeing and functioning. Subsequently, the identification of specific symptoms experienced by a patient can allow for better-tailored intervention (Sansoni, 2016).

Speech-language pathology PROMs for MS

Whilst well-established HRQoL questionnaires are readily available for a broad spectrum of patient types, speech-language pathology-related PROMs developed for PwMS are limited. PROMs used to assess speech and voice symptomology in PwMS are not validated for use with this population. These PROMs include the Quality of Life for the Dysarthric Speaker questionnaire (Piacentini et al., 2014) and the Voice Handicap Index (Jacobson et al., 1997). Similarly, commonly used PROMs to screen for dysphagia in PwMS are not MS-specific and include the Rome III questionnaire,

MD Anderson Dysphagia Inventory (MDADI), and the Northwestern dysphagia patient check sheet (Guan et al., 2015). For example, the MDADI is a PROM developed for patients with head and neck cancer, and therefore the application of this tool with PwMS may overlook issues of particular importance to this population. The use of questionnaires not validated for the population of interest can lead to measurement error and clinical relevance cannot be determined with total confidence (Dowrick, Wootten, Murphy, & Costello, 2015).

To our knowledge, there is one MS-specific speech-language pathology-related PROM, the DYsphagia in MUltiple Sclerosis (DYMUS) (Bergamaschi et al., 2008). In their initial article, Bergamaschi et al. (2008) set out to define and validate this tool and in conclusion reported a reliable and valid questionnaire. However, their methodology lacks rigour when compared to the quality criteria for evaluating HRQoL questionnaires developed by Terwee et al. (2007). The quality criteria outlined by Terwee et al. (2007) include content validity, internal consistency, test-retest reliability, construct validity, criterion validity, longitudinal validity, responsiveness, floor and ceiling effects, and interpretability. However, a brief comparison between the DYMUS and these measurement properties indicates that the DYMUS does not meet all these criteria. Clinicians should therefore exercise caution when interpreting outcomes based on the DYMUS due to concerns regarding its validity. Moreover, the DYMUS only assesses swallowing status. Dysphagia is, however, only one of the many speech-language pathology disturbances affecting PwMS.

Aims of the study

Speech-language pathology functions are impacted in PwMS and affect HRQoL, with these symptoms predominantly assessed using non-MS-specific PROMs. It is best-practice to use PROMs that have undergone rigorous psychometric evaluation for the population of interest to ensure that they are clinically useful (Terwee et al., 2007). There is a need therefore to develop a PROM that addresses a spectrum of speech-language pathology functions in PwMS, which can then be used to assess health status, evaluate treatment efficacy, and promote tailored interventions for PwMS. The aim of this study is to develop and validate such a tool using quality methods to establish a psychometrically robust PROM for PwMS.

Method

Ethical approval for this study was obtained from The University of Sydney Human Research Ethics Committee (Reference number 2017/197) and

complied with the Declaration of Helsinki ethical principles (World Medical Association, 2001).

Study design

The initial 41 items of the “Speech pathology-specific questionnaire for persons with Multiple Sclerosis” (SMS) were submitted to rigorous psychometric assessment consisting of principal component analysis (PCA), internal consistency, test-retest reliability, floor and ceiling effects, and criterion validity. Various guidelines are available that provide an outline of the statistical tests to be used when validating a HRQoL questionnaire. This study applied the criteria developed by Terwee et al. (2007), as it comprehensively outlines the measurement properties that should be present in a psychometrically robust PROM to ensure reliability and validity.

Questionnaire development

The initial items of the SMS were generated from a comprehensive literature review of speech-language pathology dimension-specific PROMs that address the core domains of speech-language pathology (speech, voice, swallowing, and language). PubMed and MEDLINE electronic databases were searched to identify existing questionnaires using the following terms: speech-language pathology, MS, voice, dysarthria, dysphagia, language, questionnaire, and survey. These terms were used in the advanced search builder to combine the search terms. The selected items that appeared in the initial SMS were taken from the following PROMs: 10 items were taken from the Voice Handicap Index-10 (VHI-10) (Rosen, Lee, Osborne, Zullo, & Murry, 2004), 13 items were taken from the Swallowing Quality of Life Questionnaire (SWAL-QOL) (McHorney et al., 2000), 10 items were taken from the Quality of Life in the Dysarthric Speaker (QOL-DyS) (Piacentini, Zuin, Cattaneo, & Schindler, 2011), and eight items were taken from the Multiple Ability Self-Report Questionnaire (MASQ) (Seidenberg, Haltiner, Taylor, Hermann, & Wyler, 1994).

Several features were considered during questionnaire selection. All the selected PROMs use a 5-point Likert response scale. In contrast to a dichotomous response method, this multi-item scale allows for increased tool sensitivity to capture differences in PwMS with milder symptomatic conditions. Accordingly, we did not use questions from the DYMUS, as this measure uses a dichotomous response method. The procedure of collapsing and recoding a Likert response scale into dichotomous variables has theoretical support. However, the inverse is not commonly practiced (Siebert & Siebert, 2017). Additionally, these PROMs address the specific symptoms experienced by PwMS as defined in the literature (Bauer et al., 2013;

De Pauw et al., 2002; Guan et al., 2015; Hartelius et al., 2000; Konstantopoulos et al., 2010; Laakso et al., 2000; Piacentini et al., 2014). Several features of the selected items were modified to ensure questionnaire ease of completion. In the items selected from the SWAL-QOL, the possessive determiner “you” was changed to the personal pronoun “I/me” to ensure the pronouns were consistent across all items. The Likert response scale for the SWAL-QOL items (first option = always and last option = never) was reversed (first option = never and last option = always) to match the response scale of the VHI-10, QOL-DyS, and MASQ to reduce respondent confusion and inaccurate responses. The wording of three items from the MASQ was changed. The original items were phrased in the affirmative (“I find it easy to...”), whilst the other selected items were phrased in the negative (“I find it difficult to...”). The use of both positively and negatively worded items in a single questionnaire can increase respondent confusion due to scale inversion (Van Sonderen, Sanderman, & Coyne, 2013). Item randomisation was used to remove existing constructs from the previous questionnaires to help overcome order bias. This was achieved using the Excel RAND function.

Criterion measure

The SMS scores were correlated against the 12-Item Short Form Health Survey (SF-12). The SF-12 is a generic multi-dimensional measure for assessing HRQoL. It is an abbreviated version of one of the most widely used measures to assess health status, the 36-Item Short Form Health Survey (SF-36). The SF-12 has been found to yield scores of considerable accuracy to the SF-36 with less respondent burden (Ware, Kosinski, & Keller, 1996). This measure has several features that make it a suitable choice to calibrate the SMS against. The SF-12 addresses health concepts of relevance to PwMS (Nortvedt, Riise, Myhr, & Nyland, 2000). It consists of 12 questions covering eight health dimensions: general health, physical functioning, role limitations due to physical health, vitality, role limitations due to emotional problems, emotional well-being, social functioning, and bodily pain (Ware et al., 1996). It also requires respondents to identify with a degree of wellbeing based on a 5-point Likert response scale. This multi-item scale is sensitive to capture differences in PwMS with milder symptoms, reducing ceiling effects (De Smedt, Clays, Annemans, & De Bacquer, 2014). The questionnaire uses a recall period covering the past four weeks. This is a favourable feature when studying the MS population due to the varying course of the disease (Tullman, 2013). The SMS should correlate with the SF-12, as they both measure compatible constructs of HRQoL, but not strongly, because one is a general measure and the other is a disease-specific measure.

Participants

Participants were recruited from different English-speaking countries using online channels. A recruitment announcement outlining the study’s aim, procedures, and the web-link to the questionnaire were e-mailed to MS support groups and speech-language pathology associations. These databases were identified through a Google search. The study was also advertised to MS societies and support groups on social media platforms. Recruited participants met the following criteria: (1) over 18 years old, (2) have MS, and (3) English-speaking. Participants whom did not meet these criteria were excluded from the study.

The sample size recommended to conduct PCA and validate a PROM is an area of debate. Various sources suggest that <100 participants is unlikely to produce a stable matrix solution (Kline, 2013; Terwee et al., 2007). In this study, 164 participants responded to the initial questionnaire, which is above the minimum recommended sample size to perform PCA. In addition, the empirically and theoretically grounded recommendations made by Guadagnoli and Velicer (1988) for determining the sample size required to produce a stable solution were also observed. This involved assessing component patterns with respect to the number of items defining a component and the magnitude of component loadings. There is no established criterion to determine the test-retest sample size of a psychometric validation study. As such, no formal sample size was calculated for the follow-up questionnaire. However, it is recognised that reliability is improved with increased numbers of participants.

Procedure

The online questionnaire was developed using SurveyMonkey, a cloud-based survey design platform. This software was selected for its user-friendly interface and reputable data security and privacy practices. Upon accessing the questionnaire, participants were asked to read the Participant Information Statement and provide informed consent to advance to the questions. Participants were able to withdraw at any point in time prior to submission of a completed questionnaire. The submission of a completed questionnaire was an indication of consent to participate. The questionnaire was comprised of three main sections. Participants completed questions concerning demographic (age, gender, country of residency, highest level of education, and employment status) and clinical information (type of MS, disease duration, medication management, and involvement with speech-language pathology services), followed by the SF-12 questions, and the 41 items of the initial SMS.

At the end of the questionnaire, participants were given the option to provide an e-mail address

to complete a follow-up questionnaire for test-retest reliability analysis. This questionnaire contained the 41 items of the initial SMS and 3 questions pertaining to demographic and clinical information (age, type of MS, and country of residency). Participants that left their e-mail address were sent a web-link to the follow-up questionnaire after a two-week timeframe. Terwee et al. (2007) acknowledge that this time period is long enough to reduce participant recall bias but short enough to minimise the possibility of notable clinical change. The follow-up e-mails were sent to participants individually to ensure confidentiality. Participants demographic and clinical information were used to permit matching of test and retest questionnaires.

Data analysis

Statistical analysis of the data was largely conducted using SPSS version 24.0 for Windows. The participants' responses to the SF-12 questions were analysed using the QualityMetric Health Outcomes™ Scoring Software version 5.0 (Saris-Baglana et al., 2010).

Item reduction procedure

PCA was applied as an extraction method followed by orthogonal rotation (varimax). The purpose of the current study was item reduction to produce a PROM, rather than identifying potential underlying causal processes. Tabachnick and Fidell (2001) recommend PCA rather than factor analysis in these circumstances.¹ The Kaiser-Meyer-Olkin (KMO) test for sampling adequacy and Bartlett's test of sphericity were performed to examine the suitability of the data for PCA. Three procedures were used to define the optimal number of components to retain. First, following extraction of principal components from a correlation matrix, all components with an eigenvalue greater than one were considered (Kaiser, 1960). The second procedure was inspection of a Cattell scree plot with the component numbers on the x-axis and the eigenvalues on the y-axis. The point where the slope of the curve leveled off was employed to indicate the number of components that should be retained (Cattell, 1966). The third procedure that was used to identify the appropriate number of components to extract was Velicer's statistically based minimum average partial (MAP) test (Fabrigar, Wegener, MacCallum, & Strahan, 1999). As defined by Comrey and Lee (2013), the rotated component loading cut-off was set at 0.55 and items that cross-loaded into another component with <0.2 difference were eliminated. Items of key clinical significance were considered for retention even if cross-loaded.

Internal consistency

Internal consistency was evaluated using Cronbach's alpha. An outcome between $\alpha = 0.70$ and $\alpha = 0.95$ was considered acceptable (Terwee et al., 2007).

Test-retest reliability

Test-retest reliability was assessed using intra-class correlation coefficients (ICC) with a two-way mixed single measures (absolute agreement) method. Test-retest correlations were done at the individual item-level rather than overall domain score. An ICC of ≥ 0.70 was considered acceptable (Terwee et al., 2007).

Floor and ceiling effects

A component was considered to exhibit a floor or ceiling effect if >15% of participants scored the minimum or the maximum possible score (Terwee et al., 2007). Items were coded as follows: 0 = never, 1 = almost never, 2 = sometimes, 3 = almost always, and 4 = always.

Criterion validity

Criterion validity was determined by calculating the correlation between the final items of the SMS and the SF-12 scores. Participants' responses to the SF-12 questions were entered into the QualityMetric Health Outcomes™ Scoring Software. This software generated two summative values: the physical component summary score and the mental component summary score. The SMS and SF-12 summary scores were formally tested for normality using the Kolmogorov-Smirnov test. As the SMS data was not normally distributed, nonparametric Spearman's rank correlation coefficient (ρ) was used as a measure of correlation. The strength of the correlation was indicated using the following guide: $\rho > 0.70$ is a strong correlation, $\rho = 0.30-0.70$ is a moderate correlation, and $\rho < 0.30$ is a weak correlation (Mukaka, 2012).

Result

One hundred and sixty four participants responded to the initial questionnaire. The characteristics of participants are detailed in Table I. As recommended by Guadagnoli and Velicer (1988), component loadings were evaluated to determine the suitability of this sample size for item reduction procedures. Guadagnoli and Velicer (1988) state that if a component consists of four or more items with loadings >0.60 or if a component consists of many items with loadings >0.40, the sample size used can be interpreted using item reduction procedures. The component loadings of this sample satisfy these conditions.

At the end of the questionnaire, participants were given the option to leave their e-mail address to

Table I. Participant characteristics.

Characteristic		Value
Sex, <i>n</i> (%) <i>n</i> = 164	Female	133 (81.1)
	Male	31 (18.9)
Age (years) <i>n</i> = 163	Mean	47.9
	Standard deviation	±13.9
	Range	21–76
	New Zealand	81 (49.4)
Country of residency, <i>n</i> (%) <i>n</i> = 164	Australia	39 (23.8)
	Ireland	17 (10.4)
	South Africa	5 (3.0)
	United States of America	5 (3.0)
	Canada	3 (1.8)
	United Kingdom	3 (1.8)
	Other	11 (6.7)
	Relapsing-remitting MS	101 (61.6)
	Secondary progressive MS	25 (15.2)
Type of MS, <i>n</i> (%) <i>n</i> = 164	Primary progressive MS	19 (11.6)
	Relapsing progressive MS	3 (1.8)
	Unsure	16 (9.8)
	Mean	12.9
	Standard deviation	±10.6
Disease duration, years <i>n</i> = 162	Range	0–52
	Yes	119 (72.6)
	No	45 (27.4)
Medication management, <i>n</i> (%) <i>n</i> = 164	Yes	1 (0.6)
	No	163 (99.4)
Receiving speech-language pathology services, <i>n</i> (%) <i>n</i> = 164	Primary education	1 (0.6)
	Secondary education	48 (29.3)
	Tertiary education	115 (70.1)
Education, <i>n</i> (%) <i>n</i> = 164	Full-time	42 (25.6)
	Part-time	15 (9.1)
	Retired	36 (22)
	Unemployed	28 (17.1)
	Other	43 (26.2)
Employment, <i>n</i> (%) <i>n</i> = 164		

complete a follow-up questionnaire. The 130 participants that left their e-mail address were sent the web-link to the follow-up questionnaire. Out of these participants, 88 responded to the follow-up questionnaire. Ten of these participants' demographic and clinical information (age, type of MS, and country of residency) on the follow-up questionnaire did not match with information from an initial questionnaire, and therefore were not considered during analysis. Hence, the sample size for test-retest reliability was 78 participants.

Item reduction procedure

The KMO measure of sampling adequacy and Bartlett's test of sphericity were applied to examine the suitability of the data for PCA. The KMO measure of sampling adequacy was 0.93 and the individual diagonal elements of the anti-image correlation matrix were >0.86, which are within the excellent range (Dziuban & Shirkey, 1974). Bartlett's test of sphericity was significant ($\chi^2(820) = 5757.80, p < 0.001$), confirming that there were significant relationships among the items (Dziuban & Shirkey, 1974). Given these indicators, PCA was considered appropriate for the dataset.

The 41 items of the initial SMS (see [Supplementary Table 1](#)) were entered into PCA. The optimal number of components to retain was determined by inspecting eigenvalues, examining

scree plots, and Velicer's MAP test. The number of components with an eigenvalue over one suggested the presence of seven underlying latent dimensions, accounting for 71.42% of the total variance. A Cattell scree plot revealed the last clear inflection occurred between components 3 and 4 followed by a flat line trend (Figure 1A), making a three- or four-component solution viable. However, Velicer's MAP test recommended a three-component solution. Components 4, 5, 6, and 7 were found to contribute minimally to the solution, accounting for 3.30%, 2.91%, 2.62%, and 2.47% of the cumulative variance, respectively. From a clinical perspective, items that loaded into component 4 pertained to salivation. Salivation is not a primary concern for PwMS, but rather is more observed in persons with Parkinson's disease (Friedman & Potulska, 2001). From this, a three-component solution was derived. After removal of components 4, 5, 6, and 7, 29 items remained, accounting for an unrotated cumulative variance of 60.13% and a rotated cumulative variance of 45.13%. The four items in component 4 that were removed were Q21, "I have thick saliva or phlegm," Q25, "I have to clear my throat," Q28, "I cough," Q35, "I have excess saliva or phlegm." The four items in component 5 that were removed were: Q1, "I run out of air when I talk," Q3, "I use the phone less often than I would like to," Q5, "I find it difficult to follow telephone conversations," and Q18, "People seem to be speaking too fast." The two items in component 6 that were removed were

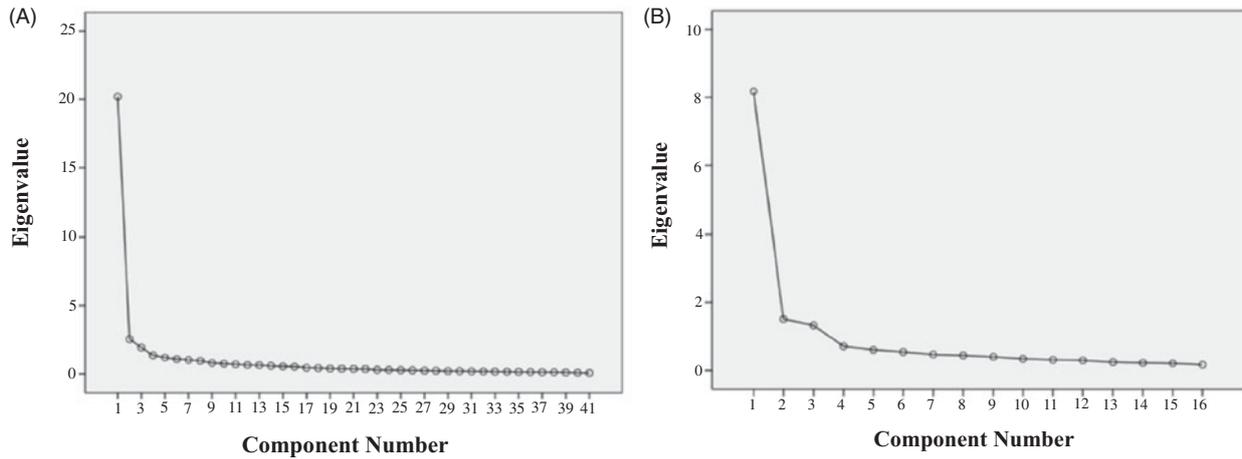


Figure 1. (A) Cattell scree plot for the initial 41 items of the SMS. (B) Cattell scree plot for the final 16 items of the SMS.

Q6, “I drool” and Q40, “Food or liquid dribble out of my mouth.” The two items in component 7 that were removed were Q32, “My speech has a nasal quality” and Q33, “Food or liquid come out of my nose.”

The minimum rotated component loading threshold was set at 0.55 (Comrey & Lee, 2013). Four items did not meet this cut-off, leaving 25 items. Q2, “I use a great deal of effort to speak” and Q36, “I have problems chewing” from component 1 were lost. Q22, “It is difficult for me to read and follow a newspaper story” from component 2 was lost and Q27, “I gag” from component 3 was lost. Three items had a high cross-loading that did not meet the 0.2 difference cross-loading cut-off point. Q26, “I have significant difficulty speaking when I am in a hurry” loaded most strongly into component 1 (0.563). However, this item produced a high cross-loading of 0.523 onto component 2, and hence was removed. Q11, “I find it difficult to make sense out of what people say to me” loaded most strongly onto component 2 (0.580). However, this item produced a high cross-loading of 0.486 onto component 5. This item is considered to be of clinical value as it is the only item that explores the receptive component of language in the questionnaire, and hence it was not eliminated. Q24 “I cough food or liquid out of my mouth when it gets stuck” loaded most strongly into component 3 (0.599). However, this item produced a high cross-loading of 0.481 onto component 4. This item is of clinical significance as coughing is a valuable self-report sign suggestive of potential penetration and/or aspiration of oral intake into the larynx, and hence it was not removed. Moreover, components 5 and 4, onto which Q11 and Q24 respectively cross-loaded, were removed from the solution, so the items did not cross-load with retained components.

Of the remaining 24 items, 15 items loaded into component 1, four items loaded into component 2, and five items loaded into component 3.

Component 1 of the rotated matrix was comprised of items pertaining to speech and voice functions. Component 2 was comprised of items pertaining to language functions. Component 3 was comprised of items pertaining to swallowing functions. The clinical interpretation of the three components led these to be labeled: speech and voice, language, and swallowing, respectively.

Internal consistency

Cronbach’s alpha was used to assess the internal consistency of the three components. The 15 items of component 1 resulted in an α of 0.96, which is over the accepted maximum alpha threshold of $\alpha = 0.95$ as defined by Terwee et al. (2007). The α value decreased to 0.91 with the removal of the seven items with the equal highest total correlation and individual α value of 0.96. The seven items that were removed were: Q7, “My voice makes it difficult for people to hear me,” Q9, “I’m tense when talking to others because of my voice,” Q13, “My speech is difficult for strangers to understand,” Q20, “The sound of my voice varies throughout the day,” Q37, “My speech sounds unnatural,” Q38, “My speech is slow or hesitant,” and Q39, “People have difficulty understanding me in a noisy room.” The four items of component 2 resulted in an α of 0.89. Removal of the item with the lowest total correlation and α increased the α value to 0.90. However, the item lost was, “I find it difficult to make sense out of what people say to me.” As previously mentioned, this item was not eliminated, as it is the only item that explores the receptive component of language in the questionnaire. This is not problematic, as an α of 0.89 is still considered acceptable according to the parameters defined by Terwee et al. (2007). The five items of component 3 resulted in an α of 0.89. The removal of any item did not increase the α value. Overall, examination of the internal consistency of each component led to the removal of seven items, reducing the SMS to 17 items.

Test-retest reliability

Test-retest reliability demonstrated strong correlations between the items on the test and retest questionnaires. Of the 17 items remaining after factor analysis and internal consistency measures, one item did not meet the set ICC cut-off of ≥ 0.70 . Q19, “My speech problem is so severe that it is difficult for my family to understand” was removed from component 1, reducing the total number of questionnaire items to 16. After the removal of this item, component 1 had an α of 0.91. This α value lies within the accepted parameters set by Terwee et al. (2007) and the removal of any item did not increase the α value.

Criterion validity

Spearman’s rank correlation coefficient identified a statistically significant weak to moderate correlation between the SF-12 physical component summary score and the three components of the SMS. The physical component score correlated most strongly with the swallowing component ($\rho = -0.359$, $\rho < 0.05$), followed by the speech and voice ($\rho = -0.238$, $\rho < 0.05$), and language ($\rho = -0.210$, $\rho < 0.05$) components. A statistically significant but weaker correlation was found between the SF-12 mental component summary score and the three components of the SMS. The mental component score correlated most strongly with the language component ($\rho = -0.245$, $\rho < 0.05$), followed by the swallowing ($\rho = -0.226$, $\rho < 0.05$), and speech and voice ($\rho = -0.004$, $\rho < 0.05$) components. All correlations were in the negative direction, that is, higher scores on the SF-12 (best health) were associated with lower scores on the SMS (absent/less severe speech-language pathology symptoms).

Final questionnaire with a three-component solution

The final SMS consists of 16 items with a three-component solution: speech and voice (component 1), which consists of seven items; language (component 2), which consists of four items; and swallowing (component 3), which consists of five items. Final component scores were calculated by taking the mean of the items in each component. Table II reports descriptive statistics.

PCA with varimax rotation was performed on the final set of 16 items to establish the percentage of variance explained uncontaminated by the removed items. The rotated cumulative variance explained by the final solution is 68.1%. Most items had a strong primary loading with small loadings on other components (see [Supplementary Table II](#)). However, whilst Q12, “My speech is slow” loaded most strongly into component 1 (0.570), it also produced a high cross-loading of 0.442 onto component 2. This item is associated with both component 1

Table II. Descriptive statistics summarising the three components of the SMS.

Component	Sum	Mean	Range	Standard deviation
1. Speech and voice	776	0.68	3.57	± 0.73
2. Language	912	1.39	4.00	± 0.90
3. Swallowing	641	0.78	3.80	± 0.78

(speech and voice) and component 2 (language), as speech motor proficiency can influence expressive language abilities. Moreover, there is clinical value in this item as dysarthria is one of the most well documented speech-language pathology symptoms affecting PwMS, including speaking rate (Hartelius et al., 2000). Accordingly, this item was not removed and was assigned to component 1 (speech and voice), as it is best fit from a clinical perspective.

A Cattell scree plot of the final 16 items validated a three-component solution with the last obvious inflection at component 3 (Figure 1B). Cronbach’s alpha for the final three components was 0.89. No floor or ceiling effects were observed in the final three components. A low percentage of participants achieved the highest or lowest score (0.6%). In all cases, only one participant achieved the highest score and one participant achieved the lowest score. This percentage is acceptable when compared with the 15% cut-off value set by Terwee et al. (2007).

Randomisation of the final 16 items of the SMS was completed to remove existing constructs and help overcome order bias (see Appendix I). This helps to ensure respondents provide the most accurate responses.

Discussion

This study presents the development and validation of the SMS, a new PROM to assess speech-language pathology symptoms in PwMS. Statistical tests were applied to determine the components of the questionnaire and to reduce the number of items. The final version of the SMS explores the following dimensions of speech-language pathology: speech, voice, language, and swallowing. The questionnaire showed strong levels of internal consistency and test-retest reliability. Floor and ceiling effects were negligible for all components. A weak to moderate correlation between the SMS and the SF-12 was identified. This finding does not cast doubt on the validity of the SMS, but quite the contrary. Scepticism towards new PROMs can arise when they strongly correlate with another measure, thus generating redundant data. On the other hand, this weaker correlation suggests the SMS provides different but complementary information to the SF-12. The final version of the SMS was shortened from 41 items to 16 items. This abbreviated tool has practical implications, including reduced respondent burden and ease of clinician interpretation.

It is important to consider the findings of this study in light of other studies. As the current study is the first to validate a PROM that assesses a spectrum of speech-language pathology symptoms in PwMS, direct comparisons with other similar studies cannot be drawn. However, it is possible to draw some comparisons between this study and the validation of the DYMUS questionnaire to assess dysphagia in PwMS (Bergamaschi et al., 2008). The demographic and clinical characteristics of this study's participants are similar to those from the DYMUS study: 81.1% of participants in this study were female versus 74% in the DYMUS study; 61.6% of participants had relapsing-remitting MS in this study versus 72% in the DYMUS study; the mean age of participants in this study was 47.9 years versus 40.5 years in the DYMUS study; the mean disease duration of participants in this study was 12.9 years versus 10.1 years in the DYMUS study. These descriptive statistics align with epidemiological data on MS (Tullman, 2013). The statistical methods applied to validate the DYMUS involved PCA as an extraction method followed by orthogonal rotation (varimax), Cattell's scree test, eigenvalues greater than one to determine the number of dimensions, Cronbach's alpha as a measure of reliability, and criterion validity by correlating the DYMUS with the Expanded Disability Status Scale. These methods were also employed in this validation study. However, this study employed additional measurement properties. These include test-retest reliability to examine if the SMS results are consistent over time and floor and ceiling effects to identify range constraints that may impede the ability to distinguish between milder and more severe symptoms. This study also employed statistical methods, such as Velicer's MAP test, to determine the number of components to retain. Therefore, it can be said that this study has performed a more robust psychometric analysis to validate this PROM.

Clinical implications

In recent years, a greater emphasis has been placed on the value and quality of life rather than the volume and quantity of life of patients in the healthcare system. With this growing emphasis on HRQoL, PROMs have gained greater prominence in clinical practice (Sansoni, 2016). To our knowledge, the DYMUS is the only speech-language pathology-related PROM validated for PwMS, which pertains to dysphagia (Bergamaschi et al., 2008). However, dysphagia symptoms are only one of the many speech-language pathology deficits impacting the HRQoL of PwMS. The SMS fills this gap, as it addresses several speech-language pathology symptoms commonly reported by PwMS (Klugman & Ross, 2002). This reliable and valid PROM will allow speech-language pathologists to provide input to the overall assessment and interdisciplinary

management of PwMS, addressing the less visible physical symptoms that may have an equal impact on HRQoL.

This study has important implications for research purposes, clinical practice, and service policy. The SMS can be used in clinical trials to evaluate treatment effectiveness from the patient's perspective and in longitudinal studies to survey the natural history of the disease. This measure can also be used in clinical practice to evaluate treatment effectiveness through direct contrast of questionnaire scores pre- and post- an episode of care, and can guide ongoing management strategies. The multi-dimensionality of the SMS is a benefit for this disease population. It can be argued that it may be unwarranted to assess PwMS across a spectrum of speech-language pathology functions when their presenting need is a single area. Nevertheless, in the early stages of this disease, symptoms can be subtle and thus completion of this questionnaire may reveal previously undetected difficulties. Early detection allows for more timely treatment options and improves intervention outcomes (De Pauw et al., 2002). The course of the disease typically changes over time and new clinical manifestations may arise (Tullman, 2013). Accordingly, there is a place for this tool to be used across the continuum of care to monitor patient stability and deterioration, helping identify the need for appropriate intervention at each stage of care—from prevention to palliation. Furthermore, the SMS can be used in service provision studies for MS to identify whether severely compromised patients receive appropriate rehabilitation services.

Limitations

A limitation of this study arises from the online Internet-based method used for data collection. This study involved a self-selection recruitment procedure whereby the questionnaire was released on the Internet through several online channels and persons who had Internet access visited the questionnaire web-link and voluntarily responded to the questionnaire. As such, we were not in control of the selection process. This sampling method did not give equal opportunity for individuals in the target population to be selected. The population of Internet users was difficult to define and the true selection probability from the population was unknown. Hence, there may be undercoverage, causing sampling bias (Greenacre, 2016). Moreover, the date of birth information of one participant yielded a negative age, and thus was not entered in descriptive analysis ($n = 163$). A similar occurrence was observed in the information of two participants regarding disease duration ($n = 162$). However, exclusion of this data should not significantly change the characteristics of the participants. Online questionnaire completion was the chosen

method of data collection, as it allowed for efficient dissemination of the questionnaire to a large number of potential participants from around the world. This is a strength of this study, as it involved the psychometric analysis of the tool across different English-speaking countries.

Another limitation that arises from using an Internet-based method for data collection is that it was unknown if all participants had a formal diagnosis of MS. However, the characteristics of the participants in this study sample reflect the well-established demographic features of PwMS (Tullman, 2013).

Future directions

Future studies should focus on confirming the psychometric properties and component structure of the SMS. This may be achieved by conducting confirmatory factor analysis in a new sample to validate the scale, as well as by using other methods such as Item Response Theory. Future studies should also evaluate correlations between the SMS and direct observation and instrumental analyses, such as modified barium swallow studies, acoustic voice analyses, and formal speech and language assessments. This information can be used to establish the sensitivity and specificity of the questionnaire to calculate a cut-off score for screening purposes. Subsequently, the SMS can be a useful tool for the preliminary selection of patients that should be referred for instrumental evaluation to provide a more comprehensive appraisal of their symptoms.

Conclusion

In conclusion, this study presents the SMS, a 16 item PROM that allows clinicians to efficiently assess a spectrum of speech-language pathology functions in persons with MS. Statistical analysis of the SMS revealed satisfactory reliability and validity measurement properties. The SMS should be integrated with instrumental diagnosis to enrich and supplement the clinical assessment of persons with MS. The SMS can be used in clinical trials, clinical practice, and service provision studies to assess health status, monitor treatment effectiveness, evaluate service accessibility, and facilitate patient-centered care.

Supplementary material

Supplemental data for this article can be accessed at <http://10.1080/17549507.2018.1499802>.

Notes

1 Hereafter the term “component” is the technically correct name for what is more generally referred to as “factor,” following a PCA.

ORCID

Sarah El-Wahsh  <http://orcid.org/0000-0002-3247-1590>

Robert Heard  <http://orcid.org/0000-0003-2089-0752>

Hans Bogaardt  <http://orcid.org/0000-0001-8413-533X>

References

- Bauer, V., Aleric, Z., Jancic, E., Knezevic, B., Prpic, D., & Kacavenda, A. (2013). Subjective and perceptual analysis of voice quality and relationship with neurological dysfunction in multiple sclerosis patients. *Clinical Neurology and Neurosurgery*, *115*, S17–S20. doi:10.1016/j.clineuro.2013.09.015
- Benito-León, J. (2011). Are the prevalence and incidence of multiple sclerosis changing? *Neuroepidemiology*, *36*, 148–149. doi:10.1159/000325368
- Bergamaschi, R., Crivelli, P., Rezzani, C., Patti, F., Solaro, C., Rossi, P., ... Cosi, V. (2008). The DYMUS questionnaire for the assessment of dysphagia in multiple sclerosis. *Journal of the Neurological Sciences*, *269*, 49–53. doi:10.1016/j.jns.2007.12.021
- Burks, J.S., Bigley, G.K., & Hill, H.H. (2009). Rehabilitation challenges in multiple sclerosis. *Annals of Indian Academy of Neurology*, *12*, 296. doi:10.4103/0972-2327.58273
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. doi:10.1207/s15327906mbr0102_10
- Cheak, E.M. (2011). Multiple sclerosis and dysphagia (Doctoral thesis), Southern Illinois University, Carbondale. Retrieved from http://opensiuc.lib.siu.edu/gs_rp
- Comrey, A.L., & Lee, H.B. (2013). *A first course in factor analysis*. Hillsdale, NJ: Psychology Press.
- De Pauw, A., Dejaeger, E., D’hooghe, B., & Carton, H. (2002). Dysphagia in multiple sclerosis. *Clinical Neurology and Neurosurgery*, *104*, 345–351. doi:10.1016/S0303-8467(02)00053-7
- De Smedt, D., Clays, E., Annemans, L., & De Bacquer, D. (2014). EQ-5D versus SF-12 in coronary patients: Are they interchangeable? *Value in Health*, *17*, 84–89. doi:10.1016/j.jval.2013.10.010
- Dowrick, A.S., Wootten, A.C., Murphy, D.G., & Costello, A.J. (2015). “We used a validated questionnaire”: What does this mean and is it an accurate statement in urologic research? *Urology*, *85*, 1304–1311. doi:10.1016/j.urology.2015.01.046
- Dziuban, C.D., & Shirkey, E.C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*, 358. doi:10.1037/h0036316
- Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272. doi:10.1037/1082-989X.4.3.272
- Friedman, A., & Potulska, A. (2001). Botulinum toxin for treatment of parkinsonian sialorrhea. *Neurologia i Neurochirurgia Polska*, *7*, 329–327. doi: 10.1016/s0022-510x(05)81959-8
- Greenacre, Z.A. (2016). The importance of selection bias in internet surveys. *Open Journal of Statistics*, *06*, 397. doi:10.4236/ojs.2016.63035
- Guadagnoli, E., & Velicer, W.F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin*, *103*, 265. doi:10.1037/0033-2909.103.2.265
- Guan, X.L., Wang, H., Huang, H.S., & Meng, L. (2015). Prevalence of dysphagia in multiple sclerosis: A systematic review and meta-analysis. *Neurological Sciences*, *36*, 671–681. doi:10.1007/s10072-015-2067-7

- Hartelius, L., Runmarker, B., & Andersen, O. (2000). Prevalence and characteristics of dysarthria in a multiple-sclerosis incidence cohort: Relation to neurological data. *Folia Phoniatrica Et Logopaedica*, 52, 160–177. doi:10.1159/000021531
- Jacobson, B.H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M.S., & Newman, C.W. (1997). The voice handicap index (VHI): development and validation. *American Journal of Speech-Language Pathology*, 6, 66–70. doi:10.1044/1058-0360.0603.66
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151. doi:10.1177/001316446002000116
- Karimi, M., & Brazier, J. (2016). Health, health-related quality of life, and quality of life: What is the difference? *Pharmacoeconomics*, 34, 645–649. doi:10.1007/s40273-016-0389-9
- Kline, P. (2013). *Handbook of psychological testing*. Abingdon, Oxon: Routledge.
- Klugman, T.M., & Ross, E. (2002). Perceptions of the impact of speech, language, swallowing, and hearing difficulties on quality of life of a group of South African persons with multiple sclerosis. *Folia Phoniatrica Et Logopaedica*, 54, 201–221. doi:10.1159/000063194
- Konstantopoulos, K., Vikelis, M., Seikel, J.A., & Mitsikostas, D.D. (2010). The existence of phonatory instability in multiple sclerosis: an acoustic and electroglottographic study. *Neurological Sciences*, 31, 259–268. doi:10.1007/s10072-009-0170-3
- Laakso, K., Brunnegard, K., Hartelius, L., & Ahlsen, E. (2000). Assessing high-level language in individuals with multiple sclerosis: A pilot study. *Clinical Linguistics & Phonetics*, 14, 329–349. doi: 10.1080/02699200050051065
- Lethlean, J.B., & Murdoch, B.E. (1994). Naming errors in multiple sclerosis: Support for a combined semantic/perceptual deficit. *Journal of Neurolinguistics*, 8, 207–223. doi:10.1016/0911-6044(94)90027-2
- Mackenzie, C., & Green, J. (2009). Cognitive–linguistic deficit and speech intelligibility in chronic progressive multiple sclerosis. *International Journal of Language and Communication Disorders*, 44, 401–420. doi:10.1080/13682820802697879
- McHorney, C.A., Bricker, D.E., Robbins, J., Kramer, A.E., Rosenbek, J.C., & Chignell, K.A. (2000). The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: II. Item reduction and preliminary scaling. *Dysphagia*, 15, 122–133. doi:10.1007/s004550010013
- Mukaka, M.M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24, 69–71. Retrieved from <http://www.mmj.mw/>
- Nadeau, S.E., & Crosson, B. (1997). Subcortical aphasia. *Brain and Language*, 58, 355–402. doi:10.1006/brln.1997.1707
- Nortvedt, M.W., Riise, T., Myhr, K.M., & Nyland, H.I. (2000). Performance of the SF-36, SF-12, and RAND-36 summary scales in a multiple sclerosis population. *Medical Care*, 38, 1022–1028. doi:10.1097/00005650-200010000-00006
- Piacentini, V., Mauri, I., Cattaneo, D., Gilardone, M., Montesano, A., & Schindler, A. (2014). Relationship between quality of life and dysarthria in patients with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation*, 95, 2047–2054. doi:10.1016/j.apmr.2014.04.023
- Piacentini, V., Zuin, A., Cattaneo, D., & Schindler, A. (2011). Reliability and validity of an instrument to measure quality of life in the dysarthric speaker. *Folia Phoniatrica Et Logopaedica*, 63, 289–295. doi:10.1159/000322800
- Rosen, C.A., Lee, A.S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the Voice Handicap Index–10. *The Laryngoscope*, 114, 1549–1556. doi:10.1097/00005537-200409000-00009
- Sansoni, J.E. (2016). Health outcomes: An overview from an Australian perspective. Australian Health Outcomes Collaboration, Australian Health Services Research Institute, University of Wollongong.
- Saris-Baglana, N., Dewey, C., Chisholm, G., Plumb, E., King, J., & Kosinski, M. (2010). *QualityMetric Health Outcomes™ Scoring Software, Version 5.0*. Lincoln, RI: QualityMetric Incorporated.
- Seidenberg, M., Haltiner, A., Taylor, M.A., Hermann, B.B., & Wyler, A. (1994). Development and validation of a multiple ability self-report questionnaire. *Journal of Clinical and Experimental Neuropsychology*, 16, 093–104. doi:10.1080/01688639408402620
- Siebert, C.F., & Siebert, D.C. (2017). *Data Analysis with Small Samples and Non-Normal Data: Nonparametrics and Other Strategies*. New York, NY: Oxford University Press.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate analysis*. California State University, Northridge: Harper Collins Publishers.
- Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., ... de Vet, H.C.W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60, 34–42. doi:10.1016/j.jclinepi.2006.03.012
- Tullman, M.J. (2013). Overview of the epidemiology, diagnosis, and disease progression associated with multiple sclerosis. *The American Journal of Managed Care*, 19, S15–S20. Retrieved from <http://www.ajmc.com/>
- Van Sonderen, E., Sanderman, R., & Coyne, J.C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS One*, 8, e68967. doi:10.1371/journal.pone.0068967
- Ware, J.E., Kosinski, M., & Keller, S.D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34, 220–233. doi:10.1097/00005650-199603000-00003
- World Medical Association. (2001). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, 79, 373. Retrieved from <http://www.who.int/bulletin/en/>