The Latent Structure of Medically Unexplained Symptoms and its Relation to Functional Somatic Syndromes

Michael Witthöft*  Wolfgang Hiller

Noelle Loch  Fabian Jasper

Johannes Gutenberg University of Mainz

In Press in
International Journal of Behavioral Medicine

*Correspondence concerning this article should be addressed to Michael Witthöft, Visiting Researcher at the Department of Psychological Medicine, Institute of Psychiatry, King’s College, London, UK; permanent address:
Michael Witthöft, Division of Clinical Psychology and Psychotherapy, Johannes Gutenberg University of Mainz, Wallstrasse 3, D-55122 Mainz, Germany. Tel +496131-3939202, E-mail: witthoef@uni-mainz.de.
Abstract

**Purpose:** Medically unexplained symptoms are the hallmark of somatoform disorders and functional somatic syndromes. Although medically unexplained symptoms represent a common phenomenon both in the general population as well as in medical settings, the exact latent structure of somatic symptoms remains largely unclear. **Methods:** We examined the latent structure of medically unexplained symptoms by means of the PHQ-15 questionnaire (i.e., a popular symptom checklist) and provide support for the construct validity of our model. The data was analyzed using confirmatory factor analysis in a general population sample (Study 1; \( N = 414 \)) and in a sample of primary care patients (Study 2; \( N = 308 \)). We compared four different latent structure models of medically unexplained symptoms: A general factor model, a correlated group factor model, a hierarchical model and a bifactor model. **Results:** In Study 1, a bifactor model with one general factor and four independent specific symptom factors (i.e., gastrointestinal, pain, fatigue, and cardio-pulmonary symptoms) showed the best model fit. This bifactor model was confirmed in the primary care sample (Study 2). Additionally, the model explained 59% of the variance of the Irritable Bowel Syndrome (IBS). In this structural equation model both the general factor (14%) as well as the gastrointestinal symptom factor (42%) significantly predicted the IBS. **Conclusions:** The findings of both studies help to clarify the latent structure of somatic symptoms in the PHQ-15. The bifactor model outperformed alternative models and demonstrated external validity in predicting IBS.

**Keywords:** medically unexplained symptoms (MUS), somatization, somatoform disorders, functional somatic syndromes, irritable bowel syndrome (IBS), confirmatory factor analysis, nested factor model, bifactor model
Introduction

Medically unexplained symptoms (MUS) are a frequent and mostly transient phenomenon in the general population. In a study by Hiller, Rief, and Brähler [1] in the general population, 81.6% of the participants reported at least one medically unexplained symptom associated with at least mild impairment during the last seven days, and 22.1% even reported one or more medically unexplained symptom associated with severe impairment. In contrast to transient and mild MUS as an everyday phenomenon, persistent, distressing, and severely impairing MUS represent the defining feature of somatoform disorders according to DSM-IV [2, 3]. The term “somatoform” is used to refer to bodily symptoms that at first seem to be indicative of organ pathology but for which no clear medical cause can be found. Among the mental disorders, somatoform disorders are the most prevalent category in primary care settings (e.g., [4, 5]) but still our knowledge regarding the etiology and pathogenesis of this condition appears rather limited (e.g., [6, 7]).

As outlined by Deary [8], one crucial prerequisite for explaining MUS and somatoform disorders is the establishment of a reasonable and empirically justified taxonomy of MUS. In statistical terms, such taxonomy implies knowledge about the latent structure of MUS. In this regard an adequate measurement model of MUS is needed as a starting point for further research on causes and consequences of MUS and related conditions such as somatoform disorders and functional somatic syndromes. Yet, such a generally accepted taxonomy, latent structure, or measurement model on which experts in MUS research and practice agree is still lacking. In primary care, screening questionnaires like, e.g., the Patient Health Questionnaire-15 (PHQ-15; [9] are used to identify people at risk for somatoform disorders by computing sum scores of MUS and comparing these sum scores to pre-defined cut-off scores. In psychometric terms, this procedure of computing sum scores of symptom lists implies the existence of a general
somatization factor, although the true latent structure of MUS might be more complex. As Mcnemar [10] pointed out, it is at least questionable to create a sum score of a questionnaire if the postulate of unidimensionality does not hold. In our case, two patients with a totally different pattern of MUS (e.g., many gastrointestinal symptoms versus many cardiological symptoms) might yield exactly the same sum score on such a questionnaire. This may be satisfying for screening purposes (i.e., for initially identifying patients suffering from MUS) but still means an unnecessary large loss of useful information when it comes to the question of specifying the exact diagnosis (i.e., the type of somatoform disorder or functional somatic syndrome) and planning an adequate treatment (i.e., treatment protocols may differ considerably according to the organ system or symptom type that is most important as e.g., in irritable bowel syndrome or chronic fatigue syndrome).

In general, the term *latent structure* does not only refer to the question whether diverse MUS might best be considered as many different phenomena or rather as one homogenous construct (e.g., [11]) but also to the question whether MUS represent a *continuous* or rather a *categorical* construct. As recently proposed by McGrath and Walters (in press), answering the second question about the *type* of latent structure (i.e., whether MUS represent a categorical or continuous construct) might actually represent a prerequisite for the development of a reasonable *model* of the latent structure. Interestingly, this question of the *type* of latent structure of MUS has largely been ignored in previous research, although a dimensional model of MUS has been theoretically assumed previously (e.g., [12, 13]). Two recent taxometric studies explicitly tested the type of latent structure of MUS and the results suggest that MUS indeed represent a dimensional rather than a categorical phenomenon [14, 15]. Following Meehl [16, 17] this finding suggests that a monocausal etiology of MUS, somatoform disorders, and functional
somatic syndromes appears rather unlikely\(^1\). Consequently, the modeling of the dimensional latent structure using factor analytic techniques appears well justified.

Previous studies using exploratory (EFA) and confirmatory factor analysis (CFA) have shown that MUS represent rather a multidimensional construct: In one of the first latent variable analyses of MUS in a sample of 686 family medicine patients, Robbins and colleagues [19] modeled symptoms of fibromyalgia (FMS), chronic fatigue syndrome (CFS), irritable bowel syndrome (IBS), somatic depression, and somatic anxiety and found evidence for a well fitting model with five correlated factors according to the described syndromes. The authors interpreted their findings as evidence for the existence of independent functional somatic syndromes. In a mixed sample of 315 healthy people and different kinds of psychosomatic patients using the DSM-III-R somatization symptoms, Deary [8] could replicate the findings of Robbins et al. [19] and reported five (medium to strongly correlated) symptoms factors labeled FMS, CFS, IBS, somatic depression, and somatic anxiety. Additionally, using second order factor analysis, Deary [8] highlighted the large degree of communality among the five symptom factors and postulated a hierarchical model of MUS consisting of a general symptom factor and five lower order symptom factors representing the functional somatic syndromes. Taylor et al. [20] replicated this five-factor model but reported an inferior model fit of the hierarchical model proposed by Deary [8]. In contrast, Nimnuan et al. [21] reported evidence of one strong general factor of MUS and a weaker second factor primarily determined by non-cardiac chest pain. However, only EFA but not CFA was used in this study. Using CFA, Liu et al. [13] found evidence for a strong general factor of MUS and two gender specific subfactors (i.e., conversion and menstrual symptoms in

\(^1\) Although this statement might appear trivial and researchers mostly agree on complex psychobiological models of MUS and related conditions, monocausal models (e.g., of virus infections) are still popular. In this regard, a study postulating that the Chronic Fatigue Syndrome is strongly linked to a certain virus [18] has been retracted and the findings are most likely attributable to contaminated laboratory reagents.
females; conversion and pain symptoms in males). In a more recent study using EFA, Fink et al. [12] reported evidence for three symptom specific factors (i.e., cardiopulmonary, musculoskeletal, and gastrointestinal symptoms) but the following cluster and latent class analysis demonstrated that these symptom patterns did not form specific classes of patients but rather belong to one latent trait. These findings of Fink et al. [12] appear all the more important since the question of symptom patterns and the identification of subgroups of patients is addressed separately whereas previous studies often have confounded these two issues.

There have also been various attempts to identify the factor structure of the PHQ-15 questionnaire. Kroenke et al. [9] used 13 of the PHQ-15 items (except menstrual pain, painful sexual intercourse) for their EFA and reported an oblique factor structure with three factors (cardiopulmonary, gastrointestinal, general pain/fatigue). Contrary, Lee et al. [22] analyzed all of the PHQ-15 items and reported an orthogonal four factor solution (cardiopulmonary, gastrointestinal, pain, neurological) that appeared in their EFA. Recently, Mewes et al. [23] conducted a CFA with 12 of the PHQ-15 items and postulated five lower order factors (cardiopulmonary, gastrointestinal, pain, neurological, fatigue/depression) and one higher order factor (general somatization). They reported CFI values of .92 or .91 depending on the particular sample. Thus, according to the guidelines by Hu and Bentler [24] the model fit was acceptable which is also the case for the EFAs we outlined above which hardly showed a so-called simple structure [25]. What the models have in common is the postulation of a gastrointestinal factor, a pain (plus fatigue) factor and a cardiopulmonary factor.

All in all, the exact latent structure of MUS remains unknown, i.e., it remains unclear how many latent factors represent MUS and if these factors are either correlated or orthogonal. Additionally, it remains unclear whether specific symptom patterns are informative to describe subgroups of patients suffering from functional somatic syndromes. Some of the difficulties to
identify a stable factor structure of MUS could be explained by a rather complex latent structure (i.e., a bifactor model). If there exists a general factor (e.g., general somatization) and several specific group factors (e.g., cardiopulmonary, pain, gastrointestinal) it is not surprising that EFAs show instable solutions with factor loading patterns which are far from the so-called simple structure.

Based on the finding that some studies found evidence of a general factor, and that in addition some studies also found evidence of symptom specific factors, we consider especially two kinds of models as promising for a better understanding of the latent structure of MUS: First, a hierarchical model consisting of first and second order latent factors [8]; second, a bifactor or nested factor model that consists of a general factor and orthogonal specific group factors (e.g., [26, 27]). In contrast to the hierarchical model, the bifactor model assumes general as well as specific sources of variances included in the manifest variables. Consequently, a comparison of the hierarchical model and the bifactor model in case of MUS allows us to answer the question whether MUS represent one homogenous latent construct (in which case, the hierarchical model would be expected to show the best model fit) or whether MUS also share symptom specific variance distinct from the general factor (in this case, the bifactor model should represent the best fitting model). Because some of the studies that we have outlined also proposed correlated factor models we also want to test this type of model. Till now, the evidence for a latent model of MUS which only features one general factor (i.e., no further group factors) seems rather weak but nevertheless we will include a general factor model (without additional group factors) for comparison purposes.

To our knowledge, a bifactor model of somatic symptoms was tested successfully only once in a study by Thomas and Locke [15] in a very specific sample of patients with either epilepsy or nonepileptic seizures using different somatic complaint subscales of the MMPI-2-RF
The authors concluded that the best fitting model to their data was a bifactor model with different orthogonal latent factors, i.e., a general latent complaint factor, as well as content-specific symptom factors (i.e., a malaise factor, a gastrointestinal factor, a head complaint factor, and a neurological complaint factor; [15]).

Based on the existing knowledge regarding the structure of MUS, the aim of the following two studies was to develop an optimal model of the latent structure of MUS as a dimensional construct (Study 1) and to replicate and validate this model (Study 2). For this purpose, the PHQ-15 [29] as one of the most widely used dimensional measures assessing the 15 most important MUS was chosen in the following two studies.

**Study 1**

The primary aim of Study 1 was to establish an adequate model of MUS based on the PHQ-15. Building on previous studies on the latent structure of MUS (e.g., [9, 22, 13, 23]), we intended to compare four different models (Figure 1): First, a general factor model which is implied by the common practice to compute a sum score of all items; second, a four factor model differentiating different symptom types; third, a hierarchical model consisting of a general factor and lower level symptom specific factors; fourth, a bifactor model (e.g., [15]) in which each symptom shares specific as well as general components of systematic variance. The four models will be compared in terms of goodness of fit statistics and, whenever applicable, by direct model comparison tests through $\chi^2$-difference tests [30].

**Method**

**Participants.** As part of a study on stress, emotion regulation, and mental and physical health, questionnaires were sent to 3000 randomly selected adult inhabitants of 10 German cities...
and communes. Participants received information regarding the study rationale and provided written informed consent. Participation was voluntary and unpaid. The study protocol was approved by the local ethic committee of the Psychological Institute of the University of Mainz and the responsible Federal State Commissioner for Data Protection and Freedom of Information. 423 questionnaires were sent back. This sample was highly comparable to the German population concerning major socio-demographic aspects. Nine questionnaires had to be excluded due to completely missing subscales of at least one instrument. The final sample comprised 414 participants (54.1% women). The mean age of the participants was 47.2 years ($SD = 16.7$; Range: 18-89 years). 31.7% were unmarried, 59.4% married, 8.4% divorced or widowed. A total of 15.7% lived alone and 77.8% lived with their partner, family or other people in one household. Only 1.9% reported to be without graduation, 22.0% finished Secondary General School, 22.2% Intermediate Secondary School, 21.1% Grammar School classes A-level, and 31.9% University.

**Measures.** We applied the PHQ-15 [9, 29] which is part of the Patient Health Questionnaire (PHQ; [31, 32]). The PHQ-15 is a continuous self-report measure of somatic symptom distress over the last four weeks and consists of 15 somatic symptoms (e.g., headaches, back pain, dizziness). In response to the question “During the past 4 weeks, how much have you been bothered by any of the following problems?” participants are asked to rate their symptom distress for each of the 15 symptoms on a three-point Likert scale that ranges from *not bothered at all* up to *bothered a lot*. The PHQ-15 represents a well accepted continuous measure of MUS and is also recommended for a dimensional assessment of bodily symptoms and associated distress by the somatoform disorder task force for DSM-5 [33] (please see Table 1A in the Appendix for the entire PHQ-15). Prior to data analyses, two items were excluded due to gender specific content (Item 4: menstrual problems) and due to a very low base rate (Item 9: fainting
Data analyses. We conducted exploratory factor analyses and parallel analyses according to Horn [34] in SPSS 18.0.1. Confirmatory factor analyses were performed with Mplus 6.1 [35]. The analyses of the measurement models were conducted with the robust mean and variance adjusted weighted least squares (WLSMV) estimation, which is less sensitive to deviations from multivariate normal distribution and even yields very good results in case of comparatively small sample sizes [36, 37]. Because the $\chi^2$ test is sensitive to the sample size and the complexity of the model (e.g., [38], we used other descriptive fit measures for the evaluation of the model fit. As an absolute fit index we chose the RMSEA (Root Mean Square Error of Approximation). Furthermore, the CFI (Comparative Fit Index) and the TLI (Tucker-Lewis Index) as incremental fit indices are reported. Based on the recommendations of Schermelleh-Engel et al. [39], the RMSEA should be smaller than .08 or .05 to indicate an acceptable or good fit. CFI and TLI values greater than .95 can be considered as an adequate fit and values greater than .97 as a good fit. The four different models that will be tested are depicted in Figure 1.

Results

First, a general factor model with all of the 13 remaining PHQ-15 items loading on one latent factor was tested by using CFA. As detailed in Table 1, the model fit of this model was poor with none of the goodness of fit indices reaching acceptable values. Second, we tested a correlated group factor model. In order to specify such a model, we first conducted an exploratory factor analysis and used the parallel analysis procedure [34] to determine the number of existing factors. Based on the results of this analysis, we tested a model with four correlated
latent factors (a gastrointestinal factor, a pain factor, a fatigue factor, and a cardio-pulmonary symptoms factor) using CFA. Although the goodness of fit values indicated a better model fit compared to the general factor model, most of the statistics still did not reach the proposed cut-off scores for good or satisfactory model fit. Similarly, a hierarchical model with one general factor and four symptom specific lower order factors did not reveal satisfactory model fit.

Finally, a bifactor model (Model IV in Figure 1) was specified by allowing every item to load on a general factor as well as on one of four symptom specific factors. Goodness of fit statistics indicated a good to excellent model fit (Table 1). A direct comparison of the hierarchical factor model with the bifactor model by using the $\chi^2$-difference test revealed a significantly better model fit of the bifactor model ($\chi^2(7) = 42.83, p < .001$). By inspecting the factor loadings in the bifactor model (Figure 2), it is noticeable that 7 of the 13 manifest variables had significant factor loadings on the general (somatization) factor as well as on the symptom specific factors.

The strongest group factor is represented by the gastrointestinal factor, whereas the cardio-pulmonary symptom factor appears rather weak (without significant factor loadings). Interestingly and rather counterintuitive, particular symptoms (e.g., headache and dizziness) do not share symptom specific variances but only load significantly on the general factor.

**Please insert Table 1 and Figure 2 about here**

**Discussion of Study 1**

The primary aim of Study 1 was to model the latent structure of MUS by using the PHQ-15 as one of the most prominent questionnaires for the dimensional assessment of MUS. As a result of testing four different models that were either implied by common clinical practice or by previous factor analytic studies, a bifactor model of the symptoms included in the PHQ-15
yielded the best model fit. This model consists of one general factor on which all of the 13 included items loaded significantly, and four symptom specific factors (a gastrointestinal factor, a pain factor, a fatigue factor, and a cardio-pulmonary factor). In the bifactor model, all factors are by definition orthogonal and therefore they account for unique parts of variance in the symptom variables. Consequently, the bifactor model actually decomposes the variance of each manifest indicator variable into a general part (accounted for by the general factor) and into a specific part (accounted for by one of the four specific symptom factors).

By inspecting the factor loadings of the bifactor model, it becomes obvious that many of the symptoms show significant loadings on both levels of modeling, i.e., on the general factor as well as on one of the specific symptom factors. This implies that the variability of symptom distress as assessed by the PHQ-15 is simultaneously determined by two sources of systematic variance, that is, by the general factor and by a symptom specific factor. At this point, we can only speculate about the question what these two different sources of variance are. Possibly, the general factor refers stronger to an affective component associated with symptom experience, whereas the symptom specific factors are closer related to the sensory part of symptom experience. At this stage of interpretation, it might also be important to consider that the items of the PHQ-15 actually represent an amalgam of two things, that is, first the existence of a specific symptom, and second the level of distress associated with a certain symptom. Although, a distinction of these two levels of information might appear trivial from a clinical point of view, it might actually be possible that people experience symptoms without any distress.

Based on this observation it might be possible that the general factor mainly represents the symptom distress variance, whereas the symptom specific factors are stronger related to the pure (sensory) existence of a respective symptom. Alternatively, the general factor might be stronger related to somatization and the symptom specific variability to physically explainable
symptoms. Additionally, the symptom specific factors might be stronger related to certain functional somatic syndromes (e.g., IBS and FMS).

All in all, the results of Study 1 suggest that the latent structure of MUS, at least as measured by the PHQ-15, is complex and best empirically represented in a bifactor model as a combination of both, common and specific sources of variance regarding symptom experiences.

**Study 2**

The aims of Study 2 were twofold: First, the bifactor model of Study 1 should be replicated in an independent sample to test stability and generalizability of the model proposed in Study 1. Second, we intended to provide additional support for the construct validity of the bifactor solution of the PHQ-15 by exploring relations between the proposed PHQ-15 factors (the specific symptom factors as well as the general somatic symptom distress factor) and functional somatic syndromes. We hypothesized that common functional somatic syndromes like IBS might show significant associations to the general PHQ-15 distress factor as well as to certain specific symptom factors. Specifically, we hypothesized that IBS would show significant associations to the gastrointestinal PHQ-15 factor. To achieve these two aims, the data of 308 primary care patients were analyzed using nonparametric structural equation modeling analogous to Study 1.

**Method**

**Participants.** The data analyzed in Study 2 was collected as part of a larger Study on the prevalence of somatoform disorders in primary care [40]. This study was conducted in two primary care practices in Mainz, Germany. Both were part of the regular German health care system. Consecutive patients were asked to take part in a short study and to answer the PHQ-15
and some demographic questions. Those who agreed answered the questionnaires by themselves in the waiting room. The sample for Study 2 comprised 308 primary care patients (71.4% women) with a mean age of 47.2 years ($SD = 16.3$; Range: 18-87 years). Of these 308 participants, 32.5% lived alone, 37.3% with one and 29.8% with two or more people in one household. 20.1% finished Secondary General School, 20.5% Intermediate Secondary School, 27.3% Grammar School classes A-level and 31.2% University. For further details concerning the sample, please refer to [40].

**Measures.** As in Study 1, the PHQ-15 was used to assess somatic symptoms and associated distress. Again, prior to data analyses, two items of the PHQ-15 were excluded due to gender specific content (Item 4 on menstrual problems) and due to a very low base rate (Item 9: fainting spells). As part of a larger interview, the criteria of IBS were checked according to the Rome II criteria [41]. Thirty-four participants (11%) met the Rome II criteria.

**Data analyses.** Analogue to Study 1, the CFA was performed with Mplus 6.1 [35] using the WLSMV estimator. In order to rule out the possibility that alternative models, that turned out as unsatisfactory in Study 1, might reveal a better model fit in Study 2, we replicated the computation of the four different models of Study 1 (see Figure 1 for details).

**Results**

**Bifactor model of the PHQ-15.** As detailed in Table 2, only the bifactor model showed a good model fit. Again, the hierarchical factor model (Model III) and the bifactor model (Model IV) were directly compared using the $\chi^2$-difference test: The bifactor model fitted the data significantly better than the hierarchical model ($\chi^2(7) = 51.68, p < .001$).

Similar to Study 1, most of the manifest variables (10 out of 13) had significant factor loadings on the general (somatization) factor as well as on the symptom specific group factors (Figure 3). Each of the four symptom specific factors at least revealed two significant loadings.
In sum, the pattern of factor loadings is highly similar to the findings in Study 1, although the cardio-pulmonary factor appears stronger in this sample of primary care patients.

***Please insert Table 2 and Figure 3 about here***

**Associations between the PHQ-15 and IBS.** In order to test possible associations between the four latent variables of the bifactor model and the diagnosis of IBS, a latent regression model was tested. In this model, the IBS diagnosis (as a dichotomous variable) was regressed onto the five latent PHQ-15 factor variables (i.e., the general factor as well as the four specific symptom factors). The goodness of fit statistics indicated a good model fit. As hypothesized, the general factor (14%) as well as the gastrointestinal factor (42%) explained unique and significant parts of variability of the IBS diagnosis (Figure 4). Interestingly, none of the other three specific symptom factors (cardio-pulmonary, fatigue, or pain symptoms) was significantly associated with the IBS diagnosis.

***Please insert Figure 4 about here***

**Discussion of Study 2**

Study 2 aimed at replicating the bifactor model of somatic symptoms in the PHQ-15 of Study 1 and was intended to gain first evidence for the validity of the proposed bifactor structure. Both aims could be achieved: First, the bifactor model again turned out as the best fitting model and the convergent validity of the gastrointestinal factor could be demonstrated by significant and specific associations to IBS. The fact that we did not observe any significant relationships between the remaining three group factors (pain, cardio-pulmonary, and fatigue symptoms) and
the IBS demonstrates the discriminant validity of our bifactor model.

We had originally intended to test associations to other functional somatic syndromes as well (i.e., FMS and CFS). Unfortunately, the base rates of these syndromes turned out to be unacceptably low (< 2 %) in our sample of primary care patients. Future studies with larger sample sizes are needed to clarify the question of whether other functional somatic syndromes show the same picture of associations to common as well as specific symptom factors as demonstrated for IBS.

**General discussion**

The aim of the two studies was to clarify the latent structure of somatic symptoms assessed by the PHQ-15. As a result, the findings of both studies (in the general population and in primary care) unambiguously favor a bifactor model in which the variability of symptom distress is explained by a general factor (most likely representing somatization or an affective component of symptom perception) and symptom specific factors (that might reflect physiological factors). The bifactor model is in accordance with previous observations of a strong general factor of MUS (e.g., [6]) as well as the existence of specific symptom patterns (e.g., [10]).

In our bifactor model, the general factor reflects the communality among the different symptoms and it appears noteworthy that certain symptoms (e.g., headache or dizziness) are entirely determined by the general factor. Because headache and dizziness represent very common and general bodily symptoms that are closely related to emotional distress and diverse psychopathologies (e.g., mood and anxiety disorders), it appears likely that the general factor primarily reflects cognitive-emotional disturbances associated with symptom experiences and elevated stress levels. Moreover, the general factor might also reflect neurobiological alterations,
especially in the stress responsive system (i.e., the autonomous nervous system and the hypothalamic-pituitary-adrenal axis) that have been proposed as a common etiological factor in functional somatic syndromes [42] and somatoform disorders [43].

In general, the observation of a strong common factor underlying many different bodily symptoms is in line with previous studies on this topic suggesting that different MUS and patterns of MUS in terms of functional somatic syndromes may reflect one underlying common phenomenon (e.g., [12, 44]). However, the bifactor model implies, that most of the covered bodily symptoms are explained by two “causes” in terms of latent variables. Given recent research on the etiology and pathogenesis of specific functional somatic syndromes [45, 46] it is tempting to speculate that the specific symptom factors might reflect specific precipitating factors (e.g., infections, organic diseases, or environmental factors more generally), whereas the general factors covers a cognitive-affective and neurobiological diathesis that might be highly relevant for the perpetuation of the symptom experiences. In order to test these hypotheses and to validate the bifactor model, further prospective studies using not only subjective measures of symptom distress but also physiological and neurobiological measures are necessary.

First evidence for the external validity of the proposed bifactor model was gained by demonstrating that different factors of the bifactor model were significantly associated with the diagnosis of IBS. More specifically, IBS was found to be associated especially with the gastrointestinal symptom factor and the general symptom distress factor, but not with other symptom factors. It has to be noted that the finding of no significant association between the other specific symptom factors (i.e., pain, fatigue, and cardio-pulmonary symptoms) and IBS does not mean that these symptoms are of no importance in patients suffering from IBS. The significant contribution of the general distress factor rather suggests that almost every symptom included in the PHQ-15 is associated with IBS (e.g., symptoms of pain, fatigue, etc.). However,
when we consider symptom specific (i.e., incremental) variance components as implied by the bifactor model, only gastrointestinal symptoms predict IBS over and above the general symptom distress factor. Future research should test the hypothesis that other functional somatic syndromes (e.g., fibromyalgia and the chronic fatigue syndrome) might show a similar pattern of associations with the latent general factor and the respective symptom specific factors of the proposed bifactor model of MUS.

For research purposes (with large sample sizes), the bifactor model represents the most accurate and detailed latent structure and offers intriguing opportunities for further research on psychological, biological, and interactive causes, correlates, and consequences of MUS. With regard to the forthcoming classification of the somatoform disorders in DSM-5, the proposed bifactor structure (and the existence of a strong general factor) is largely in agreement with the proposal of one broad category of complex somatic symptom disorder (CSSD) for DSM-5 [47]. The concept of CSSD was introduced by the DSM-5 workgroup on somatoform disorders to account for the large overlap between subtypes of somatoform disorders [47]. Accordingly, CSSD is proposed to combine the previous diagnoses of somatization disorder, undifferentiated somatoform disorder, and hypochondriasis in one single diagnostic category. According to the current proposal of the DSM-5 workgroup [33], CSSD is defined by the presence of one or more somatic symptoms that are distressing (criterion A), excessive thoughts, feelings, or behaviors related to the symptoms or the associated health anxiety (criterion B), and symptom chronicity of at least 6 months (criterion C). Within the concept of CSSD, optional “specifiers” are suggested to account for different types of CSSD that are dominated by either (diverse) somatic symptoms, health anxiety, or pain symptoms [33]. It appears tempting to speculate that the variability covered by the symptom specific factors of the bifactor model reflects relevant information regarding possible subtypes of CSSD but further research on this issue is necessary.
Several limitations of our study are noteworthy: By using a questionnaire based measure of MUS, it is not possible to rule out possible physical explanations of the reported symptoms. Therefore, we might have included symptoms which actually are medically explainable. However, the distinction between medically explained and unexplained has been criticized in the past due to the unreliability of determining whether a certain symptom is medically explainable or not [48]. Given the high percentage of medically not fully explainable symptoms in primary care practice (e.g., 76% [40]; 57.5% [49]), it can be assumed that the majority of symptoms covered in the PHQ-15 was entirely or partly unexplained. Additionally, it has to be acknowledged that the PHQ-15 does not allow for a clear differentiation between the existence of a symptom (in terms of frequency and intensity) and the associated level of distress. Both kinds of information are inextricably interwoven in the current question and response format of the PHQ-15 (Table A1). Future studies with other assessment instruments (both self-report and interview-based measures) are therefore necessary to investigate the generalizability of the proposed bifactor model of MUS.

Another important aspect stems from the fact that the latent structure (including dimensionality) of any construct does not only depend on the questionnaire that is used to assess the construct but also on the particular sample (e.g., [50]). We tested our latent models of MUS in a primary care sample and a sample from the general population. The use of a primary care sample in Study 2 represents a major limitation and further studies on possible associations between the latent MUS factors and functional somatic syndromes should be conducted in larger samples of the general population. It is also not clear if the structure will be the same in a sample of psychosomatic patients (i.e., very high symptom burden) or a student population (i.e., probably rather low symptom burden). Given the rather complex bifactor model, the sample sizes of both studies have to be considered as rather small. A more detailed analysis of gender
differences in the proposed bifactor model by conducting multiple group analysis with a rigorous
test of measurement invariance is therefore beyond the scope of this article. Concerning the
different symptom types covered by the PHQ-15, it is obvious that the presence of a neurological
symptoms group factor is hardly possible because this instrument does not include enough of
these symptoms. This is not the case with other measures of MUS like, for example, the SOMS
questionnaire by Rief and Hiller [51] which includes many items of neurological symptoms.
Finally, it remains open in how far the proposed bifactor model can be replicated in samples of
patients suffering from other functional somatic syndromes (e.g., fibromyalgia and chronic
fatigue syndrome) that were not addressed in our study. Thus, we propose further studies with
alternative instruments and preferably larger sample sizes to allow for multiple group analysis in
order to replicate and extend our findings of a latent bifactor structure underlying MUS.
Acknowledgements

The study was supported by a research grant of the German Academic Exchange Service (DAAD) awarded to Michael Witthöft.
References

Figure 1: Overview of the different models tested in Study 1 and 2 (error terms of manifest indicator variables not shown)
Figure 2

Bifactor model of medically unexplained symptoms (MUS) in Study 1 ($N = 414$) with standardized factor loadings (all factor loading coefficients printed in bold are significant at $p < .05$; error terms of manifest variables not shown)

Goodness of fit statistics:
- Chi$^2$ (df): 90.31 (54)
- $p$ (Chi$^2$): .001
- CFI: .978
- TLI: .968
- RMSEA: .040
- 90% CI RMSEA: .025 - .055
Figure 3

Bifactor model of medically unexplained symptoms (MUS) in Study 2 \((N = 308)\) with standardized factor loadings (all factor loading coefficients printed in bold are significant at \(p < .05\); error terms of manifest variables not shown)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastroenterolog. symptoms</td>
<td>0.40</td>
<td>0.47</td>
<td>0.51</td>
<td>0.51</td>
<td>0.70</td>
<td>0.51</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.56</td>
<td>0.56</td>
<td>0.61</td>
<td>0.75</td>
</tr>
<tr>
<td>Cardio-pulmon. symptoms</td>
<td>0.32</td>
<td>0.40</td>
<td>0.43</td>
<td>0.54</td>
<td>0.70</td>
<td>0.43</td>
<td>0.43</td>
<td>0.56</td>
<td>0.44</td>
<td>0.56</td>
<td>0.44</td>
<td>0.37</td>
<td>0.79</td>
</tr>
<tr>
<td>Fatigue symptoms</td>
<td>0.65</td>
<td>0.65</td>
<td>0.49</td>
<td>0.49</td>
<td>0.66</td>
<td>0.66</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Goodness of fit statistics:

- Chi^2 (df): 101.39 (54)
- \(p\) (Chi^2): <.001
- CFI: .968
- TLI: .954
- RMSEA: .053
- 90% CI RMSEA: .037 - .069
Figure 4

Structural equation model of the PHQ-15 bifactor model (Study 2) predicting Irritable Bowel Syndrome (IBS; error terms of manifest variables not shown)

Goodness of fit statistics:
- Chi² (df): 113.34 (62)
- p (Chi²): <.001
- CFI: .968
- TLI: .953
- RMSEA: .052
- 90% CI RMSEA: .036 - .067

R² = .59
Table 1

*Goodness of fit statistics for the four different models tested in Study 1 (N = 414)*

<table>
<thead>
<tr>
<th></th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-Factor</td>
<td>Correlated</td>
<td>Hierarchical</td>
<td>Bifactor</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Factor Model</td>
<td>Model</td>
<td>Model</td>
</tr>
<tr>
<td>(\chi^2) (df)*a</td>
<td>407.17 (65)</td>
<td>144.66 (59)</td>
<td>144.32 (61)</td>
<td>90.31 (54)</td>
</tr>
<tr>
<td>CFI*</td>
<td>.789</td>
<td>.947</td>
<td>.949</td>
<td>.978</td>
</tr>
<tr>
<td>TLI*</td>
<td>.746</td>
<td>.930</td>
<td>.934</td>
<td>.968</td>
</tr>
<tr>
<td>RMSEA*</td>
<td>.113</td>
<td>.059</td>
<td>.057</td>
<td>.040</td>
</tr>
<tr>
<td>90% CI RMSEA</td>
<td>.102-.123</td>
<td>.047-.072</td>
<td>.045-.070</td>
<td>.025-.055</td>
</tr>
<tr>
<td>Evaluation of</td>
<td>poor</td>
<td>acceptable</td>
<td>acceptable</td>
<td>good</td>
</tr>
<tr>
<td>model fit*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* \(a\) All \(\chi^2\)-values were highly significant, \(p < .001\). *Conventions for good model fit (Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003): CFI ≥ .95; TLI ≥ .95; RMSEA < .06; results of model comparisons with \(\chi^2\)-difference tests: Model II is to be preferred compared to Model III (\(\chi^2(2) = 3.02, p = .22\)) and Model IV should be preferred over Model III (\(\chi^2(7) = 42.83, p < .001\)).
Table 2

*Goodness of fit statistics for the four different models tested in Study 2 (N = 308)*

<table>
<thead>
<tr>
<th></th>
<th>Model I 1-Factor Model</th>
<th>Model II Correlated Factor Model</th>
<th>Model III Hierarchical Model</th>
<th>Model IV Bifactor Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ (df)</td>
<td>356.87 (65)</td>
<td>160.83 (59)</td>
<td>160.80 (61)</td>
<td>101.39 (54)</td>
</tr>
<tr>
<td>CFI</td>
<td>.803</td>
<td>.931</td>
<td>.933</td>
<td>.968</td>
</tr>
<tr>
<td>TLI</td>
<td>.763</td>
<td>.909</td>
<td>.914</td>
<td>.954</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.121</td>
<td>.075</td>
<td>.073</td>
<td>.053</td>
</tr>
<tr>
<td>90% CI RMSEA</td>
<td>.109-.133</td>
<td>.061-.089</td>
<td>.059-.087</td>
<td>.037-.069</td>
</tr>
<tr>
<td>Evaluation of model fit*</td>
<td>poor</td>
<td>poor</td>
<td>poor</td>
<td>good</td>
</tr>
</tbody>
</table>

*Note. $^a$ All $\chi^2$-values were highly significant, $p < .001$. *Conventions for good model fit (Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003): CFI $\geq$ .95; TLI $\geq$ .95; RMSEA < .06; results of model comparisons with $\chi^2$-difference tests: Model II is to be preferred compared to Model III ($\chi^2(2) = 4.38, p = .11$) and Model IV should be preferred over Model III ($\chi^2(7) = 51.68, p < .001$).
### Appendix

**Table A1**

**Patient-Health-Questionnaire 15 (PHQ-15)**

<table>
<thead>
<tr>
<th></th>
<th>During the past 4 weeks, how much have you been bothered by any of the following problems?</th>
<th>Not bothered at all</th>
<th>Bothered a little</th>
<th>Bothered a lot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stomach pain</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>2</td>
<td>Back pain</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>3</td>
<td>Pain in your arms, legs, or joints (knees, hips, etc.)</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>4</td>
<td>Menstrual cramps or other problems with your periods</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>5</td>
<td>Pain or problems during sexual intercourse</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>6</td>
<td>Headaches</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>7</td>
<td>Chest pain</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>8</td>
<td>Dizziness</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>9</td>
<td>Fainting spells</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>10</td>
<td>Feeling your heart pound or race</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>11</td>
<td>Shortness of breath</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>12</td>
<td>Constipation, loose bowels, or diarrhea</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>13</td>
<td>Nausea, gas, or indigestion</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>14</td>
<td>Trouble sleeping</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>15</td>
<td>Feeling tired or having low energy</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

*Note.* Adapted from [www.phqscreeners.com](http://www.phqscreeners.com)