

Postprint Version	1.0
Journal website	http://linkinghub.elsevier.com/retrieve/pii/S1438-4639(15)00117-0
Pubmed link	http://www.ncbi.nlm.nih.gov/pubmed/26358929
DOI	10.1016/j.ijheh.2015.08.006

This is a NIVEL certified Post Print, more info at <http://www.nivel.eu>

Assessing non-specific symptoms in epidemiological studies: Development and validation of the Symptoms and Perceptions (SaP) questionnaire

JORIS YZERMANS^a, CHRISTOS BALIATSAS^A, SANDRA VAN DULMEN^{A, B, C}, IRENE VAN KAMP^D

^a NIVEL (Netherlands Institute for Health Services Research), Utrecht, The Netherlands

^b Department of Primary and Community Care, Radboud University Medical Center, Nijmegen, The Netherlands

^c Faculty of Health Sciences, Buskerud and Vestfold University College, Drammen, Norway

^d National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

ABSTRACT

Objectives

To describe the development of the Symptoms and Perceptions (SaP) questionnaire, a new instrument measuring diversity and severity features and related perceptions of non-specific (physical and psychological) symptoms (NSS) and to test its reliability and validity within an epidemiological setting.

Methods

First, a Delphi study was performed to develop the items of the questionnaire. Its psychometric properties were then tested within the context of an epidemiological study. This study yielded questionnaire data and general practice registry data from a sample of 5933 Dutch adults. Principal component analyses (PCA) were performed to test the factorial structure of the symptom assessment subscales of the SaP. Indicators of reliability and convergent validity were examined separately for the questionnaire parts of symptoms and perceptions. Criterion validity of the symptom scores of the SaP was also explored.

Results

The factorial structure of the symptom subscales reflected meaningful sets of clusters, indicative of the representation of all relevant organ systems. The SaP showed high reliability and good convergent validity for both the symptoms and perceptions part. It was also able to identify, to a certain degree, cases of

diagnosed somatoform disorders and multiple NSS in primary care. Additionally, the scores on the “psychological” factor fairly predicted the diagnosis of common mental disorders.

Conclusions

The SaP questionnaire is a reliable and valid self-reported measure of diverse features of NSS and corresponding perceptions. It can also be a useful tool for the identification of multiple somatic symptoms and related disorders in general practice and assessment of psychological problems in epidemiological studies. Further validation of the questionnaire in different samples and settings will establish the promising psychometric properties demonstrated in the present study.

Abbreviations

SaP, Symptoms and Perceptions (questionnaire);
NSS, non-specific symptoms;
GP, general practitioner;
ICPC, International Classification of Primary Care;
B-IPQ, brief Illness Perceptions Questionnaire;
PCA, principal component analysis;
ROC, receiver operating characteristic;
AUC, area under the curve

1. INTRODUCTION

Symptoms such as fatigue, headache, sleep difficulties, gastrointestinal disturbances and musculoskeletal pain are often called non-specific because they are very common in the general population, occur in multiple organ systems and can be caused by a variety of factors, sometimes unknown (Engel and Katon, 1999). When presented in primary care, up to 50% of non-specific symptoms (NSS) are not explained by organic pathology (Barsky and Borus, 1999 and Körber et al., 2011). The term “medically unexplained symptoms” is regularly used to define such complaints in research and healthcare, although unfavorable among the patients (Henningsen et al., 2011). Clusters of persistent NSS have been mainly referred to as functional somatic syndromes or somatoform disorders (Mayou et al., 2005 and Wessely et al., 1999).

Rapid developments in numerous fields of society and technology such as increasing urbanization, use of energy resources and expansion of mobile communication networks introduce potential threats to population's health (Gavrilescu et al., 2015), underpinning the attribution of NSS to environmental agents, even at lower levels than the established safety standards. Although the harmful effects of a number of environmental pollutants are well-documented (WHO, 2009), often the attributed environmental causes of NSS are not adequately supported by scientific evidence as in the case of idiopathic environmental intolerance (IEI) and its manifestations (Das-Munshi et al., 2006, Kreutzer, 2000 and Rösli et al., 2010). This is the main reason why such symptomatic conditions are considered part of the functional somatic syndrome spectrum (Wessely et al., 1999).

In recent years, NSS in the general population have been an object of study in relation to various environmental stressors such as electromagnetic fields (EMF) and

noise (Baliatsas et al., 2015 and Zijlema et al., 2015) and also in the aftermath of disasters and environmental incidents (van den Berg, 2007 and Yzermans and Gersons, 2002). Environmental exposures related to modern technology and disasters, and perceptions of those exposed, increasingly influence health and health-related behavior (Page et al., 2006). This underlines the need to measure and monitor the health status of affected populations or populations at risk. Frequency of NSS and related perceptions in epidemiological research are primarily assessed by using self-reported symptom scales. Many of these questionnaires are not validated or they assess exclusively physical symptoms of one particular organ area; even validated symptom lists suffer from major conceptual and methodological limitations, as recently reviewed (Zijlema et al., 2013).

First, there is large between-questionnaire variation in terms of the symptoms assessed, the rating scale on which they are scored and the timeframe used. Second, most of the existing questionnaires assess either symptom prevalence or burden, while it is important to combine both diversity and severity of the reported complaints. Third, many of the existing questionnaires exclude potentially relevant symptoms use a quite long recall period and have restricted applicability to large-scale studies involving various population groups.

Besides, most of the existing symptom questionnaires do not conceptualize psychological health as part of physical health and vice versa, which is gradually becoming an outdated perspective (Noyes et al., 2008). As a result, population studies employ separate screening instruments for the assessment of psychological problems in clinical settings, which could make some participants feel uncomfortable or unwilling to respond (Dozeman et al., 2011).

Finally, some questionnaires measure general symptoms or perceived health, while others focus on the screening of somatoform disorders based on symptom count scores. Although such symptom scores comprise an established indicator of functional impairment and illness behavior (Creed et al., 2012, van den Berg et al., 2005 and van der Windt et al., 2008) and can be a useful tool in the identification of somatoform symptoms (Körber et al., 2011), only a thorough medical assessment can reliably determine whether the reported symptoms are associated with an organic cause or not (Carson et al., 2015). The use of data from primary care registries makes it feasible to involve clinical judgment in large studies (Okkes et al., 2002 and van den Berg, 2007). However, persistent presentation of NSS in primary care is not very common (van den Berg, 2007 and Verhaak et al., 2006a).

The combination of self-reported symptoms and medical registry data is a sound approach to gain insight into respondents' health and symptom characteristics and an asset toward the minimization of outcome misclassification often occurring in population studies (Baliatsas et al., 2015). This requires, though, that the methodology in terms of symptom selection and categorization of the questionnaire corresponds with routine registration in health care, especially in general practice. In addition, there are multiple nonbiomedical factors that play a role in symptom intensity and seeking medical help, such as illness perceptions in response to NSS (Barsky, 1981, Foster et al., 2008, Frosthalm et al., 2007 and Koloski et al., 2001). The evaluation of such factors is also in line with the new diagnosis for somatoform disorders "Somatic symptom and related disorders" in the fifth edition of the Diagnostic and Statistical Manual of mental disorders (DSM-5) (American

Psychiatric Association, 2013). Nevertheless, patients' beliefs about their NSS are usually underinvestigated, both in general practice and research.

The aim of this study was to introduce and test the psychometric properties of a newly developed questionnaire to: (1) include symptoms from all relevant organ systems and combine diversity with features of severity using a reliable timeframe; (2) allow the concurrent assessment of NSS and cognitive representations that people have about their most important symptoms; (3) be applicable to large-scale epidemiological studies; (4) make no a priori distinction between "explained" and "unexplained" symptoms, acknowledging the multifactorial etiology of symptomatology and the interacting physical, psychological and social components of symptom report; and (5) be efficiently combined with GP-registered data.

2. METHODS

2.1. Questionnaire development and content validity

Originally, the SaP questionnaire was designed for use in vulnerable population subgroups in the aftermath of disasters and environmental incidents. The development of the questionnaire started with a two-round Delphi study (Hasson et al., 2000) among healthcare providers in primary care. Fig. 1 illustrates the steps of the questionnaire development and results that led to the final version. The input for the panel of experts was mainly based on the International Classification of Primary Care (ICPC) (Lamberts and Wood, 1987), a registry used by GPs in several European countries, including symptoms and diagnoses. This classification was used to make sure that all health areas and organ systems were represented and that the questionnaire would be comparable to primary care electronic medical records. Symptoms that were too specific or rare (incidence $\leq 1/1000$ per year) were omitted. The list was completed with additional somatoform symptoms not covered by the ICPC. In order to obtain an indication of the nature, severity and diversity of health problems, the scoring system was formulated on prevalence/number, duration and symptom-related encounters with GPs. Illness perceptions were included in the list using an adapted version of an already established instrument, the brief Illness Perception Questionnaire (B-IPQ) (Broadbent et al., 2006 and Kaptein et al., 2004), which referred to symptoms instead of illness.

[FIG. 1.]

In the first Delphi round it was assessed whether the list of symptoms and perceptions was complete and relevant for its general intended purpose. Results from the first round were summarized and sent in the second round to all participants, along with a new score form to assess relevant symptoms to be included from the list. After authors' screening, the pre-final list contained 51 symptoms assessed by three scoring systems (number, duration and related GP contacts). There was also an optional open-ended item so that respondents could add more symptoms if needed/relevant. Consensus was also reached on linking symptoms with subsequent questions on perceptions. It was decided that perceptions should refer to a maximum of three symptoms considered as most important by respondents.

To assess the readability of the resulting questionnaire and to identify any problems in format and item interpretation, the questionnaire was piloted in a small general population sample, recruited via social and professional networks. The resulting list from the Delphi was pre-tested among thirteen women and five men with different age (mean = 40, range 18–66) and level of education ($n = 6$ of lower and $n = 12$ of higher education). Assessment was carried out based on cognitive interviewing using think-aloud and verbal probing techniques (Priede and Farrall, 2011 and Willis, 2004), and in written form using a feedback form.

The structure, response burden and relevance of the revised version of the piloted questionnaire was tested (online and in a paper version) among survivors of a firework disaster in the Netherlands, ten years after the event. The scale was sent out to 826 subjects with a response rate of 72%. Response to the questions was on average 97% per item. Fifty percent of the responses named two to three possible causes for their symptom(s).

Despite the satisfactory results in terms of response, the most consistently reported disadvantage of the scale in the pilot studies was its length. A shorter version was then produced intended for use in large-scale epidemiological studies, based on: (1) questionnaire items on which 100% consensus had been reached, (2) symptoms with the highest prevalence in the sample of disaster survivors and/or the general population (GP patients) and (3) symptoms that are often non-specific in nature (Robbins et al., 1997).

2.2. Symptom scoring

The present version of the SaP questionnaire includes 28 items which correspond to non-specific symptoms, physical as well as psychological, similar to those often reported in general practice. It assesses three symptom features represented by corresponding subscales:

(1) Symptom number/prevalence (“Has this complaint bothered you in the past month?”; potential responses: “Yes” = 1/“No” = 0).

(2) Symptom duration (“If so, for how long have you been bothered by this complaint?”; potential responses: “<1 month” = 1/“1–3 months” = 2/“4–6 months” = 3/“>6 months” = 4); no symptom report was rated as zero.

(3) GP-presentation/Health care utilization related to the reported symptom(s) (“Have you consulted a GP for this complaint in the past year?”; potential responses: “No” (or no symptom experience at all) = 0/Yes, in the past 4 weeks = 1/Yes, between 1 and 12 months ago = 1).

A higher total score indicates increased report of symptomatology, longer duration and higher symptom-related health care utilization. It is suggested that respondents with more than five symptom items missing are excluded from the analyses.

2.3. Assessment of corresponding perceptions

In the adapted version of the B-IPQ used as part of the SaP, only the perceptions regarding the most important symptom (as perceived by the respondent) are assessed.

The scoring of symptom perceptions is similar to the original version of the B-IPQ (Broadbent et al., 2006 and Kaptein et al., 2004). The B-IPQ items are scored on a 11-point Likert scale (0–10) and can be analyzed either as a sum score or individually, since each item represents a different perception dimension: consequences (perceived impact of symptom in life), timeline (perceived duration), personal control, treatment, symptom concern, symptom

comprehensibility/understanding, and emotional response. Since the perception questions are already referring to symptoms, the item “identity” of the original version of the B-IPQ was not included. Furthermore, compared to the original version, the question on treatment was slightly changed with a yes or no option, so that people who do not receive treatment for their symptom can also complete this question by sharing their own thoughts on desirability of treatment. The score for “treatment” is different for people who are and who are not receiving treatment, because the questions are about either their experience of effectiveness or about the perceived need for treatment. Finally, one open-ended question (instead of three in the original version) is included, about the main perceived cause for the most important symptom among the ones reported.

2.4. Epidemiological study

Data for the present analysis were collected within the framework of the epidemiological risk assessment study EMPHASIS (Baliatsas et al., 2015), which investigated the association between non-specific physical symptoms in relation to exposure to electromagnetic fields (EMF). A cross-sectional study was conducted from January to June 2011 among 5933 adult (≥ 18) Dutch citizens. Prior to the study, an informed consent was provided. Self-reported data was collected based on an extensive questionnaire (filled out electronically or in a paper version), consisted of four domains: (a) residential environment, (b) health, (c) well-being and (d) household/demographic information. Data from electronic medical records of health problems registered in general practices were also obtained, from the primary care database of the Netherlands Institute for Health Services Research (NIVEL). The study was carried out according to the Dutch legislation on privacy. According to the Dutch Medical Research Involving Human Subjects Act, ethical approval was not required. Details regarding the study design and sampling are presented elsewhere (Baliatsas et al., 2014 and Baliatsas et al., 2015).

2.5. Self-reported measures to test convergent validity

The following validated questionnaires were used in the present study to correlate with the SaP questionnaire:

The “general health” subscale of the RAND-36 Health Survey questionnaire (Van der Zee and Sanderman, 1993) was used to measure general well-being.

The 7-item Personal Well-being Index (PWI-A) which assesses various aspects of quality of life on a 11-point scale (International Well-being Group, 2006).

A 10-item version of the Groningen Sleep Quality Scale (Visser et al., 1978). Higher scores indicate lower sleep quality.

Perceived control was assessed using two items from a Dutch version of the Life Orientation Test (LOT) (Vinck et al., 1998): “I am always optimistic about my future” and “I hardly ever expect things to go my way”. An additional item was added, namely “If I try I can influence the quality of my living environment”, to enhance the individual sense of control that can lead to a positive outcome. The score is rated on a 5-point Likert scale ranging from strongly disagree to strongly agree (0–4). After proper reversals, a higher sum score reflects less perceived control.

The 12-item version of the General Health Questionnaire (GHQ-12) (Goldberg, 1973), which is a measure of mental distress, often used to screen for non-psychotic mental health problems. A cut-off point of >3 on the GHQ-12 (Schreuders et al.,

2007) was also used as a classification variable to assess criterion validity of the SaP scores on the psychological symptom cluster.

2.6. GP-registered non-specific symptoms and psychiatric morbidity to test criterion validity

Registry-based symptoms from general practices were coded by the GPs according to the International Classification of Primary Care (ICPC) (Lamberts and Wood, 1987). To evaluate the clinical judgment of the general practitioner on the symptoms, we used “episodes of care” (Biermans et al., 2008), which represents the period from the first presentation of a health problem to a general practice until the completion of the last encounter for the same problem (WONCA, 1995). In the present study, this period was set between July 2010 and June 2011.

To assess the prevalence of multiple NSS, a selection was made of GP-presented symptoms similar/corresponding to the symptoms included in the SaP questionnaire. When necessary, registered symptoms were clustered; for instance, the item “headache” in the self-reported list corresponded to the ICPC codes N01 (headache) and N02 (tension headache).

Regarding psychological and psychiatric symptoms the assessment of somatoform disorders diagnosed in general practice was based on episodes of somatization disorder (P75) and neurasthenia (P78).

The prevalence of non-psychotic mental problems was based on episodes of Anxiety (ICPC codes: P01, P74), Depression (P03, P76), Phobia/compulsive disorder (P79), and Personality disorder (P80).

2.7. Statistical analysis

Analysis of variance (one-way ANOVA), Pearson's correlations and unpaired samples *t*-tests were performed as descriptive analyses. Corrected item-total correlations were calculated to test whether each symptom item was consistent with the average behavior of the scale. A value of ≥ 0.2 is generally considered as acceptable (Kline, 2014). To explore the underlying structure of the symptom scale, principal component analysis (PCA) was performed; oblique (Direct Oblimin) rotation with Kaiser normalization was used, which allows for correlation among the factors. The following basic criteria were applied to assess factor solutions: (1) minimum eigenvalues of 1; (2) factor loadings of ≥ 0.35 ; (3) extraction of theoretically interpretable/meaningful set of factors. Reliability was assessed using Cronbach's alpha. Values of 0.70 or higher indicate satisfactory internal consistency (Nunnally and Bernstein, 1994). Convergent validity was evaluated based on Pearson's correlations between the total scores of the three symptom scoring methods and conceptually/theoretically related measures. A square-root transformation was applied to the sum score of the “GP-presentation” subscale because of exceeding kurtosis. Test-retest reliability could not be assessed due to the study design. Finally, we assessed criterion validity; the ability of the SaP scale to “detect” patients with somatization and multiple NSS in general practice was examined by plotting receiver operating characteristic (ROC) curves and calculating sensitivity and specificity based on different cut-off values. Analyses were carried out with IBM SPSS Statistics (SPSS Inc. version 19, Chicago, IL, USA) and MedCalc (MedCalc Software bvba version 15.6, Ostend, Belgium).

3. RESULTS

3.1. Descriptive analyses

Results of the non-response analysis and details regarding respondents' socio-demographic profile and health characteristics are described elsewhere (Baliatsas et al., 2014 and Baliatsas et al., 2015). Table 1 shows the distribution of basic demographic characteristics and outcome scores in the current sample. Missing data for a given item varied between 0.5% and 6%.

[TABLE 1]

Score on the subscales "number of symptoms" and "GP-presentation" correlated significantly with younger ($r = -0.09, p < 0.001$) and older age ($r = 0.12, p < 0.001$). No association was found between symptom duration and age.

Symptom scores were also associated with gender ("symptom number": mean women 6.8, SD = 4.8 vs. men 5.2, SD = 4.5, $p < 0.001$; "symptom duration": 16.5, SD = 15.2 vs. 12.6, SD = 14.2, $p < 0.001$; "GP-presentation": 1.99, SD = 3.0 vs. 1.5, SD = 2.8, $p < 0.001$), education level ("symptom duration": mean lower 15.5, SD = 16.3 vs. higher 14.8, SD = 14.8, $p < 0.01$; "GP-presentation": 2.3, SD = 3.4 vs. 1.3, SD = 2.3, $p < 0.001$) and foreign background (mean 79.8 SD = 19.5 vs. rest of the sample 76.1 SD = 19.0, $p < 0.001$).

There were also significant associations between symptom perceptions and demographic characteristics; these were similar for participants who followed a treatment for their most important symptom and those who did not. Indicatively, within the "no treatment" group, the sum score of the adapted B-IPQ was associated with age ($r = 0.17, p < 0.001$), education level (mean lower 38.3, SD = 13.1 vs. higher 29.9, SD = 12.1, $p < 0.001$) and foreign background (mean 36.9 SD = 13.1 vs. rest of the sample 32.4 SD = 12.6, $p < 0.001$). No gender differences were observed. There was large variation in responses to the open-ended questions of symptom perceptions. The symptoms considered as the most important by the participants were fatigue, leg/hip/knee/foot symptoms, neck or shoulder symptoms and headache. Qualitative analysis also showed that the most frequently reported attributed causes were: psychosocial factors/stress/pressure in daily life/negative life events, excessive workload, occupational stress, unexplained cause, unhealthy lifestyle/habits, aging, environmental agents/weather/temperature, congenital factors, (chronic) medical conditions and injuries.

3.2. Factorial structure

Principal component analysis identified 6 factors for the scoring methods "symptom number" and "symptom duration" and 7 factors for "GP-presentation", explaining between 43% and 46% of the total variance, respectively.

The factor structure was identical for these two subscales and mainly reflected clusters of psychological/neurovegetative, musculoskeletal, gastrointestinal, cardiac, pulmonary, and symptoms from several other organ systems. Regarding the "GP-presentation" subscale, seven factors were extracted that accounted for 49% of the variance. Factorial structure reflected similar clusters of symptoms, in line with the ICPC clustering. Table 2, Table 3 and Table 4 show the results of the PCA per scoring method.

[TABLE 2.]

[TABLE 3.]

[TABLE 4.]

3.3. Reliability

Item-total correlations ranged between 0.23 and 0.5 for the score on number of symptoms, 0.26–0.56 for symptom duration and 0.2–0.52 for symptom-related healthcare utilization. Cronbach's α coefficients for each subscale, when including all items, ranged between 0.83 and 0.86.

Regarding the perceptions part, internal consistency of the scale was $\alpha = 0.71$ among participants who did not follow a treatment for their most important symptom and $\alpha = 0.62$ for the “treatment” group. Item-total correlations varied between 0.2 and 0.64, with only the item “comprehensibility” having a lower value (0.1).

3.4. Convergent validity

As shown in Table 5, the total score for each of the three symptom features was strongly associated with worse perceived health, lower sleep quality, lower quality of life, higher levels of psychological distress, lower perceived control and more negative symptom perceptions (Table 5). The sum score of the adapted version of the B-IPQ was also correlated with the aforementioned variables, with coefficients ranging between 0.29 and 0.49 (absolute values).

[TABLE 5]

3.5. SaP scores as indicator of GP-registered somatoform disorders and multiple non-specific symptoms

Criterion validity was tested using the area under the curve (AUC). Values ranged between 0.7 and 0.79 ($p < 0.001$), indicating fair accuracy and discriminatory power (Fig. 2). Table 6 shows the cut-off points with the highest patient rates of detection of patients with somatoform disorders and multiple (>2) GP-registered NSS for each SaP subscale.

[Fig. 2.]

[TABLE 6.]

3.6. Using the “psychological” factor of the SaP as indicator of mental disorders in general practice

The AUC values are illustrated in Fig. 3, representing generally good classification results.

[FIG. 3.]

Table 6 presents the most optimal SaP thresholds for the identification of mental disorders, using the prevalence of non-psychotic mental disorders and the GHQ-12 score (≥ 4) as “gold standards”.

4. DISCUSSION

The aim of this paper was to introduce a new tool to assess NSS and related perceptions in large-scale studies and examine its psychometric properties. Results showed that the SaP questionnaire is an easy-to-use, valid and reliable instrument for measuring NSS in the general population, within the context of epidemiological research. During its development stage, the content validity of the questionnaire was established using a panel of experts in primary healthcare. After piloting the questionnaire twice, the final version was tested in a large epidemiological study. Both the symptoms and perceptions part of the questionnaire showed very good internal consistency. Convergent validity was highly satisfactory as indicated by the generally consistent and strong correlations with conceptually relevant constructs. The SaP questionnaire assessment is based on the total symptom score approach and takes into account important features of symptomatology such as number, duration and related healthcare utilization. Besides determining whether health complaints are psychological or physical, medically unexplained or not, it is more relevant to identify features of NSS that make them disabling and influence their clinical course (van den Berg, 2007). Total scores of NSS comprise an established indicator of functional impairment and healthcare utilization in the general population and different patient groups (Creed et al., 2012, Tomenson et al., 2013, van den Berg et al., 2005 and van der Windt et al., 2008) and have been shown to be prognostic for somatoform disorders (Kroenke et al., 1998).

The criterion validity of the NSS scores was also explored, with the SaP subscales being able to identify, to a certain degree, cases of diagnosed somatoform disorders and multiple NSS in primary care. Results regarding the AUC values and sensitivity and specificity of potential cut-offs were at least comparable with those of recent validation studies on broadly used questionnaires such as the PHQ (patient health questionnaire)-15 and PSC (physical symptom checklist)-51, while the SaP outperformed the somatization scale of the four-dimensional symptom questionnaire (4DSQ) (de Vroege et al., 2015 and Terluin et al., 2006). However, differences in operationalization of the employed gold standards hinder comparability. In general, the subscales “symptom number” and “symptom duration” seemed to perform better compared to “GP-presentation” against the gold standards; combination of the three subscales for diagnostic purposes might maximize the potential for correct classifications.

The factorial structure of the symptom subscales reflected a meaningful set of clusters, similar to the ICPC classification and indicative of the representation of several relevant organ systems. There were similarities with that of established symptom lists (Kroenke et al., 1998 and Lee et al., 2011), with the difference that the SaP, includes more items/symptoms and covers more organ systems, thus some differences were to be expected. The cluster with the largest explained variance was the one related to seven psychological/neurovegetative symptoms, showing a good internal consistency. We therefore explored it as an indicator of common mental disorders, based on both GP-registered and self-reported criterion/classification variables. We consider the current results as promising, given that its performance approaches that of established screening instruments which use a considerably larger number of questions (Gill et al., 2007, Holi et al., 2003 and Schmitz et al., 1999). Though a direct comparison with these instruments is not possible, since their validation was primarily based on different golden standards, such as structured

interviews. In addition to being in agreement with the abandonment of the mind-body distinction, the parallel assessment of psychological symptoms may help reduce potentially experienced stigma and related social bias by implicitly indicating that psychological health is part of general health. It also facilitates researchers to screen for psychiatric morbidity without employing additional instruments, and study interrelationships between physical and psychological symptoms.

Although the role of questionnaires measuring NSS exceeds the boundaries of a diagnostic tool, they are often solely treated or developed as such, which makes them inapplicable to large-scale studies. For example, a major use of symptom questionnaires is in risk assessment studies, where exposure-outcome associations are investigated. Considering the large variation of symptoms and physiologic reactions attributed to environmental exposures such as EMF, diversity in symptom selection and assessment is crucial. It is noteworthy though that many of the existing questionnaires are unsuitable for population studies, while only about 15% assesses more than one feature of symptoms (Zijlema et al., 2013).

In addition to its good psychometric properties discussed above, a major advantage of the SaP questionnaire is its suitability for use in large-scale epidemiological studies as indicated by the following aspects incorporated in its design: (1) The measurement of relevant (both physical and psychological) NSS and symptom features in combination with the evaluation of corresponding perceptions, which is in agreement with a biopsychosocial perspective on NSS (Creed et al., 2012). The SaP questionnaire may seem elaborate compared to shorter questionnaires; we would rather consider this an advantage, since it is part of a thorough assessment.

Commonly used scales such as the PHQ-12 do not include symptoms that might be of importance such as memory or concentration problems and skin symptoms, while the somatization scale of the 4DSQ does also not assess fatigue. Furthermore, recent systematic evaluations of health questionnaires have shown no evidence for an association between questionnaire length and response burden (Rolstad et al., 2011).

(2) The content and structure of the SaP questionnaire allows for the direct comparison between self-reported symptoms and data from health care providers. This enables the combination of the comparative advantages of both assessments which overcomes shortcomings such as outcome misclassification and can enhance between-study comparability.

(3) The consideration of health perceptions as part of the questionnaire provides the possibility to simultaneously gain useful information regarding the severity, importance and perceived causes (possible attribution from exposure to an environmental agent) of experienced health problems, and may provide a generic indication regarding prognosis of symptoms on a large scale. (4) Use of a time reference that minimizes recall bias. (5) The fact that no particular training is needed for its administration. (6) Flexibility in terms of item and subscale use, since the SaP questionnaire is not, per se, a diagnostic tool. For instance, items on physical symptoms can be used independently in risk assessment studies to assess prevalence or duration of symptoms that could be of interest in relation to a given exposure. Moreover, the sum score of the physical symptoms of SaP is a strong indicator of functional impairment and illness behavior among environmentally sensitive individuals (Baliatsas et al., 2014). Finally, as mentioned earlier, the cluster of psychological symptoms can be used to adjust for common mental problems.

Among the strengths of the present investigation are the large sample size which is an important advantage when performing psychometric/diagnostic accuracy studies (Leeflang et al., 2008) and the multiple health indicators and diagnoses against which the SaP was validated. Additionally, we used a primary care database that reflects real-life practice and a registry system with well-documented reliability (Okkes et al., 2002).

A number of potential limitations and research challenges should be acknowledged. First, test–retest reliability was not assessed because a follow-up was not part of the study design. Second, the diagnoses of the GPs can be subject to variations (Rosendal et al., 2003) while underdiagnosis of psychiatric disorders in general practice is well-documented (Nuyen et al., 2005 and Verhaak et al., 2006b), although this is unlikely for the more severe cases (Piek et al., 2012). The stability and reliability of the SaP over time should be assessed in future research. It would also be useful to test the SaP questionnaire in different samples and settings, in order to assess the latent structure of the reported symptoms based on confirmatory factor analyses (Witthöft et al., 2013) and establish its psychometric properties and cut-offs for diagnostic/classification purposes. The role of the perception part in the identification of clinically defined outcomes such as somatoform disorders, should be examined as well. Criterion validity should be confirmed using additional classification systems as “gold standards” such as the recent edition of DSM-5 (APA et al., 2013).

5. CONCLUSIONS

The SaP questionnaire is a free, reliable and valid instrument that assesses features of NSS that are commonly presented in primary care. Unlike the majority of the existing symptom questionnaires, SaP reflects multiple and clearly identifiable organ areas, while it takes into account accompanying perceptions. It is suitable for use in large-scale studies and can be useful in the identification of people with NSS and related disorders as well as common mental problems in general practice. Considering the multifactorial nature of NSS and the current measurement challenges within the context of epidemiological studies, the SaP questionnaire might contribute to a more efficient assessment.

FUNDING SOURCE

Data used in the present study were collected within the framework of the project EMPHASIS (“Non-specific physical symptoms in relation to actual and perceived exposure to EMF and the underlying mechanisms”), funded by The Netherlands Organization for Health Research and Development (ZonMw) (project number: 85100002). The study sponsor had no involvement in study design, collection, analysis, writing and interpretation of the data and in the decision to submit the study for publication. The SaP questionnaire was developed after request of the department of health assessment studies in the aftermath of disasters of the National Institute for Public Health and the Environment (RIVM, Bilthoven, The Netherlands)

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

We are grateful to the respondents and their GPs for their cooperation. We would like to thank Dr. Linda Grievink (RIVM) who took the initiative to request the development of a tool to measure symptoms and associated perceptions, Dr. Iris Nijrolder for her contribution during the questionnaire development stage and also Petra ten Veen and Elsbeth de Leeuw-Stravers (data management of the GP registrations). In addition, we would like to thank our colleagues at NIVEL for their feedback on the drafts of the paper, especially Renée Bouwman and Dr. Daphne Jansen.

REFERENCES

- American Psychiatric Association, 2013 American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (DSM-5) American Psychiatric Publ. (2013)
- Baliatsas et al., 2014 C. Baliatsas, I. Van Kamp, M. Hooiveld, J. Yzermans, E. Lebrecht Comparing non-specific physical symptoms in environmentally sensitive patients: prevalence, duration, functional status and illness behavior *J. Psychosom. Res.*, 76 (2014), pp. 405–413
- Baliatsas et al., 2015 C. Baliatsas, J. Bolte, J. Yzermans, G. Kelfkens, M. Hooiveld, E. Lebrecht, et al. Actual and perceived exposure to electromagnetic fields and non-specific physical symptoms: an epidemiological study based on self-reported data and electronic medical records *Int. J. Hyg. Environ. Health*, 218 (2015), pp. 331–344
- Barsky, 1981 A. J. Barsky Hidden reasons some patients visit doctors *Ann. Intern. Med.*, 94 (1981), pp. 492–498
- Barsky and Borus, 1999 A. J. Barsky, J. F. Borus Functional somatic syndromes *Ann. Intern. Med.*, 130 (1999), pp. 910–921
- Biermans et al., 2008 M. C. Biermans, D. H. De Bakker, R. A. Verheij, J. V. Gravestein, M. W. van der Linden, P. F. de Vries Robbé Development of a case-based system for grouping diagnoses in general practice *Int. J. Med. Inf.*, 77 (2008), pp. 431–439
- Broadbent et al., 2006 E. Broadbent, K. J. Petrie, J. Main, J. Weinman The brief illness perception questionnaire *J. Psychosom. Res.*, 60 (2006), pp. 631–637
- Carson et al., 2015 A. J. Carson, J. Stone, C. H. Hansen, R. Duncan, J. Cavanagh, K. Matthews, et al. Somatic symptom count scores do not identify patients with symptoms unexplained by disease: a prospective cohort study of neurology outpatients *J. Neurol. Neurosurg. Psychiatry*, 86 (2015), pp. 295–301
- Creed et al., 2012 F. H. Creed, I. Davies, J. Jackson, A. Littlewood, C. Chew-Graham, B. Tomenson, et al. The epidemiology of multiple somatic symptoms *J. Psychosom. Res.*, 72 (2012), pp. 311–317
- Das-Munshi et al., 2006 J. Das-Munshi, G. J. Rubin, S. Wessely Multiple chemical sensitivities: a systematic review of provocation studies *J. Allergy Clin. Immunol.*, 118 (2006), pp. 1257–1264
- de Vroege et al., 2015 L. de Vroege, W. H. Emons, K. Sijtsma, R. Hoedeman, C. M. van der Feltz-Cornelis Validation of the 4DSQ somatization subscale in the occupational health care setting as a screener *J. Occup. Rehab.*, 25 (2015), pp. 105–115
- Dozeman et al., 2011 E. Dozeman, D. J. van Schaik, H. W. van Marwijk, M. L. Stek, H. E. van der Horst, A. T. Beekman The center for epidemiological studies depression scale (CES-D) is an adequate screening instrument for depressive and anxiety disorders in a very old population living in residential homes *Int. J. Geriatr. Psychiatry*, 26 (2011), pp. 239–246
- Engel and Katon, 1999 C. C. Engel, W. J. Katon Population and need-based prevention of unexplained symptoms in the community Institute of Medicine, Strategies to Protect the Health of Deployed U.S. Forces: Medical Surveillance, Record Keeping, and Risk Reduction National Academy Press, Washington, DC (1999), pp. 173–212
- Foster et al., 2008 N. E. Foster, A. Bishop, E. Thomas, C. Main, R. Horne, J. Weinman, et al. Illness perceptions of low back pain patients in primary care: what are they, do they change and are they associated with outcome? *Pain*, 136 (2008), pp. 177–187

- Frosthalm et al., 2007L. Frosthalm, E. Oernboel, K.S. Christensen, T. Toft, F. Olesen, J. Weinman, *et al.* Do illness perceptions predict health outcomes in primary care patients? A 2-year follow-up study *J. Psychosom. Res.*, 62 (2007), pp. 129–138
- Gavrilescu et al., 2015M. Gavrilescu, K. Demnerová, J. Aamand, S. Agathos, F. Fava Emerging pollutants in the environment: present and future challenges in biomonitoring, ecological risks and bioremediation *N. Biotechnol.*, 32 (2015), pp. 147–156
- Gill et al., 2007S.C. Gill, P. Butterworth, B. Rodgers, A. Mackinnon Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population *Psychiatry Res.*, 152 (2007), pp. 63–71
- Goldberg, 1973D.P. Goldberg The Detection of Psychiatric Illness by Questionnaire Maudsley Monograph, vol. 21 Oxford University Press, Oxford (1973)
- Hasson et al., 2000F. Hasson, S. Keeney, H. McKenna Research guidelines for the Delphi survey technique *J. Adv. Nurs.*, 32 (2000), pp. 1008–1015
- Henningens et al., 2011P. Henningens, P. Fink, C. Hausteiner-Wiehle, W. Rief Terminology, classification and concepts F. Creed, P. Henningens, P. Fink (Eds.), *Medically Unexplained Symptoms, Somatization and Bodily Distress: Developing Better Clinical Services*, Cambridge University Press (2011), pp. 43–68
- Holi et al., 2003M.M. Holi, M. Marttunen, V. Aalberg Comparison of the GHQ-36, the GHQ-12 and the SCL-90 as psychiatric screening instruments in the Finnish population *Nord. J. Psychiatry*, 57 (2003), pp. 233–238
- International Well-being Group, 2006 International Well-being Group Personal Wellbeing Index-Adult (PWI-A) (fourth ed.) Australian Centre on Quality of Life, Deakin University, Melbourne (2006) <http://www.deakin.edu.au/research/acqol/instruments/wellbeing-index/pwi-adult-english.pdf>
- Kaptein et al., 2004A.A. Kaptein, I.M. van Korlaar, M. Scharloo IPQ-K (2004) <http://www.uibno/ipq/pdf/BIPQ-Dutch.pdf>
- Kline, 2014P. Kline An Easy Guide to Factor Analysis Routledge (2014)
- Koloski et al., 2001N.A. Koloski, N.J. Talley, P.M. Boyce Predictors of health care seeking for irritable bowel syndrome and nonulcer dyspepsia: a critical review of the literature on symptom and psychosocial factors *Am. J. Gastroenterol.*, 96 (2001), pp. 1340–1349
- Körber et al., 2011S. Körber, D. Frieser, N. Steinbrecher, W. Hiller Classification characteristics of the Patient Health Questionnaire-15 for screening somatoform disorders in a primary care setting *J. Psychosom. Res.*, 71 (2011), pp. 42–47
- Kreutzer, 2000R. Kreutzer Idiopathic environmental intolerance: case definition issues *Occup. Med.*, 15 (2000), pp. 511–517
- Kroenke et al., 1998K. Kroenke, R.L. Spitzer, F.V. de Gruy, R.A. Swindle A symptom checklist to screen for somatoform disorders in primary care *Psychosomatics*, 39 (1998), pp. 263–272
- Lamberts and Wood, 1987H. Lamberts, M. Wood International Classification of Primary Care Oxford University Press, Oxford (1987)
- Lee et al., 2011S. Lee, L.M. Yee, T. Adley Psychometric properties of the Chinese 15-item Patient Health Questionnaire in the general population of Hong Kong *J. Psychosom. Res.*, 71 (2011), pp. 69–73
- Leeflang et al., 2008M.M. Leeflang, K.G. Moons, J.B. Reitsma, A.H. Zwinderman Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions *Clin. Chem.*, 54 (2008), pp. 729–737
- Mayou et al., 2005R. Mayou, L.J. Kirmayer, G. Simon, K. Kroenke, M. Sharpe Somatoform disorders: time for a new approach in DSM-VA *Am. J. Psychiatry*, 162 (2005), pp. 847–855
- Noyes et al., 2008R. Noyes, S.P. Stuart, D.B. Watson A reconceptualization of the somatoform disorders *Psychosomatics*, 49 (2008), pp. 14–22
- Nunnally and Bernstein, 1994J.C. Nunnally, I.H. Bernstein Psychometric Theory (third ed.) McGraw-Hill, New York (1994)
- Nuyen et al., 2005J. Nuyen, A.C. Volkers, P.F. Verhaak, F.G. Schellevis, P.P. Groenewegen, G.A. Van Den Bos Accuracy of diagnosing depression in primary care: the impact of chronic somatic and psychiatric co-morbidity *Psychol. Med.*, 35 (2005), pp. 1185–1190
- Okkes et al., 2002I.M. Okkes, S.K. Oskam, H. Lamberts The probability of specific diagnoses for patients presenting with common symptoms to Dutch family physicians *J. Fam. Pract.*, 51 (2002), pp. 31–36

- Page et al., 2006L.A. Page, K.J. Petrie, S.C. WesselyPsychological responses to environmental incidents: a review and a proposed typology*J. Psychosom. Res.*, 60 (2006), pp. 413–422
- Piek et al., 2012E. Piek, W.A. Nolen, K. van der Meer, K.J. Joling, B.J. Kollen, B.W. Penninx, et al.Determinants of (non-)recognition of depression by general practitioners: results of the Netherlands Study of Depression and Anxiety*J. Affect. Disord.*, 138 (2012), pp. 397–404
- Priede and Farrall, 2011C. Priede, S. FarrallComparing results from different styles of cognitive interviewing: ‘Verbal Probing’ vs. ‘Thinking Aloud’*Int. J. Soc. Res. Methodol.*, 14 (2011), pp. 271–287
- Robbins et al., 1997J.M. Robbins, L.J. Kirmayer, S. HemamiLatent variable models of functional somatic distress*J. Nerv. Ment. Dis.*, 185 (1997), pp. 606–615
- Rolstad et al., 2011S. Rolstad, J. Adler, A. RydénResponse burden and questionnaire length: is shorter better? A review and meta-analysis*Value Health*, 14 (2011), pp. 1101–1108
- Röösli et al., 2010M. Röösli, P. Frei, E. Mohler, K. HugSystematic review on the health effects of exposure to radiofrequency electromagnetic fields from mobile phone base stations*Bull. World Health Organ.*, 88 (2010), pp. 887–896
- Rosendal et al., 2003M. Rosendal, F. Bro, P. Fink, K.S. Christensen, F. OlesenDiagnosis of somatisation: effect of an educational intervention in a cluster randomised controlled trial*Br. J. Gen. Pract.*, 53 (2003), pp. 917–922
- Schmitz et al., 1999N. Schmitz, J. Kruse, C. Heckrath, L. Alberti, W. TressDiagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments*Soc. Psychiatry Psychiatr. Epidemiol.*, 34 (1999), pp. 360–366
- Schreuders et al., 2007B. Schreuders, H. van Marwijk, J. Smit, F. Rijmen, W. Stalman, P. van OppenPrimary care patients with mental health problems: outcome of a randomised clinical trial*Br. J. Gen. Pract.*, 57 (2007), pp. 886–891
- Terluin et al., 2006B. Terluin, H.W. van Marwijk, H.J. Adèr, H.C. de Vet, B.W. Penninx, M.L. Hermens, et al.The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization*BMC Psychiatry*, 6 (2006), p. 34
- Tomenson et al., 2013B. Tomenson, C. Essau, F. Jacobi, K.H. Ladwig, K.A. Leiknes, R. Lieb, et al.Total somatic symptom score as a predictor of health outcome in somatic symptom disorders*Br. J. Psychiatry*, 203 (2013), pp. 373–380
- van den Berg et al., 2005B. van den Berg, L. Grievink, R.K. Stellato, C.J. Yzermans, E. LebrechtSymptoms and related functioning in a traumatized community*Arch. Intern. Med.*, 165 (2005), pp. 2402–2407
- van den Berg, 2007B. van den BergPhysical symptoms that are frequently unexplained among survivors of the Enschede fireworks disasterDoctoral dissertation. Utrecht University (2007)
- Van der Zee and Sanderman, 1993K.I. Van der Zee, R. SandermanMeasuring General Health with the RAND-36: A Manual. Dutch VersionNoordelijk Centrum voor Gezondheidsvraagstukken, Groningen, The Netherlands (1993)
- van der Windt et al., 2008D.A.W.M. van der Windt, Z.M. Dunn, M.N. Spies-Dorgelo, C.D. Mallen, A.H. Blanckstein, W.A.B. StalmanImpact of physical symptoms on perceived health in the community*J. Psychosom. Res.*, 64 (2008), pp. 265–274
- Verhaak et al., 2006aP.F.M. Verhaak, S.A. Meijer, A.P. Visser, G. WoltersPersistent presentation of medically unexplained symptoms in general practice*Fam. Pract.*, 23 (2006), pp. 414–420
- Verhaak et al., 2006bP.F. Verhaak, F.G. Schellevis, J. Nuijen, A.C. VolkersPatients with a psychiatric disorder in general practice: determinants of general practitioners’ psychological diagnosis*Gen. Hosp. Psychiatry*, 28 (2006), pp. 125–132
- Vinck et al., 1998J. Vinck, G. Wels, M. Arickx, S. VinckAssessment of optimism*Gedrag Gezond.*, 26 (1998), pp. 79–90 (in Dutch)
- Visser et al., 1978P. Visser, W.F. Hofman, A. Kumar, R. Cluydts, I.P.F. de Diana, P. Marchant, et al.Sleep and mood: measuring the sleep qualityR.G. Priest, Pletscher, A., Ward, J. (Eds.), *Proc. of the North Eur. Symp. on Sleep Research*, MTP Press Limited, Basel, Lancaster (1978), pp. 135–14

- Wessely et al., 1999S. Wessely, C. Nimnuan, M. SharpeFunctional somatic syndromes: one or many?*Lancet*, 354 (1999), pp. 936–939
- Willis, 200G.B. WillisCognitive Interviewing: A Tool for Improving Questionnaire DesignSage Publications (2004)
- Witthöft et al., 2013M. Witthöft, W. Hiller, N. Loch, F. JasperThe latent structure of medically unexplained symptoms and its relation to functional somatic syndromes*Int. J. Behav. Med.*, 20 (2013), pp. 172–183
- WHO, 2009World Health OrganizationGlobal Health Risks: Mortality and Burden of Disease Attributable to Selected Major RisksWorld Health Organization, Geneva (2009)
- WONCO Classification Committee, 1995WONCA Classification CommitteeAn international glossary for general/family practice*Fam. Pract.*, 12 (1995), pp. 341–369
- Yzermans and Gersons, 2002C.J. Yzermans, B.P.R. GersonsThe Chaotic Aftermath of an Airplane Crash in AmsterdamSpringer US (2002)
- Zijlema et al., 2013W.L. Zijlema, R.P. Stolk, B. Lowe, W. Rief, P.D. White, J.G.M. RosmalenHow to assess common somatic symptoms in large-scale studies: a systematic review of questionnaires*J. Psychosom. Res.*, 74 (2013), pp. 459–468
- Zijlema et al., 2015W.L. Zijlema, D.W. Morley, R.P. Stolk, J.G.M. RosmalenNoise and somatic symptoms: A role for personality traits?*J. Hyg. Environ. Health*, 218 (2015), pp. 543–54

TABLES AND FIGURES

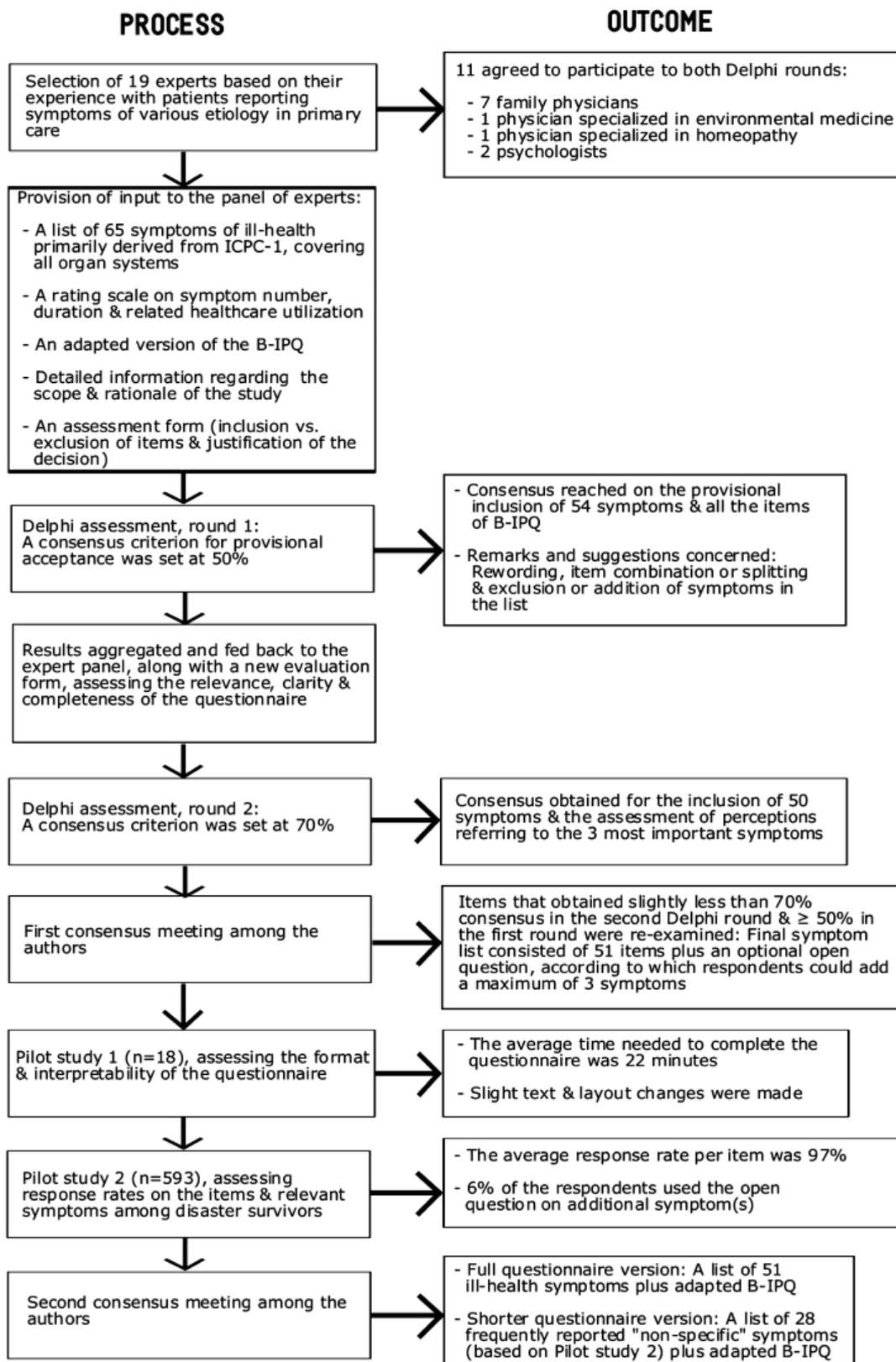


Fig. 1. Overview of the questionnaire development stage.

Table 1

Basic demographic and symptom characteristics in the sample.^a

Characteristics	Sample (n = 5933)
Age (%)	
18–24	5.8
25–44	30.4
45–64	39.5
65–74	13.0
75+	11.1
Mean age (SD)	52.2 (17.3)
Female gender (%)	58.4
Education ^b (%)	
Lower	24.0
Middle	44.3
Higher	31.6
Foreign background (%)	12.6
GP-registered somatoform disorders (%)	1.2
>2 GP-registered non-specific symptoms (%)	4.0
GP-registered mental disorders (%)	7.1
GHQ-12 score >3 (%)	21.3
Mean score on NSS features (SD)	
Number (range: 0–28)	6.1 (4.8)
Duration (range 0–112)	14.8 (14.9)
Related healthcare utilization (range: 0–28)	1.8 (2.9)
Mean score on symptom perceptions per item (SD) (range: 0–10)	
Consequences	5.7 (2.6)
Timeline	6.9 (2.9)
Personal control	5.5 (3.1)
Treatment control ^c	4.1 (2.5)
Perceived need for treatment ^d	4.9 (3.4)
Concern	4.4 (3.1)
Comprehensibility	2.9 (3.0)
Emotional response	4.5 (3.1)
Total mean score on symptom perceptions (SD) (range 0–70)	
Total score on symptom perceptions (t) ^c	37.1 (11.0)
Total score on symptom perceptions (n.t) ^d	32.9 (12.8)

Abbreviation: NSS, non-specific symptoms.

^a Valid cases.

^b Lower: no education or primary school or lower secondary education; middle: intermediate vocational or intermediate general secondary or higher general secondary education; higher: higher vocational or university education.

^c Participants following treatment for their most important symptom.

^d Participants not following treatment for their most important symptom.

Table 2
Factor analysis with different factor loadings for the assessment method "symptom number".

Symptoms	Extracted factors					
	F1	F2	F3	F4	F5	F6
Feeling down/depressed	0.77					
Acute (intense) stress or crisis	0.72					
Feeling anxious/nervous/tense	0.72					
Feeling irritable/angry	0.68					
Memory- or concentration problems	0.51					
Sleep problems	0.43					
Fatigue/tiredness	0.36					
Arm/elbow/hand/wrist symptoms		0.63				
Back problems		0.62				
Neck- or shoulder symptoms		0.62				
Leg/hip/knee/foot symptoms		0.6				
Pain in muscles		0.46				
Tingling of fingers, feet or toes		0.41				
Abdominal/stomach pain			0.76			
Nausea			0.74			
Diarrhea or constipation			0.64			
Dizziness or feeling light-headed			0.38			
Headache			0.37			
Pain or pressure in chest				0.69		
Heart palpitations/awareness				0.63		
Shortness of breath or wheezing				0.56		
Cough					0.75	
Nasal symptoms					0.7	
Ear symptoms						0.57
Eye irritation						0.49
Hypersensitivity to light or noise						0.48
Skin problems						0.39
Weight change ^a						
Explained variance (%)		19.2	6.4	4.9	4.5	4.2
Total explained variance (%)	42.9%					
Cronbach's α		0.76	0.63	0.62	0.56	0.47
Total (all items included) Cronbach's α	0.83					0.36

^a Note: Item excluded due to loading <0.35.

Table 3
Factor analysis with different factor loadings for the assessment method "symptom duration".

Symptoms	Extracted factors					
	F1	F2	F3	F4	F5	F6
Feeling down/depressed	0.8					
Acute (intense) stress or crisis	0.77					
Feeling anxious/nervous/tense	0.76					
Feeling irritable/angry	0.73					
Memory- or concentration problems	0.54					
Sleep problems	0.43					
Fatigue/tiredness	0.39					
Arm/elbow/hand/wrist symptoms		0.7				
Leg/hip/knee/foot symptoms		0.65				
Neck- or shoulder symptoms		0.64				
Back problems		0.63				
Pain in muscles		0.55				
Tingling of fingers, feet or toes		0.46				
Abdominal/stomach pain			0.73			
Nausea			0.72			
Diarrhea or constipation			0.69			
Headache			0.44			
Dizziness or feeling light-headed ^a						
Pain or pressure in chest				0.75		
Heart palpitations/awareness				0.67		
Shortness of breath or wheezing				0.5		
Cough					0.79	
Nasal symptoms					0.63	
Ear symptoms						0.63
Eye irritation						0.52
Hypersensitivity to light or noise						0.36
Skin symptoms						0.36
Weight change ^a						
Explained variance (%)		22.5	6.3	4.9	4.5	4.2
Total explained variance (%)	46.2%					
Cronbach's α		0.81	0.71	0.62	0.6	0.43
Total (all items included) Cronbach's α	0.86					0.39

^a Note: Item excluded due to loading <0.35.

Table 4
Factor analysis with different factor loadings for the assessment method "GP-presentation".

Symptoms	Extracted factors						
	F1	F2	F3	F4	F5	F6	F7
Feeling down/depressed	0.82						
Acute (intense) stress or crisis	0.81						
Feeling anxious/nervous/tense	0.74						
Feeling irritable/angry	0.73						
Sleep problems	0.55						
Memory- or concentration problems	0.52						
Fatigue/tiredness ^a							
Back problems		0.67					
Arm/elbow/hand/wrist symptoms		0.66					
Neck- or shoulder symptoms		0.65					
Leg/hip/knee/foot symptoms		0.62					
Tingling of fingers, feet or toes		0.51					
Pain in muscles			0.49				
Cough			0.82				
Nasal symptoms			0.66				
Shortness of breath or wheezing			0.58				
Abdominal/stomach pain				0.78			
Diarrhea or constipation				0.75			
Nausea				0.72			
Pain or pressure in chest					0.82		
Heart palpitations/awareness					0.74		
Hypersensitivity to light or noise						0.66	
Headache						0.48	
Ear symptoms						0.46	
Dizziness or feeling light-headed						0.39	
Skin symptoms							0.7
Eye irritation							0.66
Weight change ^a							
Explained variance (%)		20.0	6.9	5.2	4.9	4.3	4.0
Total explained variance (%)	49.0%						
Cronbach's α		0.81	0.69	0.54	0.62	0.61	0.41
Total (all items included) Cronbach's α	0.83						0.26

^a Note: Item excluded due to loading <0.35.

Table 5
The associations of symptom subscales with corresponding perceptions and conceptually related measures.

	Symptom number	Symptom duration	Symptom-related healthcare utilization
General health	-0.49**	-0.54**	-0.46**
Low sleep quality	0.48**	0.50**	0.38**
Quality of life	-0.47**	-0.48**	-0.34**
Psychological distress	0.49**	0.47**	0.35**
Low perceived control	0.29**	0.30**	0.23**
Symptom perceptions (t) ^a	0.36**	0.44**	0.34**
Symptom perceptions (n.t) ^b	0.39**	0.47**	0.41**

Note: ** $p < 0.001$.

^a Participants following treatment for their most important symptom.

^b Participants not following treatment for their most important symptom.

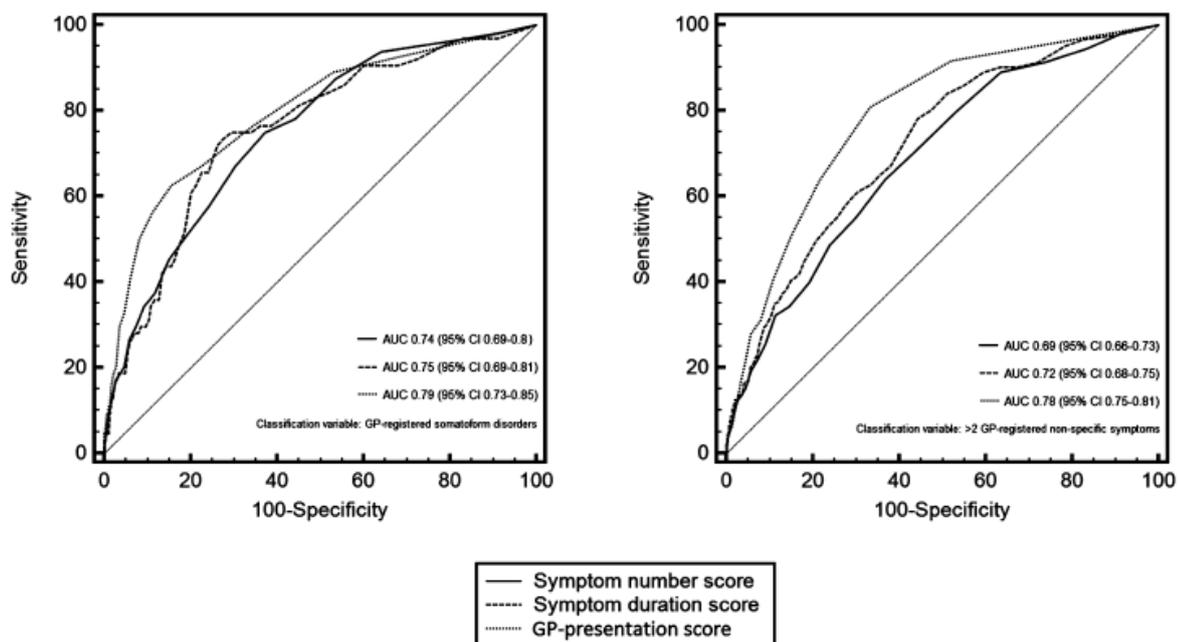


Fig. 2. Receiving operator characteristic curves for the total SaP scores as indicators of somatoform disorders and multiple NSS in general practice.

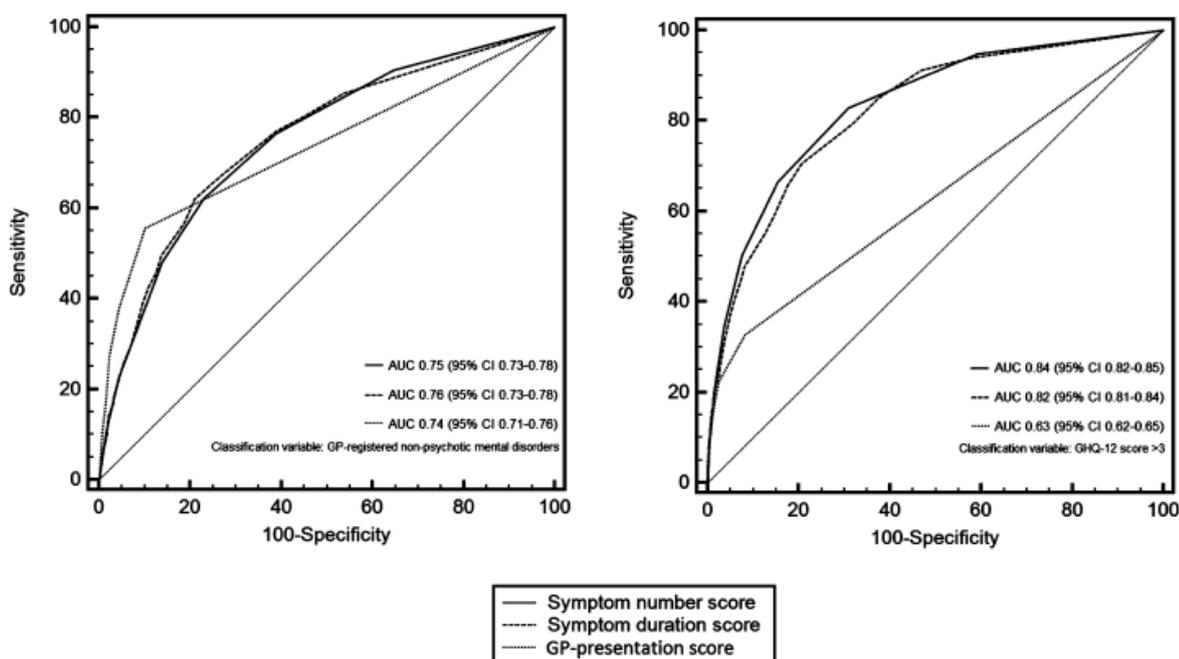


Fig. 3. Receiving operator characteristic curves for the scores of the "psychological" factor of the SaP, as indicators of mental disorders.

Table 6

Sensitivity and specificity for the sum scores of the symptom subscales; the two cut-off points with the highest correct classification rates are presented.

Classification variable	SaP score cut-off points (sensitivity, specificity)		
	Symptom number ^a	Symptom duration ^a	GP-presentation ^a
GP-registered somatoform disorders	≥6 (80.3%, 56.6%) ≥7 (77.5% , 70%)	≥12 (81%, 54.1%) ≥18 (75% , 70%)	≥2 (77.3%, 65.2%) ≥4 (63.6% , 84%)
Multiple (>2) GP-registered non-specific symptoms	≥6 (73.9% , 54.2%) ≥7 (65.2%, 61.3%)	≥12 (78.8% , 54.9%) ≥18 (60.1%, 70.5%)	≥2 (79.5% , 66.4%) ≥3 (62.6%, 77.9%)
Classification variable	SaP score cut-off points (sensitivity, specificity)		
	Symptom number ^b	Symptom duration ^b	GP-presentation ^b
GP-registered mental disorders	≥2 (78.6%, 59%) ≥3 (65.1% , 75%)	≥3 (81%, 54.5%) ≥7 (63% , 78%)	≥1 (56.2% , 89.2%) ≥2 (38.3%, 95.5%)
GHQ-12 score (>3)	≥2 (84.4%, 67.2%) ≥3 (69.1% , 83.1%)	≥4 (80%, 67.5%) ≥5 (71.1% , 78.6%)	≥1 (34% , 91.3%) ≥2 (23.4%, 97.4%)

Note: Optimal cut-off point in bold.

^a Sum score including all symptom items.

^b Sum score on the "psychological" symptom cluster.