


# BMJ Open Identifying persistent somatic symptoms in electronic health records: exploring multiple theory-driven methods of identification

Willeke M Kitselaar <sup>1,2</sup>, Mattijs E Numans,<sup>2</sup> Stephen P Sutch,<sup>2,3</sup> Ammar Faiq,<sup>2</sup> Andrea WM Evers,<sup>1,4</sup> Rosalie van der Vaart<sup>1</sup>

**To cite:** Kitselaar WM, Numans ME, Sutch SP, *et al*. Identifying persistent somatic symptoms in electronic health records: exploring multiple theory-driven methods of identification. *BMJ Open* 2021;**11**:e049907. doi:10.1136/bmjopen-2021-049907

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-049907>).

Received 04 February 2021  
Accepted 05 September 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Drs Willeke M Kitselaar;  
w.m.kitselaar@fsw.leidenuniv.nl

## ABSTRACT

**Objective** Persistent somatic symptoms (PSSs) are defined as symptoms not fully explained by well-established pathophysiological mechanisms and are prevalent in up to 10% of patients in primary care. The present study aimed to explore methods to identify patients with a recognisable risk of having PSS in routine primary care data.

**Design** A cross-sectional study to explore four identification methods that each cover part of the broad spectrum of PSS was performed. Cases were selected based on (1) PSS-related syndrome codes, (2) PSS-related symptom codes, (3) PSS-related terminology and (4) Four-Dimensional Symptom Questionnaire scores and all methods combined.

**Setting** Coded electronic health record data were extracted from 76 general practices in the Netherlands.

**Participants** Patients who were registered for at least 1 year during 2014–2018, were included (n=169 138).

**Outcome measures** Identification methods were explored based on (1) PSS sample sizes and demographics, (2) presence of chronic conditions and (3) healthcare utilisation (HCU) variables. Overlap between methods and practice specific differences were examined.

**Results** The percentage of cases identified varied between 0.3% and 7.0% across the methods. Over 58.1% of cases had chronic physical condition(s) and over 33.8% had chronic mental condition(s). HCU was generally higher for cases selected by any method compared with the total cohort. HCU was higher for method B compared with the other methods. In 26.7% of cases, cases were selected by multiple methods. Overlap between methods was low.

**Conclusions** Different methods yielded different patient samples which were general practice specific. Therefore, for the most comprehensive data-based selection of PSS cases, a combination of methods A, C and D would be recommended. Advanced (data-driven) methods are needed to create a more sensitive algorithm for identifying the full spectrum of PSS. For clinical purposes, method B could possibly support screening of patients who are currently missed in daily practice.

## INTRODUCTION

In the general population, approximately 20% of adults experience persistent or

## Strengths and limitations of this study

- Our study provides insight into the value and limitations of multiple methods to identify patients with recognisable risk of persistent somatic symptoms (PSS) in routine care data.
- Combining our explored methods decreases the likelihood of missing patients with PSS in clinical practice.
- Not excluding ‘explained’ chronic conditions provides insight in PSS in the general population and gauges the immensity of the problem for healthcare.
- Our large dataset from multiple general practices in a highly versatile area increases ecological validity and generalisability of the results.
- We drew on electronic medical records that rely on accuracy of registration, which is not flawless but represents the robustness of clinical practice.

recurring disabling physical symptoms.<sup>1–4</sup> These physical symptoms are generally not fully explained by established biomedical pathology, and cannot be fully attributed to objectively determined anatomical or functional disease severity.<sup>4–7</sup> This is often the case for both patients with well-documented diseases such as cancer<sup>8,9</sup> and cardiovascular disease,<sup>10,11</sup> as well as for patients with so-called medically ‘unexplained’ physical symptoms.<sup>12–15</sup> ‘Unexplained’ symptoms account for up to 50% of all primary care consultations in western populations.<sup>16–18</sup> While most of these symptoms are self-limiting, they persist in 2.5%–10% of cases.<sup>12,13,15</sup> Due to conceptual and domain specific differences, these conditions have been described using a wide range of labels, including medically unexplained physical symptoms, functional disorders, somatisation and somatic symptom disorder (SSD).<sup>15,19,20</sup> Alternatively, patients with a specific set of symptoms are classified as having a syndrome (eg, chronic fatigue



syndrome (CFS), fibromyalgia (FM) or irritable bowel syndrome (IBS)). In general, symptoms are often classified into one of the ‘unexplained’ categories, based on exclusion of physical conditions with well-documented biomedical pathology.<sup>21</sup> In this paper, the term ‘persistent somatic symptoms’ (PSS) is used, since recent research has found that this term is generally preferred over other terms.<sup>22</sup> Moreover, the term PSS is in line with recent advances in the field, specifically related DSM and ICD classifications, which no longer require exclusion based on the presence of a medical condition but instead focus on positive symptomatology (eg, the presence and burden of symptoms).<sup>20</sup>

This broad spectrum of PSS, whether or not accompanied by a physical condition, either directly or indirectly affect a major part of the population and are generally accompanied by an increasing burden of disease for both the patient and healthcare systems.<sup>23</sup> Although widely discussed, consensus on classification, diagnostic procedure and treatment approaches is still lacking.<sup>24–27</sup> This impedes early recognition and proactive clinical intervention of patients at risk of developing persistent problems, resulting in inappropriate and relatively high healthcare utilisation (HCU) and costs.<sup>28–29</sup> Particularly in primary care, which serves as a gatekeeper for healthcare in many (western) countries,<sup>30–32</sup> earlier recognition is desirable as it could help to prevent unnecessary referrals and could enable the initiation of proactive interventions, aiming to avoid problems becoming permanent or other adverse health consequences.

Recent advancements in data science have shown that routine primary care data can be responsibly used for epidemiological research,<sup>33–34</sup> predictive modelling<sup>35</sup> and population health management purposes.<sup>36</sup> The use of routine primary care data for research on PSS, however, is currently hampered by ambiguous registration of diagnoses in the domain of PSS,<sup>37–40</sup> which has led to individual general practitioner (GP) and/or general practices recording PSS differently. Nonetheless, several methods for identifying patients with PSS in the electronic medical records (EMR) of patients in primary care have been explored in previous research.<sup>26–39–40</sup> Yet, none of those seem fully satisfactory due to the need for additional diagnostics, limited sensitivity and exclusion of patients with mental or physical conditions.

This study aims to gain better insight into the most comprehensive data-based options for identifying the full spectrum of patients carrying the risk of having PSS in routine primary care data. A more comprehensive method of data-based identification of patients with PSS will make it possible to feedback an individual risk score to physicians that might help to increase awareness of PSS, but it might also improve future research on specific interventions. We explored the differences between previously used identification methods, focusing on (A) PSS-related syndromes<sup>26–41</sup> and (B) PSS-related symptomatology<sup>26–39–40</sup> and adding new options. First, findings from a recent survey among GPs undertaken by our group,<sup>27</sup>

indicated the use of (C) PSS-related terminology in free-text areas. Second, we found results from the validated Four-Dimensional Symptom Questionnaire (4DSQ), which screens for PSS,<sup>42</sup> to be registered in Dutch primary care health records. The 4DSQ is most likely to be administered by the mental health nurse practitioner, when patients are referred by their GP for psychological complaints. Recorded results of the (D) somatic symptoms subscale of the 4DSQ were included as another method for identifying PSS. Lastly, all methods (A–D) were combined. For all methods, outcomes relating to sample characteristics, presence of diagnosed chronic conditions and HCU were assessed.<sup>43–45</sup>

## METHOD

### Study design

In the Netherlands, all residents are enlisted with a GP in their neighbourhood and general practice care is covered by the mandatory health insurance. In the Dutch healthcare system, the GP acts as the gatekeeper to hospital services. Routine EMR data from primary care are a valuable source of information for research, healthcare organisation and population health as well as quality management.

For this study, we reused anonymously extracted routine care data<sup>46</sup> from 76 general practice centres that were affiliated with the Extramural Leiden University Medical Center Academic Network (ELAN) primary care network, the Netherlands. All practices were located in the greater Leiden and The Hague area.

For the current study, coded EMR patient data were used, including demographics, enrolment information, consultation types and dates, symptoms and diagnoses coded according to the International Classification of Primary Care (ICPC)<sup>47</sup> in the episode and contact registration (in the Netherlands, ICPC-I is used with nationally relevant adjustments<sup>48</sup>); textual episode descriptions; coded information of laboratory tests, dates and results; Anatomical Therapeutic Chemical (ATC) classification<sup>49</sup> of medications and prescription dates; and coded correspondence with other healthcare professionals and dates. For this paper, the Strengthening the Reporting of Observational Studies in Epidemiology cross sectional reporting guidelines was used.<sup>50</sup>

### Study population

All patients enrolled for at least 1 year with one of the affiliated general practices between January 2014 and December 2018, who were born before 1989 (25 years of age) and born after 1914 (100 years of age) were included in the study. Length of enrolment was primarily determined on quarterly payment data. When payment data were unavailable or enrolment and unenrolment dates indicated that the patient was enrolled for a longer period, the enrolment and unenrolment dates were used.

### Identification methods

While data-driven research may circumvent healthcare professionals’ difficulties with identifying patients at

**Table 1** Description of methods for PSS identification

	Method A	Method B	Method C	Method D
General PSS-related criteria	Irritable bowel syndrome, fibromyalgia and chronic fatigue syndrome	Frequent consultation for multiple PSS-related symptoms	Reported PSS-related terminology*	Somatic symptom subscale of the 4DSQ†
Criteria translated to EMR data	ICPC codes for chronic fatigue syndrome (A04.01), fibromyalgia (L18.01) and irritable bowel syndrome (D93)	Symptoms based on the 'Robbins list'.‡ The symptoms have been linked to ICPC-codes	Terminology is based on a cross-search of the data with ICPC codes from methods A and B	The 4DSQ is a validated questionnaire available in EMRs of Dutch GPs
Duration/cut-off	At least 6 months, or at least two contact registrations.	At least six registrations with one or more relevant ICPC codes within a 6-month period	At least 6 months	A score of $\geq 20$ on the somatic symptom subscale of the 4DSQ
Registration type	Based on episode and contact registration	Based on contact registration	Based on description in the episode registration	Laboratory test results

The table was created by the author and permission for reuse is granted to *BMJ Open*.

\*Examples of included terms: somatisation, psychosomatic, central sensitisation, atypical low back pain, stress related pain, interstitial cystitis, extreme fatigue, tension headache: good CT, functional.

†Four-dimensional symptom questionnaire.<sup>42</sup>

‡Robbins list: Back pain, joint pain, extremity pain, headaches, fatigue/weakness, sleep disturbance, difficulty concentrating, loss of appetite, weight change, restlessness, thoughts slower, chest pain, shortness of breath, palpitations, dizziness, lump in throat, numbness, nausea, loose bowels, gas/bloating, constipation, abdominal pain.<sup>52</sup>

4DSQ, Four-Dimensional Symptom Questionnaire; EMR, electronic medical records; GP, general practitioner; ICPC, International Classification of Primary Care; PSS, persistent somatic symptoms.

risk of PSS, the debate on definitions and terminology remains.<sup>6</sup> While some earlier developed definitions required physicians to classify patients primarily on the basis of exclusion of any medical explanation for the symptoms, recently developed classifications favour focusing on common behavioural similarities related to PSS instead.<sup>6 20 51</sup> The latter explicitly do not exclude patients with known medical illnesses. In line with these recent developments, our patient group is defined as having PSS when their complaints are not fully explained by established biomedical pathology. However, these symptoms and the accompanying behaviour, can also exist alongside other chronic physical conditions that are explained by established biomedical pathology. To reach the aim of our study, four methods were included (see [table 1](#)). Two methods (A and B) were based on identification methods used in previous studies, one was derived from these two existing methods (C), and one was based on expert knowledge about the available data in the ELAN-database (D). Method A identifies patients with CFS, FM and IBS based on their available ICPC codes (codes for CFS and FM are specific to the Dutch ICPC system)<sup>26 41</sup>; method B identifies patients with PSS-related symptoms which were extracted from a latent class analysis on symptoms highly prevalent in patients with PSS and has been previously used in research<sup>39 40 52</sup>; method C identifies patients based on PSS-related terminology in the episode description (the episode description is adjustable for GPs; that

is, in case a GP registers A04.01, this automatically gives the description 'CFS', but the description can be adjusted to any term the GP prefers. Our available data were systematically searched by cross-checking ICPC codes and related descriptions),<sup>27</sup> and, method D identifies patients based on recorded results of the somatic symptoms subscale of the 4DSQ.<sup>42</sup> Additionally, besides exploring overlap between methods, all four methods were integrated, selecting all patients identified by any of the methods.

### Outcomes

For all methods we calculated the following outcomes: (1) number of patients with PSS and their demographics, (2) presence of chronic physical and mental illness, and (3) HCU. Demographic variables consist of gender and age in 2014. Presence of chronic physical or mental illness was defined based on the list of ICPC codes for chronic conditions, by the Dutch institute of research in primary care (Nederlands Instituut Voor onderzoek van de EersteLijnsgezondheidszorg).<sup>53</sup> HCU was operationalised using consult frequency, number of lab tests, number of prescribed medications and number of referrals.<sup>43–45</sup>

For all HCU frequencies, mean 1-year frequencies were calculated based on the total frequency during the study period, divided by the length of enrolment of the patient. Mean consultation frequency was calculated based on the type of registration in the contact registration per patient, with the exclusion



of administrative contacts (such as making appointments). Lab tests was calculated based on the number of referrals registered for each patient to a laboratory test centre. For the mean number of medications, ATC codes were reduced to four characters which specify up to the pharmacological group a medication belongs to.<sup>49</sup> Each unique pharmacological group registered in the patients EMR was recoded as one medication. Referrals are divided into primary care and secondary care referrals and each unique referral was recorded as one referral per patient.

### Data analysis

Statistical analyses were carried out using R (V.4.0.2).<sup>54</sup> First, patients were selected based on each unique identification method. Descriptive statistics were reported on gender, age, chronic mental and physical conditions, and HCU variables for each method. Second, in order to identify overlap between methods, the percentage of patients being selected by a combination of methods was explored and depicted in a Venn diagram. A graphical display of the number of patients selected by each method per general practice was produced and depicted in a histogram with reported skewness and kurtosis. patient and public involvement

GPs were consulted during the development phase of the research design.

## RESULTS

### Number of patients with PSS and their demographics per identification method

Table 2 shows an overview of the complete cohort which includes 168 682 primary care patients with a mean age of 51.4 (SD=16.4), of whom 52.9% are female. Patients were enlisted in their general practice for an average of 4.6 years (SD=1.0) between January 2014 and December 2018. The 4DSQ, used for identifying patients (method D), was administered and registered for 1102 (0.7%) patients of the total cohort from 2017 to 2019. The

number of cases identified with each method separately varied between 482 (0.3%) for method D and 11 893 (7.0%) for method B. Integrating all methods identified 20 855 cases (12.3%).

### Presence of chronic physical and mental illness per identification method

Cases selected by methods A, B and C are more likely to have a chronic physical condition than the total cohort (60.4% vs  $\geq 66.9\%$ ). Cases selected by method B are most likely to have a chronic physical condition (79.4%). Cases selected by all four methods are more likely to have a chronic mental condition compared with the total cohort (18.2% vs  $\geq 33.8\%$ ). Cases selected by method D are most likely to have a chronic mental condition (60.0%) (table 3).

### HCU per identification method

HCU is generally higher among cases selected by any of the methods, compared with the total cohort. Cases selected by method A, C and D show similar patterns regarding most of the HCU variables. Cases selected by method B show higher average frequencies on the HCU variables compared with cases selected by the other methods, except for primary care referrals, which are similar to cases selected by method D ( $0.17\pm 0.21$  and  $0.17\pm 0.19$ , respectively) (table 4).

### Overlap on outcomes between identification methods

In all, 12.3% of patients are selected by all methods combined, which is less than the cumulative percentage (15.6%) of patients selected by method A, B, C and D separately (see table 2). Thus, 3.3% of the total cohort is selected by more than one method—which is a total of 26.8% of all selected patients. Relative to other methods (all  $\leq 11.6\%$ ), patients are selected by method A and C are most likely to be selected by both methods (34.4%). The likelihood that patients selected by method D are also selected by any other methods is lowest ( $\leq 1.3\%$ ) (see

**Table 2** Number of patients with PSS and their demographics per identification method

	Total cohort n (%) or mean $\pm$ SD	Method A n (%) or mean $\pm$ SD	Method B n (%) or mean $\pm$ SD	Method C n (%) or mean $\pm$ SD	Method D n (%) or mean $\pm$ SD	All methods combined n (%) or mean $\pm$ SD
Patients	169 138 (100.0)	8407 (5.0)	11 893 (7.0)	5574 (3.3)	482 (0.3)	20 855 (12.3)
Female	89 432 (52.9)	6276 (74.7)	8020 (67.4)	4164 (74.7)	340 (70.5)	14 490 (69.5)
Age	51.4 $\pm$ 16.4	48.6 $\pm$ 15.0	57.6 $\pm$ 18.2	49.6 $\pm$ 14.6	42.6 $\pm$ 12.0	53.6 $\pm$ 17.4
Years enrolled	4.6 $\pm$ 1.0	4.6 $\pm$ 0.9	4.5 $\pm$ 1.0	4.6 $\pm$ 0.9	4.7 $\pm$ 0.8	4.6 $\pm$ 1.0

The table was created by the author and permission for reuse is granted to *BMJ Open*

Method A: patients with recorded FM, IBS and/or CFS based on ICPC-codes.

Method B: patients with at least six ICPC codes that correspond to the Robbins list in any 6-month period.

Method C: patients with reported PSS-related terminology.

Method D: patients with  $\geq 20$  points on the somatisation subscale of the 4DSQ.

CFS, chronic fatigue syndrome; 4DSQ, Four-Dimensional Symptom Questionnaire; FM, fibromyalgia; IBS, irritable bowel syndrome; ICPC, International Classification of Primary Care; PSS, persistent somatic symptoms.

**Table 3** Presence of chronic physical and mental conditions per identification method

	Total cohort n (%)	Method A n (%)	Method B n (%)	Method C n (%)	Method D n (%)	All methods combined n (%)
Chronic physical condition	101 868 (60.2)	5624 (66.9)	9446 (79.4)	3957 (70.0)	280 (58.1)	15 313 (73.4)
Chronic mental condition	30 750 (18.2)	2838 (33.8)	4188 (35.2)	2020 (36.2)	289 (60.0)	7080 (33.9)

The table was created by the author and permission for reuse is granted to *BMJ Open*

Method A: patients with recorded FM, IBS, and/or CFS based on ICPC-codes.

Method B: patients with at least six ICPC codes that correspond to the Robbins list in any 6 month period.

Method C: patients with reported PSS-related terminology.

Method D: patients with  $\geq 20$  points on the somatisation subscale of the 4DSQ.

CFS, chronic fatigue syndrome; 4DSQ, Four-Dimensional Symptom Questionnaire; FM, fibromyalgia; IBS, irritable bowel syndrome; ICPC, International Classification of Primary Care; PSS, persistent somatic symptoms.

figure 1 for an elaborate overview of overlap between all the methods).

### Overlap between practices for selecting patients with PSS

We also explored the proportion of cases selected by each general practice ( $n=76$ ). Case selection based on method A and B is most evenly distributed between practices (skewness= $-0.17$  and  $-0.10$ , and kurtosis= $3.01$  and  $3.06$ , respectively). For method C, a moderate left skewed distribution (skewness= $0.99$  and kurtosis= $5.21$ ) shows that many practices contribute a small number of cases and some practices contribute a moderately large number of cases. Method D is highly left skewed (skewness= $2.22$  and kurtosis= $8.29$ ), indicating that many practices contribute no cases or a limited number of cases, while few practices contribute a large number of cases (figure 2).

## DISCUSSION

### Statement of principal findings

This paper describes a comprehensive study on identifying patients with PSS in routine primary care data, in which four different identification methods are explored. The different methods identify a wide range in proportions of cases: from 0.3% selected by method D (ie, recoded 4DSQ

assessments), to 7.0% selected by method B (ie, based on PSS-related ICPC codes and consult frequency). When all separate identification methods are combined, a total of 12.3% of the complete cohort is selected, of which 26.8% is selected by multiple methods. In line with findings from previous studies on PSS, selected cases are more often female (in all methods) and younger (in three out of four methods) compared with the total cohort (which approximates the general population). This study shows that the use of any single method will inevitably lead to underestimation of the number of patients with recognisable risk of PSS recognised.

Detailed analysis of the selected samples reveals some notable results. First, patients selected by any of the methods are generally more likely to have a chronic physical or mental condition, compared with the total cohort. These findings corroborate previous observations that PSS are highly prevalent in patients with chronic physical conditions and emphasises the undesirability of classifying PSS based on exclusion of a chronic physical condition.<sup>14</sup> Furthermore, in line with recommended practice to administer the 4DSQ among patients with psychological complaints,<sup>42</sup> cases selected by method D have a markedly high likelihood of having a chronic mental

**Table 4** Healthcare utilisation per identification method

	Total cohort mean $\pm$ SD	Method A mean $\pm$ SD	Method B mean $\pm$ SD	Method C mean $\pm$ SD	Method D mean $\pm$ SD	All methods combined mean $\pm$ SD
Consult frequency (per year)	4.75 $\pm$ 5.43	7.74 $\pm$ 7.32	11.70 $\pm$ 9.47	8.21 $\pm$ 8.30	8.39 $\pm$ 5.89	9.42 $\pm$ 8.47
Lab tests (per year)	0.28 $\pm$ 0.50	0.50 (0.56)	0.69 $\pm$ 0.88	0.51 $\pm$ 0.68	0.43 $\pm$ 0.56	0.57 $\pm$ 0.75
Medications (per year)	1.82 $\pm$ 1.88	2.65 $\pm$ 2.12	3.64 $\pm$ 2.70	2.81 $\pm$ 2.17	2.62 $\pm$ 1.78	3.10 $\pm$ 2.48
Primary care referrals (per year)	0.09 $\pm$ 0.15	0.14 $\pm$ 0.19	0.17 $\pm$ 0.21	0.15 $\pm$ 0.20	0.17 $\pm$ 0.19	0.15 $\pm$ 0.20
Secondary care referrals (per year)	0.30 $\pm$ 0.42	0.46 $\pm$ 0.54	0.62 $\pm$ 0.61	0.50 $\pm$ 0.57	0.48 $\pm$ 0.45	0.53 $\pm$ 0.57

The table was created by the author and permission for reuse is granted to *BMJ Open*

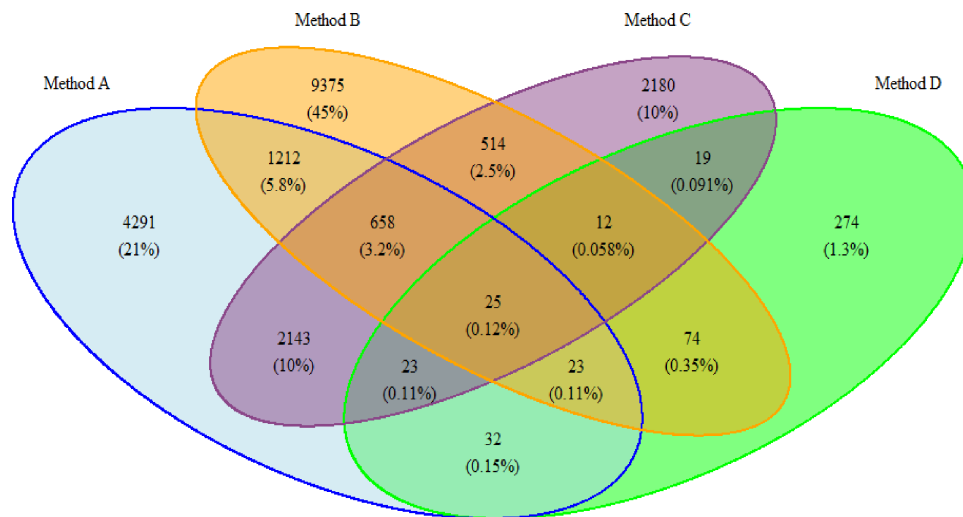
Method A: patients with recorded FM, IBS and/or CFS based on ICPC-codes.

Method B: patients with at least six ICPC codes that correspond to the Robbins list in any 6-month period.

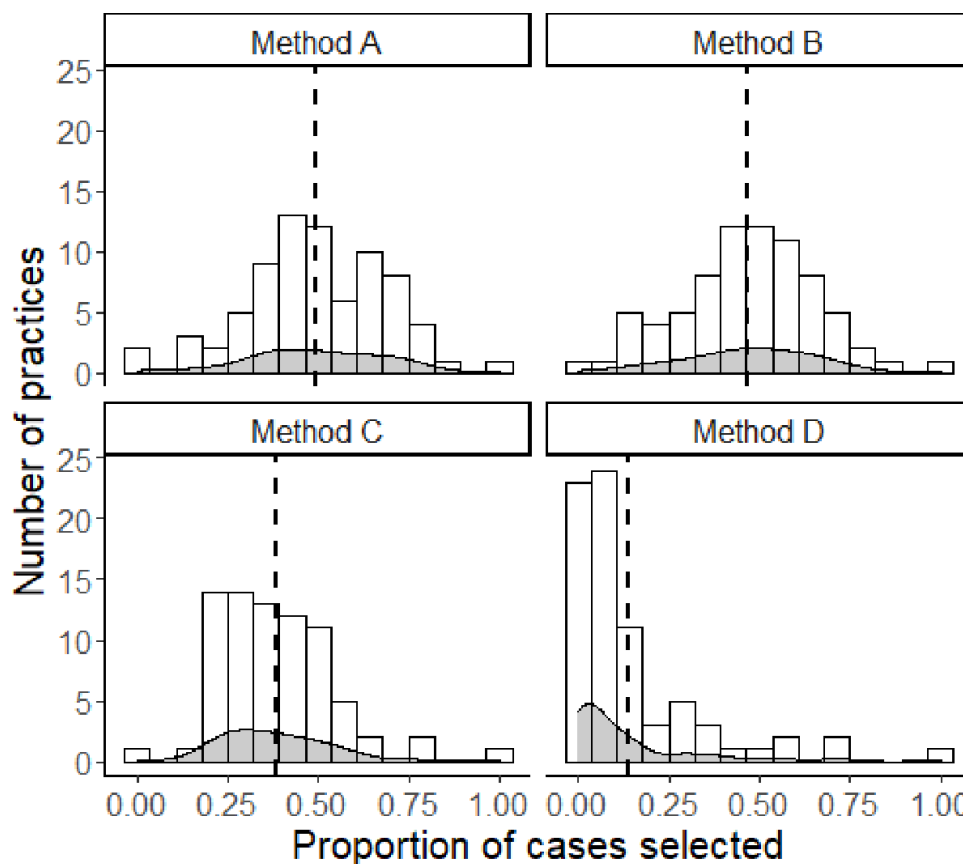
Method C: patients with reported PSS-related terminology.

Method D: patients with  $\geq 20$  points on the somatisation subscale of the 4DSQ.

CFS, chronic fatigue syndrome; 4DSQ, Four-Dimensional Symptom Questionnaire; FM, fibromyalgia; IBS, irritable bowel syndrome; ICPC, International Classification of Primary Care; PSS, persistent somatic symptoms.



**Figure 1** Overlap between selected patient samples per identification method. The figure was created by the author and permission for reuse is granted to BMJ Open. Method A: Patients with recorded FM, IBS and/or CFS based on ICPC-codes. Method B: Patients with at least six ICPC codes that correspond to the Robbins list in any 6-month period. Method C: Patients with reported PSS-related terminology. Method D: Patients with  $\geq 20$  points on the somatisation subscale of the 4DSQ. 4DSQ, Four-Dimensional Symptom Questionnaire; CFS, chronic fatigue syndrome; FM, fibromyalgia; IBS, irritable bowel syndrome; ICPC, International Classification of Primary Care; PSS, persistent somatic symptoms.



**Figure 2** Variation between practices in applying methods of registration. The figure was created by the author and permission for reuse is granted to BMJ Open. Method A: Patients with recorded FM, IBS, and/or CFS based on ICPC-codes. Method B: Patients with at least six ICPC codes that correspond to the Robbins list in any 6-month period. Method C: Patients with reported PSS-related terminology. Method D: Patients with  $\geq 20$  points on the somatisation subscale of the 4DSQ. 4DSQ, Four-Dimensional Symptom Questionnaire; CFS, chronic fatigue syndrome; FM, fibromyalgia; IBS, irritable bowel syndrome; ICPC, International Classification of Primary Care; PSS, persistent somatic symptoms.

condition. Cases selected by method B are most likely to have a chronic physical condition, which indicates that differentiating which complaints are PSS and which complaints are strictly related to a physical condition may be most challenging for cases selected by method B.

Second, HCU is higher for all samples compared with the total cohort. However, HCU spikes and deviates for cases selected by method B, compared with all other cases. Several reasons for this could be plausible. Most notably, that high HCU is expected in this selected group since consultation frequency is part of the inclusion criteria for this method and increased consultation frequency implies higher frequencies for all HCU variables. The higher HCU is also presumed to be related to the heightened likelihood of these cases to have a chronic physical condition. Finally, one could theorise that these patients seek healthcare more frequently because their PSS is not yet recognised. Remarkably, cases selected by method D, among whom chronic mental conditions spike, HCU is not much different from cases selected by method A (ie, CFS, FM and IBS) and C (ie, PSS-related terminology in episode description). Thus, even though cases selected by method D are more likely to have a chronic mental condition, HCU indicates that healthcare seeking behaviour in cases selected by methods A, C and D is more similar than in cases selected by method B.

Finally, our results show a relatively low likelihood that patients are selected by multiple methods. High variance between general practices in using one of the registration methods, especially method D, indicates that the limited overlap is explained by GPs not applying all methods equally. This is consistent with previous research which demonstrated high degrees of discordance between healthcare professionals regarding defining and classifying patients with PSS<sup>27</sup> and an ambiguous coding scheme for PSS.<sup>27 37</sup> From this finding we can conclude that the need for using either a single or multiple methods to identify PSS cases may depend on the aim of the identification. For instance, when calculating exact prevalence rates, using a single method will not be sufficient, since prevalence rates of PSS in the general Dutch population most likely range from 10% to 15%.<sup>55</sup> However, using a single method (eg, method C) may be sufficient to identify risk factors for persistence of PSS, although this should be confirmed by further research.

### Strengths and limitations

The results of this study should be viewed in context of several strengths and limitations. Using multiple methods to identify the PSS patient group, exploring their outcomes on a variety of clinically relevant variables, and exploration of general practice specific variations, results in a very comprehensive review. The use of a large set of routine EMR data from multiple general practices in a highly versatile area of the Netherlands increases ecological validity and generalisability to other populations. Additionally, the inclusion of patients with chronic conditions provides insight in PSS in the general population

and gauges the immensity of the problem for healthcare. Nonetheless, the use of routine care data comes with challenges and limitations. While registration quality is increasingly promoted and improved, it is reliant on many factors specific to the healthcare provider and general practice. Another limitation of this study is the lack of an external validation of the patient group. This seems primarily problematic for method B, which relies on ICPC codes which—while empirically related to PSS<sup>52</sup>—can also be fully explained by biomedical pathology. Notably, the small number of registrations of the 4DSQ in the EMRs reduces the usability of method D. Besides, since some ICPC codes (method A; A04.01 and L18.01), specific (Dutch) terminology (method C), and incorporation of questionnaires evaluating PSS-related problems (method D) are specific to Dutch EMRs, tailored solutions are needed to generalise the results to other countries.

### Implications for clinicians and future research

The current study provides unique insight into the complexity of identifying patients with PSS in routine care data. While the results indicate that current classification and coding of PSS is highly scattered, it shows that a data-based screening of patients with PSS in routine care data is possible. Depending on the desirable goal, single or multiple methods can be used for identification.

From a research perspective, in the first place, replicability of the methods to non-Dutch EMRs should be examined. Second, although the combination of method A, C and D improved earlier approaches towards accurate prevalence rate based on routine primary care data,<sup>43</sup> some steps still need to be taken to get accurate prevalence rates. Nonetheless, combining method A, C and D decreases the portion of patients with PSS that are misclassified as non-PSS, which may enhance the possibilities for data-driven predictive modelling of patients at risk of the broad spectrum of PSS. Finally, while it was beyond the scope of this study to investigate this further, our results regarding practice specific differences in registration may be specifically relevant for identifying GPs who need support for PSS consultations. Especially because previous research shows that a large group of GPs require additional support.<sup>27</sup> Future research should investigate whether the need for support can be linked or tailored to GPs with specific registration methods.

While the present study was primarily methodological, some clinical implications may be relevant to discuss which could enable data-based support for PSS identification (which could promote awareness among GPs regarding PSS-risk). First, clinicians may need to improve registration of the 4DSQ, because this—per suggestion by our expert panel of GPs—is the most likely cause of the limited usability of the method for databased identification. Alternatively, in line with the implications for research, since patients identified with method A, C and D are most likely on the clinicians' 'radar'—that is, they have a clear PSS-related indicator recorded, patients that are currently missed can be screened by method B.



Method B is supported by previous studies which successfully used a similar method for screening routine care data for patients with PSS.<sup>26 39 40</sup> Subsequently, validated questionnaires such as the 4DSQ<sup>42</sup> or the SSD B-criteria scale (SSD-12)<sup>4</sup> can be used to identify those patients selected by method B who need additional attention/proactive intervention. Future research should be aimed at monitoring patients selected based on method B—both towards verifying the effectiveness of this method and whether merely identifying these patients influences the health trajectory of the patients, or gauging if other interventions are needed. Ultimately, all the above could encourage the use of advanced computer systems to support the diagnostic process and subsequent decision making in practice.<sup>56</sup>

## CONCLUSION

In all, the results indicate that the theory-driven methods identify different samples of patients with PSS. A combination of methods A, C and D can form a basis for identifying the full spectrum of patients with PSS, for example, for calculating prevalence rates. Henceforth, additional advanced (data-driven) methods and validation may help to create more sensitive algorithms. These algorithms might be used in clinical practice to increase awareness of physicians on the risk of PSS, thus potentially opening possibilities to proactive interventions. For method B, the relatively high number of cases with chronic physical conditions and HCU indicates the need for additional diagnostics. Further research should focus on investigating whether method B combined with subsequent screening can be a way to identify patients with unidentified PSS who are not yet on the GPs radar.

## Author affiliations

<sup>1</sup>Health, Medical and Neuropsychology, Leiden University Faculty of Social and Behavioural Sciences, Leiden, The Netherlands

<sup>2</sup>Public Health and Primary Care / LUMC-Campus The Hague, Leiden University Medical Center, Den Haag, The Netherlands

<sup>3</sup>Health Policy and Management, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA

<sup>4</sup>Medical Delta, Leiden University, Delft University of Technology & Erasmus University, Leiden / Delft/ Rotterdam, The Netherlands

**Contributors** WMK conducted the study under the guidance of all other authors. WMK and AF preprocessed the data and WMK analysed the data. WMK drafted the manuscript. RvdV reviewed and provided critical comments on all early stage drafts of the manuscript. WMK, RvdV, SPS, AWE and MEN reviewed and the provided critical comments on drafts of the full manuscript. All authors approved the submitted version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not applicable.

**Ethics approval** The ethics committee of Leiden University Medical Center supplied a waiver of ethical approval (G19.045/SB/ib), as ethical approval was not necessary for this study.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No data are available. No additional data available.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iD

Willeke M Kitselaar <http://orcid.org/0000-0003-0914-0168>

## REFERENCES

- Kroenke K, Arrington ME, Mangelsdorff AD. The prevalence of symptoms in medical outpatients and the adequacy of therapy. *Arch Intern Med* 1990;150:1685–9.
- Kroenke K. A practical and evidence-based approach to common symptoms: a narrative review. *Ann Intern Med* 2014;161:579–86.
- Rief W, Martin A. How to use the new DSM-5 somatic symptom disorder diagnosis in research and practice: a critical evaluation and a proposal for modifications. *Annu Rev Clin Psychol* 2014;10:339–67.
- Kop WJ, Toussaint A, Mols F, et al. Somatic symptom disorder in the general population: associations with medical status and health care utilization using the SSD-12. *Gen Hosp Psychiatry* 2019;56:36–41.
- Katon W, Lin EHB, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. *Gen Hosp Psychiatry* 2007;29:147–55.
- Rief W, Burton C, Frosthalm L, et al. Core outcome domains for clinical trials on somatic symptom disorder, bodily distress disorder, and functional somatic syndromes: European network on somatic symptom disorders recommendations. *Psychosom Med* 2017;79:1008–15.
- Henningsen P, Zipfel S, Herzog W. Management of functional somatic syndromes. *Lancet* 2007;369:946–55.
- Teunissen SCCM, Wesker W, Kruiwagen C, et al. Symptom prevalence in patients with incurable cancer: a systematic review. *J Pain Symptom Manage* 2007;34:94–104.
- Grassi L, Caruso R, Nanni MG. Somatization and somatic symptom presentation in cancer: a neglected area. *Int Rev Psychiatry* 2013;25:41–51.
- Kop WJ, Synowski SJ, Gottlieb SS. Depression in heart failure: biobehavioral mechanisms. *Heart Fail Clin* 2011;7:23–38.
- Kohlmann S, Gierk B, Hümmelgen M, et al. Somatic symptoms in patients with coronary heart disease: prevalence, risk factors, and quality of life. *JAMA Intern Med* 2013;173:1469–71.
- Toft T, Fink P, Oernboel E, et al. Mental disorders in primary care: prevalence and co-morbidity among disorders. results from the functional illness in primary care (FIP) study. *Psychol Med* 2005;35:1175–84.
- Verhaak PFM, Meijer SA, Visser AP, et al. Persistent presentation of medically unexplained symptoms in general practice. *Fam Pract* 2006;23:414–20.
- Klaus K, Rief W, Brähler E, et al. The distinction between "medically unexplained" and "medically explained" in the context of somatoform disorders. *Int J Behav Med* 2013;20:161–71.
- Rosendal M, Olde Hartman TC, Aamland A. "Medically unexplained" symptoms and symptom disorders in primary care: prognosis-based recognition and classification. *BMC Fam Pract* 2017;18:1–9.
- Khan AA, Khan A, Harezlak J, et al. Somatic symptoms in primary care: etiology and outcome. *Psychosomatics* 2003;44:471–8.
- Kroenke K. Patients presenting with somatic complaints: epidemiology, psychiatric comorbidity and management. *Int J Methods Psychiatr Res* 2003;12:34–43.
- Eikelboom EM, Tak LM, Roest AM, et al. A systematic review and meta-analysis of the percentage of revised diagnoses in functional somatic symptoms. *J Psychosom Res* 2016;88:60–7.
- Page LA, Wessely S. Medically unexplained symptoms: exacerbating factors in the doctor-patient encounter. *J R Soc Med* 2003;96:223–7.
- American Psychiatric Association, DSMT Force. Diagnostic and statistical manual of mental disorders: DSM-5, 2013. Available: <http://dsm.psychiatryonline.org/book.aspx?bookid=556>
- Smith BJ, McGorm KJ, Weller D, et al. The identification in primary care of patients who have been repeatedly referred to hospital for



- medically unexplained symptoms: a pilot study. *J Psychosom Res* 2009;67:207–11.
- 22 Marks EM, Hunter MS. Medically unexplained symptoms: an acceptable term? *Br J Pain* 2015;9:109–14.
- 23 Sirri L, Grandi S, Tossani E. Medically unexplained symptoms and general practitioners: a comprehensive survey about their attitudes, experiences and management strategies. *Fam Pract* 2017;34:201–5.
- 24 Burton C. Beyond somatisation: a review of the understanding and treatment of medically unexplained physical symptoms (MUPS). *Br J Gen Pract* 2003;53:231–9.
- 25 van Dessel N, den Boeft M, van der Wouden JC, et al. Non-pharmacological interventions for somatoform disorders and medically unexplained physical symptoms (MUPS) in adults. *Cochrane Database Syst Rev* 2014;11:CD011142.
- 26 van Westrienen PE, Pisters MF, Gerrits M, et al. Identifying treatment modalities for a multidisciplinary and blended care intervention for patients with moderate medically unexplained physical symptoms: qualitative study among professionals. *JMIR Ment Health* 2019;6:e12203.
- 27 Kitselaar WM, van der Vaart R, van Tilborg-den Boeft M, et al. The general practitioners perspective regarding registration of persistent somatic symptoms in primary care: a survey. *BMC Fam Pract* 2021;22:182.
- 28 Konnopka A, Schaefer R, Heinrich S, et al. Economics of medically unexplained symptoms: a systematic review of the literature. *Psychother Psychosom* 2012;81:265–75.
- 29 Zonneveld LNL, Sprangers MAG, Kooiman CG, et al. Patients with unexplained physical symptoms have poorer quality of life and higher costs than other patient groups: a cross-sectional study on burden. *BMC Health Serv Res* 2013;13:520.
- 30 Franks P, Clancy CM, Nutting PA. Gatekeeping revisited--protecting patients from overtreatment. *N Engl J Med* 1992;327:424–9.
- 31 Boerma WG, van der Zee J, Fleming DM. Service profiles of general practitioners in Europe. European GP task profile study. *Br J Gen Pract* 1997;47:481–6.
- 32 Loudon I. The principle of referral: the gatekeeping role of the GP. *Br J Gen Pract* 2008;58:128–30.
- 33 Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing* 2016;45:353–60.
- 34 Merriel SWD, May MT, Martin RM. Predicting prostate cancer progression: protocol for a retrospective cohort study to identify prognostic factors for prostate cancer outcomes using routine primary care data. *BMJ Open* 2018;8:e019409.
- 35 Koning NR, Büchner FL, Vermeiren RRJM, et al. Identification of children at risk for mental health problems in primary care-Development of a prediction model with routine health care data. *EClinicalMedicine* 2019;15:89–97.
- 36 Smeets HM, Kortekaas MF, Rutten FH, et al. Routine primary care data for scientific research, quality of care programs and educational purposes: the Julius general practitioners' network (JGPN). *BMC Health Serv Res* 2018;18:735.
- 37 Schaefer R, Laux G, Kaufmann C, et al. Diagnosing somatisation disorder (p75) in routine general practice using the International classification of primary care. *J Psychosom Res* 2010;69:267–77.
- 38 Morris LD, Grimmer-Somers KA, Louw QA, et al. Cross-cultural adaptation and validation of the south african pain catastrophizing scale (SA-PCS) among patients with fibromyalgia. *Health Qual Life Outcomes* 2012;10:13.
- 39 den Boeft M, van der Wouden JC, Rydell-Lexmond TR, et al. Identifying patients with medically unexplained physical symptoms in electronic medical records in primary care: a validation study. *BMC Fam Pract* 2014;15:109.
- 40 Sitnikova K, Pret-Oskam R, Dijkstra-Kersten SMA, et al. Management of patients with persistent medically unexplained symptoms: a descriptive study. *BMC Fam Pract* 2018;19:88.
- 41 den Boeft M. *Risk assessment models for patients with persistent medically unexplained physical symptoms in primary care using electronic medical records. In: medically unexplained physical symptoms in primary care: identification, management and societal aspects*, 2012: 43–62. <https://research.vu.nl/en/publications/medically-unexplained-physical-symptoms-in-primary-care-identific>
- 42 Terluin B, van Marwijk HWJ, Adèr HJ, 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* 2006;6:34.
- 43 Smith RC, Gardiner JC, Armatti S, et al. Screening for high utilizing somatizing patients using a prediction rule derived from the management information system of an HMO: a preliminary study. *Med Care* 2001;39:968–78.
- 44 Masters ET, Mardekian J, Emir B, et al. Electronic medical record data to identify variables associated with a fibromyalgia diagnosis: importance of health care resource utilization. *J Pain Res* 2015;8:131–8.
- 45 Emir B, Masters ET, Mardekian J, et al. Identification of a potential fibromyalgia diagnosis using random forest modeling applied to electronic medical records. *J Pain Res* 2015;8:277–88.
- 46 STIZON - Stichting Informatievoorziening voor Zorg en Onderzoek. Available: <https://www.stizon.nl/>
- 47 ICPC-online. Het Nederlands Huisartsen Genootschap is de wetenschappelijke vereniging van huisartsen: ICPC-online, 2020. Available: <https://www.nhg.org/themas/artikelen/icpc-online>
- 48 nhg. Het Nederlands Huisartsen Genootschap is de wetenschappelijke vereniging van huisartsen: ICPC, 2020. Available: <https://www.nhg.org/themas/artikelen/icpc>
- 49 Methodology WCCfDS. Guidelines for ATC classification and DDD assignment 2013. Oslo, 2012. Available: [https://www.whocc.no/filearchive/publications/1\\_2013guidelines.pdf](https://www.whocc.no/filearchive/publications/1_2013guidelines.pdf)
- 50 von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement. *Epidemiology* 2007;18:800–4.
- 51 World Health Organisation. *International statistical classification of diseases and related health problems*. 11th ed. Geneva: World Health Organisation, 2020.
- 52 Robbins JM, Kirmayer LJ, Hemami S. Latent variable models of functional somatic distress. *J Nerv Ment Dis* 1997;185:606–15.
- 53 Nielen M, Verheij R. Morbiditeit in de huisartsenpraktijk: NIVEL Zorgregistraties eerste lijn. *TSG* 2015;93:287.
- 54 R: the R project for statistical computing. Available: <https://www.r-project.org/>
- 55 de Waal MWM, Arnold IA, Eekhof JAH, et al. Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *Br J Psychiatry* 2004;184:470–6.
- 56 Nilsson G, Åhlfeldt H, Strender L-E. Computerisation SLE. Computerisation, coding, data retrieval and related attitudes among Swedish General practitioners-a survey of necessary conditions for a database of diseases and health problems. *Int J Med Inform* 2002;65:135–43.