PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Identifying persistent somatic symptoms in electronic health records: exploring multiple theory-driven methods of identification
AUTHORS	Kitselaar, Willeke; Numans, Mattijs; Sutch, Stephen; Faiq, Ammar; Evers, Andrea; van der Vaart, R

VERSION 1 – REVIEW

REVIEWER	Zghebi, Salwa University of Manchester, Centre for Primary Care and Health Services Research
REVIEW RETURNED	03-Apr-2021

GENERAL COMMENTS	This is a good study exploring five methods to identify PSS in adults using EHRs from 76 general practices in the Netherlands between 2014 and 2018. • Abstract: in general, the abstract needs a review for consistency as in some parts it seems different from the rest of the paper. For example, the labels of the methods A-D differs in the abstract from the other sections of the paper, e.g. method B is defined as the PSS-related terminology in the abstract but as the PSS-related symptoms in the main text. Also, in the abstract the authors refer to 'comorbid' chronic physical and mental conditions which is unclear they are co-morbid to which index condition – are PSS symptom/syndrome considered as the index 'condition' here? In addition, in the rest of the paper, there is no mention of 'comorbid' conditions, and instead these are referred to as just chronic physical/mental conditions, which is the appropriate way to describe them given the lack of a defined index condition in the included patients. • Identification methods section: the authors state "….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." So how the possibility of including cases where their PSS symptoms are linked to other diagnosed conditions was accounted for in the analysis? • Identification methods section: this is the first time that the fifth method is mentioned as it is not mentioned in the abstract or introduction. Please add it to all sections appropriately. • Method B was based on symptoms on the 'Robbins list', but it is unclear what PSS symptoms this lists comprises? Please clarify on its first mention on Table 1. • Table 1: how PSS terminology in method C is defined is not very clear. So was is based on combined methods A +B? i.e search by ICPC codes of both PSS symptoms and PSS syndromes or by

	searching for PSS-related text terms? Please make it clearer in the text. It would help if an example is provided. • Discussion: can the authors comment on the how are the results are explained in the context of Dutch primary care? For example, is there knowledge on which of the examined method(s) is commonly-used by general practitioners in the country or in the area where the 76 general practices were based? Also, the findings in relation to national prevalence of PSS, if known. • Is there potential generalisability of the reported results to the identification of PSS in other countries? • Discussion: "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." This observation implies using the consultations frequency to define method B and also included as a criteria to measure the HCU outcome. Did the authors consider how to overcome this in the analysis or via a sensitivity analysis after dropping this mutual criterion from either side? • The study concludes the need for using either a single or multiple methods to identify PSS cases may depend on the aim of the identification. May the authors please provide 1-2 working examples per method, in addition to the prevalence rates example mentioned? • As recommendations and implications for future work, would external validation in non-Dutch primary care be considered? this would depend on the replicability of the methodology to non-Dutch EHRs.

REVIEWER	Gulliford, Martin King's College London, UK
REVIEW RETURNED	12-Apr-2021

GENERAL COMMENTS	This is a well written paper that describes analyses to evaluate the frequency of persistent somatic symptoms using four definitions in a primary care population using electronic health records. The work will be of interest to readers with closely related interests who, even if they do not agree with the approach, may find it informative. The abstract could be clearer. Most of the cases had physical and mental comorbid conditions, so can we really say that symptoms were 'unexplained' as discussed in the Introduction. The title refers to theory driven approaches but these theories are not described in the Introduction or Methods. Table 1 and associated text needs to ensure that abbreviations are clearly defined. Table 2, it may be useful to show a Venn diagram of the overlap of different classifications. Table 5. It does not appear to be reasonable to refer to sensitivity and specificity when there is no agreed reference method. Methods for agreement (reliability) would be more appropriate. Confidence intervals for key metrics should be presented.

	More attention shyould be given to missing values. For example, what proportion of population had key measures such as 4DSQ recorded and how was this addressed in the analysis.

REVIEWER	Qureshi, Nadeem University of Nottingham, of Primary Care
REVIEW RETURNED	12-Apr-2021

GENERAL COMMENTS	A review of “Identifying persistent somatic symptoms in electronic health records: exploring multiple theory-driven methods of identification” BMJ open Abstract A good summary of the research that has been undertaken. Suggested amendment I appreciate that the word count may be limited but it would be helpful to provide more information about PSS Introduction clear presentation of the background to the study Method A thorough presentation of the approach used. Suggested amendments • Page 8: as highlighted in discussion, some overlap between identified methods and outcomes but not clear from text that the authors has covered this in the method section • Page 8: please provide more details about the context of method B and the basis of method C. • page 9 the metrics used for the health care utilisation are not clear Results A straightforward presentation of the findings has been presented Suggested amendments • page 11: in table 2, it would be useful to have a summary of each approach but just simply labelled them as method eight, B, C • page 13: not clear of the value of describing measurements as specificity and sensitivity when not compared to a criterion standard. Better presenting as percentage overlap rather than predictive accuracy. To exemplify this, it would be helpful to include a Venn diagram showing the overlap • page 14: it is not clear the value of an analysis of overlap of practices. What does this indicate? Discussion The authors have summarises the overall descriptive data are well Suggested amendments • page 17: I am not convinced by the clinical implications. Further details need to be provided about the actual clinical utility of this tool or accept that this is simply a methodological paper • page 17 : other research recommendations do not really make sense. The researchers are simply increasing the proportion of individuals who would fit the PSS categories but not indicate which individuals comply with the optimal criteria. Which criteria would have the greatest clinical utility? The discussion of more advanced

	computer systems is irrelevant without a clear definition of optimal PSS criteria (unless authors are stating that the method B is the optimal approach, but need to justify this) Overall this was a very difficult paper to assess. It may be worth consideration with major revisions to improve clarity and key message

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Salwa Zghebi, University of Manchester
Comments to the Author:
This is a good study exploring five methods to identify PSS in adults using EHRs from 76 general practices in the Netherlands between 2014 and 2018.

1. **Abstract**: in general, the abstract needs a review for consistency as in some parts it seems different from the rest of the paper. For example, the labels of the methods A-D differs in the abstract from the other sections of the paper, e.g. method B is defined as the PSS-related terminology in the abstract but as the PSS-related symptoms in the main text. Also, in the abstract the authors refer to 'comorbid' chronic physical and mental conditions which is unclear they are co-morbid to which index condition – are PSS symptom/syndrome considered as the index 'condition' here? In addition, in the rest of the paper, there is no mention of 'comorbid' conditions, and instead these are referred to as just chronic physical/mental conditions, which is the appropriate way to describe them given the lack of a defined index condition in the included patients.
We thank the reviewer for his attentiveness on these matters and made changes according to his suggestions. Method B and C have been switched and comorbid has been removed from the abstract.

2. **Identification methods section**: the authors state *"….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."* So how the possibility of including cases where their PSS symptoms are linked to other diagnosed conditions was accounted for in the analysis?
Each individual symptom could not specifically be linked to other diagnosed chronic conditions, since we basically made use of coded data and rarely symptoms are coded alongside the chronic conditions they are potentially linked with in routine daily practice. It would also be very difficult to see if these symptoms (their seriousness or their frequency of occurrence) are within 'expected' ranges, because this is not clearly reported in the data. The main goal being to identify patients with a recognizable risk of having persistent problems, we did not consider details on symptoms paramount to the study's success. Exploring the different outcome measures (demographics, HCU, other conditions) is what is meant to (approach) validation of each method.

3. **Identification methods section**: this is the first time that the fifth method is mentioned as it is not mentioned in the abstract or introduction. Please add it to all sections appropriately.
Again, we thank the reviewer for his attentiveness. This has been added to the abstract (p2: 'and all methods combined.') and introduction (p5: 'Lastly, all methods (A-D) were combined.').

4. Method B was based on symptoms on the 'Robbins list', but it is unclear what PSS symptoms this lists comprises? Please clarify on its first mention on Table 1.

We thank the reviewer for this remark and have now added this to Table 1. See page 9: ‘c 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, palpations, dizziness, lump in throat, numbness, nausea, loose bowels, gas/bloating, constipation, abdominal pain.[52]’

5. **Table 1**: how PSS terminology in method C is defined is not very clear. So was it based on combined methods A +B? i.e search by ICPC codes of both PSS symptoms and PSS syndromes or by searching for PSS-related text terms? Please make it clearer in the text. It would help if an example is provided.
We agree that this method could be more clarified. Terms used for selecting patients with method C were searched by looking at the terms to which GPs changed an episode description to when using an ICPC-code indicating PSS. If we found a new term, we would do a search for ICPC’s that were also linked to the term. Then we would search these ICPC-codes for new terms, etc. Until we did not find any new terms – i.e., a cross-search. This has been further clarified on page 7 and the text has been changed accordingly: ‘(the episode description is adjustable for GPs; i.e., in case a GP registers A04.01, this automatically gives the description ‘chronic fatigue syndrome’, but the description can be adjusted to any term the GP prefers. Our available data was systematically searched by cross-checking ICPC codes and related descriptions)’

6. **Discussion**: can the authors comment on the how are the results are explained in the context of Dutch primary care? For example, is there knowledge on which of the examined method(s) is commonly-used by general practitioners in the country or in the area where the 76 general practices were based? Also, the findings in relation to national prevalence of PSS, if known.
Information on the national prevalence (incl. reference) has been added at page 14: (‘since prevalence rates of PSS in the general Dutch population most likely range from 10-15%.[55]’). The results on the distribution of the methods used by specific practices (figure 2) show that there is variance between practices using the methods. This has been further elaborated in the discussion on at page 14: ‘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.’
7. Is there potential generalizability of the reported results to the identification of PSS in other countries?
Indeed, this is an important point. We have added a sentence on this in strengths and limitations section (p.15): ‘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 may be needed to generalize the results to other countries.’

8. **Discussion**: *"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."* This observation implies using the consultations frequency to define method B and also included as a criteria to measure the HCU outcome. Did the authors consider how to overcome this in the analysis or via a sensitivity analysis after dropping this mutual criterion from either side?
We did consider dropping consultation frequency as a criterion for method B. However, this would include too many patients who may have non-persistent problems, or whom might have consulted only incidentally. Furthermore, we aimed to stay as closely as possible to methods used in previous research (and this was a criterion in the other studies).

9. The study concludes the need for using either a single or multiple methods to identify PSS cases may depend on the aim of the identification. May the authors please provide 1-2 working examples

per method, in addition to the prevalence rates example mentioned?
Another example has been added at page 14: ‘However, using a single method (e.g., method C) may be sufficient to identify risk factors for persistence of PSS, although this should be confirmed by further research.’

10. As recommendations and implications for future work, would external validation in non-Dutch primary care be considered? this would depend on the replicability of the methodology to non-Dutch EHRs.
Thank you for suggesting. This has been added on page 15: ‘From a research perspective, in the first place, replicability of the methods to non-Dutch EMRs should be examined.’

Reviewer: 2
Dr. Martin Gulliford, King's College London, UK
Comments to the Author:
This is a well written paper that describes analyses to evaluate the frequency of persistent somatic symptoms using four definitions in a primary care population using electronic health records.

The work will be of interest to readers with closely related interests who, even if they do not agree with the approach, may find it informative.

11. The abstract could be clearer. Most of the cases had physical and mental comorbid conditions, so can we really say that symptoms were 'unexplained' as discussed in the Introduction.
This is a much debated and difficult point. Are symptoms ever really unexplained when they can be explained by a combination of factors from different health domains (bio-psycho-social-behavior)? We believe that for this reason the current stance in the field of PSS is leaning towards not focusing on symptoms being explained or unexplained pathophysiologically, but rather focus on the somatic presentation, which means that positive criteria are used instead of previously used negative criteria of classification. This is also the reason for using the term PSS (for elaboration on this please see page 4). At present, the DSM and ICD classification do not require exclusion based on chronic physical conditions or comorbidity, which stems from the knowledge that patients with a well-established chronic condition can have more/more severe symptoms then other patients with the same condition, due to the interplay between the earlier mentioned health domains. In that sense, the symptoms that are outside of the ‘normal’ range for a specific (‘explained’) disease (severity), are therefore seen as unexplained, in this case meaning that they are not fully explained by biomedical pathology. This makes sense, since the ‘unexplained’ symptoms in patients with an ‘explained’ disorder are affected by multidomain factors, similar to patients with PSS without an ‘explained’ disorder. We added a couple of sentences accordingly in the introduction:
‘Moreover, the term PSS is in line with recent advance 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 symptomology (e.g., the presence and burden of symptoms).[20]’ (p.4)
‘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.’ (p.5)

12. The title refers to theory driven approaches but these theories are not described in the Introduction or Methods.
Theory-driven refers to – based on the current literature (i.e., contrary to data-driven; AI-methods). While this is stated in the introduction, page 6, it was not clearly stated in the methods section and

this has been corrected on page 8:
‘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 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’

13. Table 1 and associated text needs to ensure that abbreviations are clearly defined.
We have adjusted this:
‘a Examples of included terms: somatization, psychosomatic, central sensitization, atypical low back pain, stress related pain, interstitial cystitis, extreme fatigue, tension headache: good CT, functional.
b Four-dimensional symptom questionnaire.[42]
c 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, palpations, dizziness, lump in throat, numbness, nausea, loose bowels, gas/bloating, constipation, abdominal pain.[52]’
For more details, please see page 9.

14. Table 2, it may be useful to show a Venn diagram of the overlap of different classifications.
We thank the reviewer for this suggestion. Since reviewer 3 also requested a Venn diagram regarding the overlap, a Venn diagram has been added (figure 1; replacing table 5).

15. Table 5. It does not appear to be reasonable to refer to sensitivity and specificity when there is no agreed reference method. Methods for agreement (reliability) would be more appropriate.
In accordance with this remark and reviewers 3 suggestion we have simplified the results to percentages of overlap and added a Venn diagram. This led to the following correction in the methods section:
‘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.’ (p.9)
And the following correction of the results:
‘Relative to other methods (all ≤ 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 (≤ 1.3%) (see Figure 1 for an overview of overlap between all the methods).’ (p.12)

16. Confidence intervals for key metrics should be presented.
This paper is primarily meant to be descriptive. Because methods were not statistically tested/compared to one another, no confidence intervals are available.

17. More attention should be given to missing values. For example, what proportion of population had key measures such as 4DSQ recorded and how was this addressed in the analysis.
Thank you for this attentive remark. The recoding of the 4DSQ has been reported now in the results section (p.10: ‘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.’). This was not further addressed in the analysis, because of the descriptive nature of this paper. Due to large sample sizes and the large number of missing values in EMR-data (for which it is unclear if they are missing because they were not relevant or because of inadequate diagnostics/reporting), statistical testing is deemed inappropriate. In general, the lacking overlap between the methods (A-D) imply the quantity of relevant missing data (related to identification).

Reviewer: 3
Dr. Nadeem Qureshi, University of Nottingham
Comments to the Author:
A review of “Identifying persistent somatic symptoms in electronic health records: exploring multiple theory-driven methods of identification” BMJ open

Abstract
A good summary of the research that has been undertaken.

Suggested amendment
18. I appreciate that the word count may be limited but it would be helpful to provide more information about PSS
Adjustment has been added (p.2): ‘Persistent somatic symptoms (PSS) are defined as symptoms not fully explained by well-established pathophysiological mechanisms and are prevalent in up to 10% of patients in primary care.’

Introduction
clear presentation of the background to the study

Method
A thorough presentation of the approach used.

Suggested amendments
19. Page 8: as highlighted in discussion, some overlap between identified methods and outcomes but not clear from text that the authors has covered this in the method section
We thank the reviewer for this comment and have amended it on page 7: ‘Additionally, besides exploring overlap between methods, all four methods were integrated, selecting all patients identified by any of the methods.’

20. Page 8: please provide more details about the context of method B and the basis of method C.
We thank the reviewer for this comment and have amended it on page 8:
Method B: ‘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’
Method C: method C identifies patients based on PSS-related terminology in the episode description (the episode description is adjustable for GPs; i.e., in case a GP registers A04.01, this automatically gives the description ‘chronic fatigue syndrome’, but the description can be adjusted to any term the GP prefers. Our available data was systematically searched by cross-checking ICPC codes and related descriptions)’

21. page 9 the metrics used for the health care utilisation are not clear
This has been improved on page 9: ‘For all HCU frequencies, mean one-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.[49] 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.

Results
A straightforward presentation of the findings has been presented

Suggested amendments
22. page 11: in table 2, it would be useful to have a summary of each approach but just simply labelled them as method eight, B, C
Thank you kindly for this suggestion. This has been adjusted for all tables.

23. page 13: not clear of the value of describing measurements as specificity and sensitivity when not compared to a criterion standard. Better presenting as percentage overlap rather than predictive accuracy. To exemplify this, it would be helpful to include a Venn diagram showing the overlap
Due to the lack of a golden-standard for identifying PSS, it was aimed to clarify how many patients we would miss when using only one method. While we feel that the specificity and sensitivity is the best way to do this, we do understand that the former presentation makes it too complex. Instead of table 5, a Venn diagram has now been added.
24. page 14: it is not clear the value of an analysis of overlap of practices. What does this indicate?
The variation between practices indicates that implementation of any method in a specific practice may result in missing patients at risk of having PSS because the GP in the specific practice may not use this specific method to register a patient with PSS. As your question indicates that this was not clearly described, and it became clear that an explanation in the discussion was missing, therefore we have added this on page 14: '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. '
And page 15: 'Finally, while it was beyond the scope of this study to investigate 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.[27] Future research should investigate whether the need for support can be linked or tailored to GPs with specific registration methods.'

Discussion
The authors have summarises the overall descriptive data are well

Suggested amendments
25. page 17: I am not convinced by the clinical implications. Further details need to be provided about the actual clinical utility of this tool or accept that this is simply a methodological paper
We agree that it should be clearer from the text that this is primarily a methodological paper. However, the final goal would be to find an algorithm based on routine EMR analyses, that can be used as feedback to physicians to increase awareness of the risk of individual patients encountering them. Therefore, we have added a sentence to clarify this on page 15: '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 amongst GPs regarding PSS-risk).'
Method B is relevant for screening as the present study corroborates previous research using it. However, this needed a clearer statement of scientific support – which has also been added to the section at page 16: 'Method B is supported by previous studies which successfully used a similar method for screening routine care data for patients with PSS.[26, 39, 40]'

26. page 17 : other research recommendations do not really make sense. The researchers are simply increasing the proportion of individuals who would fit the PSS categories but not indicate which individuals comply with the optimal criteria. Which criteria would have the greatest clinical utility? The discussion of more advanced computer systems is irrelevant without a clear definition of optimal PSS criteria (unless authors are stating that the method B is the optimal approach, but need to justify this)

We thank the reviewer for their comment. The research recommendations have been rewritten as follows (p.15):
'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, [43] 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 miss-classified as non-PSS, which may enhance the possibilities for data-driven predictive modeling 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.[27] Future research should investigate whether the need for support can be linked or tailored to GPs with specific registration methods.'
The discussion of more advanced computer systems was misplaced and has been added to the clinical implications sections (p.15-16):
'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 amongst GPs regarding PSS-risk). Firstly, 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" – i.e., 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.[26, 39, 40] Subsequently, validated questionnaires such as the 4DSQ [42] or the somatic symptom disorder B-criteria scale (SSD-12) [4] can be used to identify those patients selected by method B who need additional attention/pro-active 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.[56]'

Overall this was a very difficult paper to assess. It may be worth consideration with major revisions to improve clarity and key message

VERSION 2 – REVIEW

REVIEWER	Zghebi, Salwa University of Manchester, Centre for Primary Care and Health Services Research
REVIEW RETURNED	23-Jul-2021

GENERAL COMMENTS	Thank you for addressing my previous comments. I have no further comments. Good luck.