PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Research Protocol for a Diagnostic Study: Identifying and Measuring the Severity of Somatic Symptom Disorder Using the Self-reported Somatic Symptom Scale-China (SSS-CN)
AUTHORS	Jiang, Meng; Zhang, Weituo; Su, Xuan; Gao, Chuang; Chen, Bingxu; Feng, Zehao; Mao, Jialiang; Pu, Jun

VERSION 1 - REVIEW

REVIEWER	Johannes A.C. Laferton Friedrich-Alexander-Universität Erlangen-Nürnberg & Psychologische Hochschule Berlin
REVIEW RETURNED	02-Jul-2018

GENERAL COMMENTS	Peer Review – BMJ Open – 2018-024290 Title: Somatic Symptom Scale-China (SSS-Ch) study: protocol for measurement and severity evaluation of a self-report version of a somatic symptom questionnaire in a general hospital in China In their manuscript Su et al. report on a study protocol of a diagnostic accuracy study to assess the validity of the self-report Somatic Symptom Scale-China (SSS-Ch) in screening for the new DSM-5 Somatic Symptom Disorder (SSD). Patients with physical discomfort presenting to internal medicine departments in a tertiary hospital in China will be recruited after a negative physical exam. They will successively fill out the SSS-Ch (and other self-report measures for the sake of validity estimation) and a clinical interview to diagnose SSD according to the criteria of DSM-5. The reviewer commends the efforts of the authors to tackle the important issue of providing (time) effective screening tools to identify patients at risk of SSD by conducting such a large scale study. However, the manuscript at its' current form lacks precision related to the description of the study rationale, the previous evidence on diagnostic tools to identify SSD and most importantly related to study methodology. Several aspects of methodology are missing or are unprecise according to the standardized reporting recommendations of diagnostic accuracy studies (STARD 2015 guidelines; (Cohen et al., 2016). The most pressing of which is the lack of detail in describing the reference standard. Therefore, this manuscript needs major revisions to better inform about the study rational and methodology. Detailed points can be seen below. I hope these might aid the authors in order to improve the presentation of their important scientific undertaking. General:

	Point 1: The reviewer is not a native English speaker himself and would therefore like refer the judgment to a native speaking reviewer or editor. However, from my judgment, it might be possible that some of the lacking precision might be improved by English language editing. Introduction: Point 2: First paragraph. It appears unclear whether the authors are referring to the precisely defined term of Somatic Symptom Disorder or to a more general concept i.e. somatoform symptoms, medically unexplained symptoms and the like. P5L8-19 rather describe a broader concept and not specifically the criteria of SSD in DSM-5. Further, the epidemiological references partly refer to different (older) conceptualizations (L19-25). Moreover, there is no SSD in ICD-10 (L19). Point 3: (P5L25-29) The authors state that there is a lack of effective screening instruments for SSD. This does not reflect the current state of the literature. (Laferton et al., 2017) have provided diagnostic accuracy estimates of identifying SSD using the screening self report questionnaires asking about physical and psychological symptoms (PHQ-15, the WI-7 and the SAIB). (Gierk et al., 2014, 2015) and (A Toussaint et al., 2015; Anne Toussaint, Löwe, Brähler, & Jordan, 2017) have assessed validity and reliability for the SSS-8 and the SSD-12 for assessing somatic symptoms and psychological symptoms of SSD respectively. (Tu et al., 2016) have reported on using the WI-7 in screening for SSD. Putting the current study in to the context of what is known so far would be helpful to get a clearer understanding of the exact study rationale. It appears to the reviewer, that the authors do want to assess the utility of a screening tool that assesses physical symptoms AND psychological symptoms and whether the utility is greater than just assessing physical symptoms. If so, it would clearly inform the reader to state the up to date knowledge on studies evaluating both psychological and physical symptom screenings for the detection of SSD. Point 4: (P5L49-54) “The DSM 5 is currently the “gold standard” for the diagnosis of SSD,….” I am not sure if this is entirely correct. The DSM 5 is the entity defining the new diagnosis of SSD. It is not, however, a specific assessment instrument (e.g. interview or questionnaire) and can therefore not be the gold standard for assessment. This also relates to P6L3-8, where the authors state that the DSM-5 includes an interview longer than half an hour. To my knowledge, the DSM-5 section on SSD does not give specific instructions on how to conduct an interview to diagnose SSD. There are several published interviews to diagnose SSD based on the DSM-5 criteria, including the SCID for DSM5 interview Module J or an interview published by (Axelsson, Andersson, Ljótsson, Wallhed Finn, & Hedman, 2016). Do the authors refer to one or any other interview at this part of the introduction? Point 5: (P6 L11-13) “The most available questionnaires are the PHQ-9 and GAD-7.” I am not entirely sure what the authors do wish to express? Do they mean the most available to assess depressive and anxiety symptoms? If so it might be an even stronger argument to state that their extensively psychometrically evaluation.

	Point 6: (P6L20) The SSS-8 is “a simplified version” of the PHQ-15. I suggest using “abbreviated” instead of simplified. Moreover, since this part of introduction serves to justify the authors use of the respective questionnaires it might be informative to the reader to state the current evidence regarding those questionnaires in screening for SSD (the research question of this manuscript). (see Point 3) Point 7: (P6L33-P7L44) The authors state that the newly developed SSS-Ch is designed to aid SSD diagnosis, specifically by including items assessing depressive and anxiety symptoms. The improvement of screening tools is an important endeavor, especially for a new diagnostic category like the SSD. However, the specific goal of the research outlined in the manuscript does not become entirely clear from this section of the manuscript. The authors state that depression and anxiety is frequently co-occurring in patients with SSD. This might suggest, that the SSS-Ch aims to assess patients SSD diagnosis AND co-morbid anxiety or depressive disorders? However, that is not explicitly state here. Also, it that case the reference standard would have to include anxiety and depression related diagnosis. The alternative research focus, I assume, would be to improve screening for SSD by including psychological criteria into the SSS-Ch to test whether this improves screening accuracy over and above screening for somatic symptoms (e.g. SSS-8 or PHQ-15). In that case, the argument of the authors would appear stronger, if the would include current evidence on such research (Point 3 and Point 6). Further, the SSS-Ch includes only one item (#17) that appears to assess SSD specific anxiety, the other psychological items refer to depressive or anxiety related symptoms that are more commonly found in other DSM-5 disorders. It would be interesting to read the authors reasoning how these more general anxiety or depressive symptoms are expected to increase SSD screening accuracy. This might still be a valid research hypothesis, yet the reader might expect why the authors do not focus on psychological symptoms of SSD as formulated in the DSM-5 or expressed in other research on this topic (e.g. Klaus et al., 2015; Rief & Martin, 2014). Assuming that improving SSD screening by adding psychological items is the authors research focus, wouldn’t the primary objective be to assess whether the SSS-Ch (including psychological items) outperforms the PHQ-15 (including only somatic symptoms)? Further, it is not clear to me which hypothesis the authors formulate by primary objective 2: ‘We expect to use the SSS-Ch for measuring SSD severity. Methods Point 8: (P8L38) “Since you have been unwell, how often have you been bothered by any of the following problems”. The SSS-Ch appear to assess symptoms for a much longer (and more unspecified) time period as other self assessment questionnaires (e.g. PHQ-15: past 4 weeks). It would be interesting to hear the reasoning behind that choice. Also, the introductory sentence asks about frequency (‘how often have you been bothered’) but the scoring is using a scaling referring to severity of influence (e.g. ‘unbearable’). Is this due to a translation error? Moreover, since the original manuscript on the validation of the SSS-Ch is not available in English, it would be informative to read more on details of the psychometric characteristics. Especially on those that will be used

	in the current investigation like the severity categories based on percentile ranks. Point 9: (P9L5-10). Only patients after a negative physical exam will be included into the diagnostic accuracy study. Although the authors already state that this is a limitation, the limitation might warrant more explicitly. Given that the assessment of absence of somatic disease is unreliable (Dimsdale et al., 2013) this might result in bias related to the epidemiologic results and the screening accuracy results of this study. Moreover, the external validity will be limited to patients after a negative physical exam. This is underlined by the assumption referred to in the power calculation (P11L48) where the authors expect a huge prevalence of SSD (76.9%). Point 10: The authors do not mention an informed consent procedure. Point 11: Exclusion criteria (P9L44-54) '4) patients who are unable to complete at least 1 follow up.' How is that predetermined before the inclusion and especially the follow up takes place? Point 12: The most pressing issue is the lack of description of the reference standard. First, it might be helpful to the reader to use the standardized structure of method sections as described by the STARD guidelines. Second, there is only minimal description of the reference standard, which does not allow for replication of the study and does not give information about the reliability or validity of the reference standard. Point 13: Patient involvement statement. The authors do state that patients were involved in certain areas of study design, however they do not describe how. Did the authors invite patients to comment on study design and procedure? Did the authors conduct a test phase including patients feedback? Discussion: Point 14: Several points mentioned above do also refer to the discussion section. Point 15: Figure 1 in my file is a bit blurry and has quite small fond. Point 16: Several points mentioned above, and especially a clearer more explicit statement about the study rational might also needed to be adapted in the Strengths and limitations statement (P4) References Axelsson, E., Andersson, E., Ljótsson, B., Wallhed Finn, D., & Hedman, E. (2016). The health preoccupation diagnostic interview: inter-rater reliability of a structured interview for diagnostic assessment of DSM-5 somatic symptom disorder and illness anxiety disorder. Cognitive Behaviour Therapy, 45(4), 259–269. http://doi.org/10.1080/16506073.2016.1161663 Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., … Bossuyt, P. M. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open, 6(11), e012799. http://doi.org/10.1136/bmjopen-2016-012799

	Dimsdale, J. E., Creed, F., Escobar, J., Sharpe, M., Wulsin, L., Barsky, A., … Levenson, J. (2013). Somatic symptom disorder: An important change in DSM. Journal of Psychosomatic Research, 75(3), 223–228. http://doi.org/10.1016/j.jpsychores.2013.06.033 Gierk, B., Kohlmann, S., Kroenke, K., Spangenberg, L., Zenger, M., Brähler, E., & Löwe, B. (2014). The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Internal Medicine, 174(3), 399–407. http://doi.org/10.1001/jamainternmed.2013.12179 Gierk, B., Kohlmann, S., Toussaint, A., Wahl, I., Brünahl, C. a, Murray, A. M., & Löwe, B. (2015). Assessing somatic symptom burden: a psychometric comparison of the patient health questionnaire-15 (PHQ-15) and the somatic symptom scale-8 (SSS-8). Journal of Psychosomatic Research, 78(4), 352–5. http://doi.org/10.1016/j.jpsychores.2014.11.006 Klaus, K., Rief, W., Brähler, E., Martin, A., Glaesmer, H., & Mewes, R. (2015). Validating psychological classification criteria in the context of somatoform disorders: A one- and four-year follow-up. Journal of Abnormal Psychology, 124(4), 1092–101. http://doi.org/10.1037/abn0000085 Laferton, J. A. C., Stenzel, N. M., Rief, W., Klaus, K., Brähler, E., & Mewes, R. (2017). Screening for DSM-5 Somatic Symptom Disorder: Diagnostic Accuracy of Self-Report Measures Within a Population Sample. Psychosomatic Medicine, 79(9). http://doi.org/10.1097/PSY.0000000000000530 Rief, W., & Martin, A. (2014). How to use the new DSM-5 somatic symptom disorder diagnosis in research and practice: a critical evaluation and a proposal for modifications. Annual Review of Clinical Psychology, 10, 339–67. http://doi.org/10.1146/annurev-clinpsy-032813-153745 Toussaint, A., Löwe, B., Brähler, E., & Jordan, P. (2017). The Somatic Symptom Disorder - B Criteria Scale (SSD-12): Factorial structure, validity and population-based norms. Journal of Psychosomatic Research, 97(June), 9–17. http://doi.org/10.1016/j.jpsychores.2017.03.017 Toussaint, A., Murray, A. M., Voigt, K., Herzog, A., Gierk, B., Kroenke, K., … Lowe, B. (2015). Development and Validation of the Somatic Symptom Disorder-B Criteria Scale (SSD-12). Psychosomatic Medicine, 78, 5–12. http://doi.org/10.1097/PSY.0000000000000240 Tu, C.-Y., Liao, S.-C., Liu, C.-Y., Chen, T.-T., Chen, I.-M., Lin, K.-F., & Huang, W.-L. (2016). Application of the Chinese Version of the Whiteley Index-7 for Detecting DSM-5 Somatic Symptom and Related Disorders. Psychosomatics, 57(3), 283–291. http://doi.org/10.1016/j.psym.2015.12.010

REVIEWER	Katarzyna Nowicka-Sauer Department of Family Medicine, Medical University of Gdańsk, Poland
REVIEW RETURNED	16-Jul-2018

GENERAL COMMENTS	This is a very interesting study design and in my opinion the attempt to create a method allowing to assess somatization is really needed. However, I have many major concerns related to the manuscript itself as well as the way in which the study design is presented. First of all, in many places English is very poor and it makes the reader confused, since he/she is not able to recognize/understand

	the authors' aims, intentions, and way of understanding the problem.…. For example in manuscript title it is said: "severity evaluation of a somatic symptom questionnaire" – it is not a "severity of questionnaire" assessed, but "severity of symptoms". Many information are not clear and not clearly specified/described or described differently in different sections of the manuscript – these make the study design not clear for the readers. The other concern is that it is proved that psychotherapy is an effective method of treatment for somatization. Would it be considered ? Why the only intervention planned is biological treatment with the use of drugs? The authors use the term "gold standard" for physicians' diagnosis. I have doubts if it can be called with this term… The whole manuscript should be checked according to SPIRIT checklist for study protocols (including interventional studies). Not every required points are met. Please find some detailed comments: Title: According to the SPIRIT checklist title: "Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym" – I think the current title does not meet these criteria. See also the above comment related to title. Abstract: Abstract needs improvement. It is not prepared according to BMJ Open guidelines for abstract, e.g.:, it should be section Methods and analysis not Methods/Design. Additional comments to abstract: 1. PHQ-15, PHQ-9 GAD-7 are not "checklists" – they are questionnaires. 2. Severity assessment accuracy of SSD – does it mean one of the outcomes would be assessment of the accuracy of SSS-Ch in the severity of symptoms ? it is not clear… 3. It is said that primary care physician will diagnose patients and that the study would be conducted in three "centres". What does it mean "centres"? Are they primary care outpatients clinics? In the text we can see also "outpatients in internal medicine" and "3 sites". Please specify the setting in which the study is/would be continued. Strenghts and limitations 1. Page 4 of 22, line 5 and 13. the first bullet point as well as the second are not in accord with study protocol of this type of publication. They are in fact results (point 1) and conclusions (point 2). According to Editors' instruction: "There should be no results or conclusions present in the study protocol". This section should relate to methods mainly. Introduction It is said that to date no questionnaire evaluates somatic symptoms. However PRIME-MD includes the list of symptoms and allows the preliminary diagnosis of somatoform disorders in primary care. Please comment on this. In fact you have cited the PRIME-MD study in your text. More current literature should be cited (2016-2018). Study objectives Page 7 of 22, line 42: in which circumstances SSD is accompanied by anxiety and depression. What "circumstances" do you mean? Would they be some correlates ? Which variables? Clinical state? Please specify. Study overview (page 8 of 22)

	"Prospective interventional diagnostic design conducted in internal medicine department in 3 sites of tertiary hospitals". Confusion: what is the study specification? I understand that is it prospective and interventional, however what does mean "diagnostic" ? This is not clear who would evaluate and diagnose the patient because in abstract it is primary care physician and here we see "internal medicine department". Is this department outpatient clinic? Does it employ primary care physicians? Again, please specify. Study design: 1. page 9 of 22: "corresponding examinations" – which examination will be performed? How the "systematic disease" would be excluded? Inclusion criteria: patients who agree to complete checklists. What are checklists? Are they PHQ questionnaires? The word checklist should not be used because these questionnaires are standardized methods. Intervention: In abstract only SSRI and SNRI are mentioned. Here we can see also SARI. Please specify treatment agents. Page 10 of 22 line 41: "SF-20 survey will be conducted as the healthy reference to evaluate patient status". SF-20 is used to evaluate self-reported quality of life but not "patient status"…again be specific what would be the aim of using SF-20 ? Would SF-20 be given at baseline assessment. Follow-up Page 10 of 22, Line 35: "Face to face …for patients taking medicines". Only these patients would be reassessed? Why patients not taking drugs would not be assessed twice ? References: The references do not include up-to-date literature. We can see articles from 1994, 1996, 1999, 2001 and the newest from 2014. Please include current literature. Is any of the references systematic review? Statistical analysis In my opinion the description of statistical methods should be more detailed and contain specific tests which are to be used. For example: Which test will be used to check the normality of the variables? In some cases median is required not mean, according to the results of normality checking. Thus I think professional statistical review would be demanded. Page 12 of 22 line 12: not only number but also percentage of patients ? In all the questionnaires used ? Because you mentioned "questionnaire"? As far as I know Spearman's correlation can be used in case of normal distribution of variable. What if the distribution would not be normal… ? The description opf statistical analysis should be closely related to the study objectives described earlier. Page 12 of 22, line 29: AUC, ROC - abbreviations, please give full name of the methods. Patient and public involvement statement Outcome measures informed by patients' priorities, experience and preferences" – what do you mean? It is not clear… Special clinic prepared, patient privacy protection – again what do you mean? It is not clear… Thank you for an opportunity to review this manuscript.

REVIEWER	Francesca Chappell University of Edinburgh, United Kingdom
REVIEW RETURNED	02-Oct-2018

GENERAL COMMENTS	The authors have written a protocol for a diagnostic study to test the performance of a scale to measure somatic symptoms in a Chinese population. Please note that my review is restricted to the statistical aspects of the paper as I am not a clinician. There were several things I liked about this paper - for example, the presence of a sample size calculation, the blinding of the physician (gold standard) to the results of the SSD scale - and I hope the authors feel that my comments will help to improve it further. In no particular order: 1. The authors have not mentioned the STARD statement (http://www.equator-network.org/reporting-guidelines/stard/). Although there is not (yet) a protocol version of the STARD statement, I would like the authors to say that they will report their study according to STARD. 2. The primary outcome will be "diagnosis and severity assessment of SSD" and they plan to use Spearman's correlation coefficient to assess severity assessment between the SSD scale and the physician's assessment. I think this is less than ideal. The authors do not say which software they will be using for the analysis (this should be added), but have they considered R? Please https://web.expasy.org/pROC/ for free software to compare ROC curves ("statistical tests based on U-statistics or bootstrap"). Please also note the list of useful references at the bottom of the webpage. 3. On page 12 lines 12-15 the authors say, "We will compute the mean (SD) questionnaire scores and the number of patients (%) in each diagnostic category as descriptive statistics". As these are not interval data, the median and interquartile range would be more appropriate. 4. On pages 15 (line 56) - 16 (lines 3-7), the authors say, "Fortunately, each subject in our study will undergo the same set of questionnaires for the entire scale, so the missing samples who are lost to follow-up will not differ among the groups and therefore will not produce significant bias and will not affect our assessment." I disagree. Lost to follow-up is very likely to change the disease spectrum in the sample and will therefore affect the diagnostic performance of the scale and hence be a source of bias. See https://www.bmj.com/content/353/bmj.i3139. The data are unlikely to be MCAR or MAR. The authors need to add a plan to compare people who are and are not lost to follow-up and attempt to ascertain how the consequent bias has affected their results. 5. The authors propose to use a "state of the art" method to impute data (page 12 lines 55-56) and cite reference 16 in support. I had a look at reference 16, and it proposes hot deck imputation for people who use SPSS. Hot deck imputation is not "state of the art", and is not preferable to multiple imputation. Reference 16 simply says that hot deck imputation (which is a type of single imputation) can be better than listwise deletion but is not as sophisticated as some other methods, such as multiple imputation. At the time of writing reference 16 (2011), multiple imputation was not available in SPSS. Please do not use hot deck imputation. Please consider if the

	assumptions of imputation are met (MAR), and if they are, use multiple imputation. The authors have not said which software they are using for the analysis, but I can recommend R and its packages for multiple imputation, please see https://www.analyticsvidhya.com/blog/2016/03/tutorial-powerful-packages-imputing-missing-values/. R is freely available from https://cran.r-project.org/. 6. On page 7, lines 39-40, the authors says that they "intend to evaluate whether the current cut-off value needs to be optimized". How? They need to read this: http://clinchem.aaccjnls.org/content/54/4/729 and refrain from data-driven methods to derive cut-offs. My preference would be to pre-specify cut-offs and assess their performance (sensitivity, specificity with confidence intervals). 7. The authors intend to see whether the SSD scale is "effective in monitoring treatment efficacy". I'm not sure how they are going to do this. They will ask patients to redo the questionnaire, but how this equates to monitoring treatment efficacy I do not know. Are they assuming that if people's scores become better over the 2, 6, and 10 week follow-up assessment that it means the scale can be used for monitoring? Does that mean they are also assuming that the SSRIs/SNRIs work for SSD? Supposing there is no change, does that mean the SSD scale is not good for monitoring or that the patient is unresponsive to drug therapy? What about regression to the mean? To be sure that any changes seen in these patients is a true change, they will need to get the patients reassessed by a physician. 8. How will the three centres be accounted for in the analysis? 9. Will the patients complete the SSD-Ch questionnaire on the same day as being assessed by the physician? 10. Will they be using logistic regression to adjust for age and sex (page 12 lines 55-56)? Please give details.

VERSION 1 – AUTHOR RESPONSE

Detailed Responses to Reviewer #1

Patients with physical discomfort presenting to internal medicine departments in a tertiary hospital in China will be recruited after a negative physical exam. They will successively fill out the SSS-Ch (and other self-report measures for the sake of validity estimation) and a clinical interview to diagnose SSD according to the criteria of DSM-5. The reviewer commends the efforts of the authors to tackle the important issue of providing (time) effective screening tools to identify patients at risk of SSD by conducting such a large scale study. However, the manuscript at its' current form lacks precision related to the description of the study rationale, the previous evidence on diagnostic tools to identify SSD and most importantly related to study methodology. Several aspects of methodology are missing or are unprecise according to the standardized reporting recommendations of diagnostic accuracy studies (STARD 2015 guidelines; (Cohen et al., 2016). The most pressing of which is the lack of detail in describing the reference standard. Therefore, this manuscript needs major revisions to better inform about the study rational and methodology. Detailed points can be seen below. I hope these might aid the authors in order to improve the presentation of their important scientific undertaking.

General:

Point 1: The reviewer is not a native English speaker himself and would therefore like refer the judgment to a native speaking reviewer or editor. However, from my judgment, it might be possible that some of the lacking precision might be improved by English language editing.

Response:

English-language and grammar editing have been performed by American Journal Experts. We hope our current version provides clear information to the reader.

AMERICAN JOURNAL EXPERTS

EDITORIAL CERTIFICATE

This document certifies that the manuscript listed below was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts.

Manuscript title:
Research Protocol for a Diagnostic Study: Identify and Measure Severity of Somatic Symptom Disorder using Self-reported Somatic Symptom Scale-China (SSS-Ch)

Authors:
Meng Jiang

Date Issued:
November 12, 2018

Certificate Verification Key:
001A-826B-BFCE-1A76-3A1A

This certificate may be verified at www.aje.com/certificate. This document certifies that the manuscript listed above was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts. Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final AJE edited version, please visit our verification page. If you have any questions or concerns about this edited document, please contact American Journal Experts at support@aje.com.

American Journal Experts provides a range of editing, translation and manuscript services for researchers and publishers around the world. Our top-quality PhD editors are all native English speakers from America's top universities. Our editors come from nearly every research field and possess the highest qualifications to edit research manuscripts written by non-native English speakers. For more information about our company, services and partner discounts, please visit www.aje.com.

Introduction:

Point 2: First paragraph. It appears unclear whether the authors are referring to the precisely defined term of Somatic Symptom Disorder or to a more general concept i.e. somatoform symptoms, medically unexplained symptoms and the like. P5L8-19 rather describe a broader concept and not specifically the criteria of SSD in DSM-5. Further, the epidemiological references partly refer to different (older) conceptualizations (L19-25). Moreover, there is no SSD in ICD-10 (L19).

Response:

Thank you very much for helping us improve our manuscript. The previous version of our paper did not express the definition clearly. We now state that the SSD concept is a broader concept in the Introduction: SSD refers to symptoms that are often difficult to explain after adequate evaluation; even when significant medical disease is present, the patients' symptoms may nonetheless be unrelated to their disease. Diagnosis of SSD emphasizes the existence of positive symptoms and signs (one or multiple somatic symptoms plus abnormal thoughts, feelings, and behaviours in response to these symptoms).

The SSD that we refer to includes both SSD in the DSM-5 (300.82 (F45.1)) and unspecified somatic symptom and related disorder (300.82 (F45.9)). To enable the reader to obtain a clear idea of SSD, we added Supplementary Figure 1 to illustrate this complex diagnosis. Regarding the epidemiological references, we updated the data according to the DSM-5. The disorder has an estimated current prevalence in the general population of 5-7%. Individuals with somatic symptoms are commonly encountered in general hospitals and primary care as well as in psychiatric and other mental health settings.

Point 3: (P5L25-29) The authors state that there is a lack of effective screening instruments for SSD. This does not reflect the current state of the literature. (Laferton et al., 2017) have provided diagnostic accuracy estimates of identifying SSD using the screening self-report questionnaires asking about physical and psychological symptoms (PHQ-15, the WI-7 and the SAIB). (Gierk et al., 2014, 2015) and (A Toussaint et al., 2015; Anne Toussaint, Löwe, Brähler, & Jordan, 2017) have assessed validity and reliability for the SSS-8 and the SSD-12 for assessing somatic symptoms and psychological symptoms of SSD respectively. (Tu et al., 2016) have reported on using the WI-7 in screening for SSD. Putting the current study into the context of what is known so far would be helpful to get a clearer understanding of the exact study rationale. It appears to the reviewer, that the authors do want to assess the utility of a screening tool that assesses physical symptoms AND psychological symptoms and whether the utility is greater than just assessing physical symptoms. If so, it would clearly inform the reader to state the up to date knowledge on studies evaluating both psychological and physical symptom screenings for the detection of SSD.

Response:

Thank you again for helping us to comprehensively review the current research status. We added all the above-mentioned references to our manuscript in the Introduction (paragraph 2): A series of studies has focused on this issue, using various self-reported questionnaires asking about either physical or psychological symptoms to screen for SSD (1-6). Laferton et al. used the Patient Health Questionnaire 15-item somatic scale (PHQ-15), the Whiteley Index-7 and the Scale for the Assessment of Illness Behaviour questionnaires to identify SSD (1). The Somatic Symptom Scale-8 and Somatic Symptom Scale-12 have been used to assess the validity and reliability of somatic symptoms and the psychological symptoms of SSD, respectively (2-5). Tu et al. have reported using the Whiteley Index-7 to screen for SSD (6). Then, we proposed that our SSS-Ch is characterized by the combined assessment of physical and psychological symptoms, which might make it more accessible and time saving for clinicians. Based on the published studies, we aim to develop a self-administered questionnaire to provide a more comprehensive reflection of the true clinical picture than can be achieved by assessing the somatic complaints alone. Our Somatic Symptom Scale-China (SSS-Ch) integrates somatic symptoms with depression and anxiety items. It incorporates affective, cognitive, and behavioural components. It is designed to be used in general medical facilities and to provide a tool for clinicians to quickly detect a suspected somatic burden. It aims to establish a more accessible and time-saving way to assess the status of subjects.

References

1 Laferton, J. A. C., Stenzel, N. M., Rief, W., Klaus, K., Brähler, E., & Mewes, R. (2017). Screening for DSM-5 Somatic Symptom Disorder: Diagnostic Accuracy of Self-Report Measures Within a Population Sample. Psychosomatic Medicine, 79(9). http://doi.org/10.1097/PSY.0000000000000530

2 Gierk, B., Kohlmann, S., Kroenke, K., Spangenberg, L., Zenger, M., Brähler, E., & Löwe, B. (2014). The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Internal Medicine, 174(3), 399–407. http://doi.org/10.1001/jamainternmed.2013.12179

3 Gierk, B., Kohlmann, S., Toussaint, A., Wahl, I., Brünahl, C. a, Murray, A. M., & Löwe, B. (2015). Assessing somatic symptom burden: a psychometric comparison of the patient health questionnaire-15 (PHQ-15) and the somatic symptom scale-8 (SSS-8). Journal of Psychosomatic Research, 78(4), 352–5. http://doi.org/10.1016/j.jpsychores.2014.11.006

4 Toussaint, A., Murray, A. M., Voigt, K., Herzog, A., Gierk, B., Kroenke, K., … Lowe, B. (2015). Development and Validation of the Somatic Symptom Disorder-B Criteria Scale (SSD-12). Psychosomatic Medicine, 78, 5–12. http://doi.org/10.1097/PSY.0000000000000240

5 Toussaint, A., Löwe, B., Brähler, E., & Jordan, P. (2017). The Somatic Symptom Disorder - B Criteria Scale (SSD-12): Factorial structure, validity and population-based norms. Journal of Psychosomatic Research, 97(June), 9–17. http://doi.org/10.1016/j.jpsychores.2017.03.017

6 Tu, C.-Y., Liao, S.-C., Liu, C.-Y., Chen, T.-T., Chen, I.-M., Lin, K.-F., & Huang, W.-L. (2016). Application of the Chinese Version of the Whiteley Index-7 for Detecting DSM-5 Somatic Symptom and Related Disorders. Psychosomatics, 57(3), 283–291. http://doi.org/10.1016/j.psym.2015.12.010

Point 4: (P5L49-54) “The DSM 5 is currently the “gold standard” for the diagnosis of SSD,….” I am not sure if this is entirely correct. The DSM 5 is the entity defining the new diagnosis of SSD. It is not, however, a specific assessment instrument (e.g. interview or questionnaire) and can therefore not be the gold standard for assessment. This also relates to P6L3-8, where the authors state that the DSM-5 includes an interview longer than half an hour. To my knowledge, the DSM-5 section on SSD does not give specific instructions on how to conduct an interview to diagnose SSD. There are several published interviews to diagnose SSD based on the DSM-5 criteria, including the SCID for DSM5 interview Module J or an interview published by (Axelsson, Andersson, Ljótsson, Wallhed Finn, & Hedman, 2016). Do the authors refer to one or any other interview at this part of the introduction?

Response:

We removed the phrase “gold standard”. As the reviewer said, it is not appropriate. Instead, we use the special physician assessment as the reference standard based on the DSM-5 criteria (Methods – Reference standard): As in Axelsson et al., judgement by a physician is set as the reference standard to test consistency. The physician team is composed of both general hospital “specified physicians” (that is, physicians qualified as national psychological counsellors) and psychologists. The status of the subject will be assessed by the physician or psychologist using the DSM-5 SSD criteria (SSD, 300.82 (F45.1), and unspecified somatic symptom and related disorder, 300.82 (F45.9)) (Supplementary Figure 1), anxiety disorder criteria and depression disorder criteria. In addition, we removed the phrase "DSM-5 includes an interview longer than half an hour”, again since it is not accurate. In the Methods – Reference standard section, we referred to the paper of Axelsson et al (7).

Reference

7. Axelsson, E., Andersson, E., Ljótsson, B., Wallhed Finn, D., & Hedman, E. (2016). The health preoccupation diagnostic interview: inter-rater reliability of a structured interview for diagnostic assessment of DSM-5 somatic symptom disorder and illness anxiety disorder. Cognitive Behaviour Therapy, 45(4), 259–269. http://doi.org/10.1080/16506073.2016.1161663

Point 5: (P6 L11-13) “The most available questionnaires are the PHQ-9 and GAD-7.” I am not entirely sure what the authors do wish to express? Do they mean the most available to assess depressive and anxiety symptoms? If so it might be an even stronger argument to state that their extensively psychometrically evaluation.

Response:

Yes, we mean that the PHQ-9 and GAD-7 are the most available to assess depressive and anxiety symptoms. We have further discussed the current questionnaires in relation to SSD assessment; thus, we removed the introduction of the PHQ-9 and GAD-7.

Point 6: (P6L20) The SSS-8 is “a simplified version” of the PHQ-15. I suggest using “abbreviated” instead of simplified. Moreover, since this part of introduction serves to justify the authors use of the respective questionnaires it might be informative to the reader to state the current evidence regarding those questionnaires in screening for SSD (the research question of this manuscript). (see Point 3)

Response:

We rewrote the introduction section, and we added all of the above-listed references (including SSS-8) to our manuscript in the Introduction (paragraph 2). Then, we proposed our SSS-Ch idea. We believe the revised version more clearly outlines our study aim.

Point 7: (P6L33-P7L44) The authors state that the newly developed SSS-Ch is designed to aid SSD diagnosis, specifically by including items assessing depressive and anxiety symptoms. The improvement of screening tools is an important endeavor, especially for a new diagnostic category like the SSD. 1) However, the specific goal of the research outlined in the manuscript does not become entirely clear from this section of the manuscript. The authors state that depression and anxiety is frequently co-occurring in patients with SSD. This might suggest, that the SSS-Ch aims to assess patients SSD diagnosis AND co-morbid anxiety or depressive disorders? However, that is not explicitly state here. Also, in that case the reference standard would have to include anxiety and depression related diagnosis. 2) The alternative research focus, I assume, would be to improve screening for SSD by including psychological criteria into the SSS-Ch to test whether this improves screening accuracy over and above screening for somatic symptoms (e.g. SSS-8 or PHQ-15). In that case, the argument of the authors would appear stronger, if they would include current evidence on such research (Point 3 and Point 6). 3) Further, the SSS-Ch includes only one item (#17) that appears to assess SSD specific anxiety, the other psychological items refer to depressive or anxiety related symptoms that are more commonly found in other DSM-5 disorders. It would be interesting to read the authors reasoning how these more general anxiety or depressive symptoms are expected to increase SSD screening accuracy. This might still be a valid research hypothesis, yet the reader might expect why the authors do not focus on psychological symptoms of SSD as formulated in the DSM-5 or expressed in other research on this topic (e.g. Klaus et al., 2015; Rief & Martin, 2014). Assuming that improving SSD screening by adding psychological items is the authors research focus, wouldn’t the primary objective be to assess whether the SSS-Ch (including psychological items) outperforms the PHQ-15 (including only somatic symptoms)? 4) Further, it is not clear to me which hypothesis the authors formulate by primary objective 2: ‘We expect to use the SSS-Ch for measuring SSD severity.

Response:

We apologize for not expressing our SSD concept and our study goal more clearly.

1) Our SSD is more of a general concept. It combines somatic symptoms and related disorders and somatoform disorders. The SSS-Ch is also designed to evaluate depression and anxiety. We divided the items into categories to diagnose somatic symptoms (items 1, 5, 9, 10, 12, 13, 16, and 18-20), anxiety (items 6, 14, 15, and 17), depression (items 3, 4, 7, and 11), and anxiety and depression (items 2 and 8) to depict a clear disorder distribution in our sample subjects. In the current version, we described the SSD in the Introduction, paragraph 4: The SSS-Ch questionnaire was developed based on the DSM-5. Additionally, it simultaneously evaluates depression and anxiety. It introduces illness anxiety disorder, which was previously not included in the DSM-IV. For the first time, an organ-based evaluation is used.

2) Additionally, we use a physician or psychologist assessment as our reference standard to determine whether we can produce a questionnaire that is consistent with a physician assessment. The status of the subject will be assessed by the physician or psychologist using the DSM-5 SSD criteria (SSD, 300.82 (F45.1) and unspecified somatic symptom and related disorder, 300.82 (F45.9)) (Supplementary Figure 1), anxiety disorder criteria and depression disorder criteria. Additionally, as the reviewer suggested, we plan to test whether our SSS-Ch improves screening accuracy or is non-inferior to screening for somatic symptoms (PHQ-15). The protocol flow chart can be found in Figure 2.

3) The reason that we include more general anxiety or depressive symptoms in our SSS-Ch is that in China, the reality is that patients with both physical symptoms and psychological symptoms prefer to go to a general medical hospital rather than a mental health centre. This situation means that clinicians must face more patients with psychological symptoms. Our primary study intention is not to create a questionnaire that is superior to the current available questionnaires. We will introduce a tool to facilitate our daily clinical work. The tool provides clinicians with an easy-to-use, questionnaire that can be completed quickly to improve the physician's comfort level in screening suspected SSD patients and to refer them to specific doctors. SSD is classified as a "mental" disorder", and it is reasonable to incorporate more psychological symptoms. The specific and non-specific items from the SSS-Ch are intended for this purpose. Later, we will investigate the utility of our SSS-Ch compared with that of the PHQ-15, PHQ9 and GAD-7 to determine whether it makes any additional contribution of non-specific psychological symptoms to the diagnosis of SSD.

4) Primary objective 2: to identify the accuracy of the SSS-Ch questionnaire in measuring the severity of symptoms.

Methods

Point 8: (P8L38) "Since you have been unwell, how often have you been bothered by any of the following problems". The SSS-Ch appear to assess symptoms for a much longer (and more unspecified) time period as other self-assessment questionnaires (e.g. PHQ-15: past 4 weeks). It would be interesting to hear the reasoning behind that choice. Also, the introductory sentence asks about frequency ('how often have you been bothered') but the scoring is using a scaling referring to severity of influence (e.g. 'unbearable'). Is this due to a translation error? Moreover, since the original manuscript on the validation of the SSS-Ch is not available in English, it would be informative to read more on details of the psychometric characteristics. Especially on those that will be used in the current investigation like the severity categories based on percentile ranks.

Response:

We set the time period as longer than 6 months, which corresponds to criterion C in the DSM-5 for SSD. In this version, we re-worded our questionnaire (Figure 1) to make each question more easily understood.

A summary of the original manuscript on the validation of the SSS-Ch appears in the Introduction: The SSS-Ch was composed of 4 dimensions: physical disorder, anxiety disorder, depression disorder, and anxiety and depression disorder. The test-retest reliability was 0.9. The correlation coefficient between each dimension and the total was between 0.76 and 0.88, and the correlation coefficient within dimensions was 0.56-0.70. In Methods: Each question has a score ranging from 1 to 4, corresponding to the problem occurrence frequency, and the total score ranges from 20 to 80. The severity categories are assessed according to the sum of the scores. The SSS-Ch scores range from 20 to 29, 30 to 39, 40 to 59, and ≥60 and represent normal, mild, moderate, and severe SSD, respectively. The selection of these cut-off values takes into account the results of our previous study (a cut-off score of 30 was obtained from the ROC, reaching a sensitivity of 0.97 and a specificity of 0.96) and clinical experience.

Point 9: (P9L5-10). Only patients after a negative physical exam will be included into the diagnostic accuracy study. Although the authors already state that this is a limitation, the limitation might warrant more explicitly. Given that the assessment of absence of somatic disease is unreliable (Dimsdale et al., 2013) this might result in bias related to the epidemiologic results and the screening accuracy results of this study. Moreover, the external validity will be limited to patients after a negative physical exam. This is underlined by the assumption referred to in the power calculation (P11L48) where the authors expect a huge prevalence of SSD (76.9%).

Response:

Yes, the reviewer is correct about the population of the study. Our study population is different from common outpatients. Only patients without a positive physical exam will be referred to the special clinic, and the referral bias does exist due to the nature of our clinic. This is the reason that we have a high prevalence of SSD. Thus, we cannot determine any epidemiologic results from our study and cannot extend our results to patients whose SSD is accompanied by organic disease. However, it is the population that we want to observe, that is, it is consistent with our daily clinical work. We explicitly describe this issue in our Strengths and limitations of this study section: A potential limitation of this study is that it represents the efficacy of the SSS-Ch only in patients without organic disease. Therefore, the further application of the SSS-Ch to patients with specific diseases should be separately investigated. Since only patients without a positive physical examination will be referred to the special clinic, a referral bias exists due to the nature of our clinic. Moreover, the epidemiology of health care facilities is different from that of general hospitals; therefore, the diagnostic accuracy in a health care sample needs additional investigation.

Point 10: The authors do not mention an informed consent procedure.

Mention informed consent procedure

Response:

We have added the informed consent procedure to the article:

Assessing capacity and obtaining informed consent

Informed consent will be sought by a trained researcher who will provide all necessary information about this study to the potential participants. It will be made clear to participants that they are under no obligation to take part, their usual care will not be affected by their decision and they can withdraw consent without giving a reason. Participants will be given a sheet with contact details for the research team and instructions on what to do if they wish to withdraw or require further information.

Point 11: Exclusion criteria (P9L44-54) '4) patients who are unable to complete at least 1 follow up.' How is that predetermined before the inclusion and especially the follow up takes place?

Response:

We schedule face-to-face interviews at 2, 6, and 10 weeks. Patients who are deemed unable to attend the interviews will be excluded, such as those who live abroad.

Point 12: The most pressing issue is the lack of description of the reference standard. First, it might be helpful to the reader to use the standardized structure of method sections as described by the STARD guidelines. Second, there is only minimal description of the reference standard, which does not allow for replication of the study and does not give information about the reliability or validity of the reference standard.

Response:

Thank you for raising this issue. We added a separate paragraph describing the reference standard in our Methods section. Reference standard: judgement by a physician is set as the reference standard to test consistency. The physician team is composed of both general hospital "specified physicians" (that is, physicians qualified as national psychological counsellors) and psychologists. The status of the subject will be assessed by the physician or psychologist using the DSM-5 SSD criteria (SSD, 300.82 (F45.1) and unspecified somatic symptom and related disorder, 300.82 (F45.9)) (Supplementary Figure 1), anxiety disorder criteria and depression disorder criteria.

Point 13: Patient involvement statement. The authors do state that patients were involved in certain areas of study design, however they do not describe how. Did the authors invite patients to comment on study design and procedure? Did the authors conduct a test phase including patients feedback?

Response:

Patients were involved at the design stage of the trial, including clarifying the understandability of the SSS-Ch questionnaire and discussing the length of the consultation time, the manner of notification of the disease condition, the follow-up method, and the dissemination of the results. Before the formal recruitment started, we received feedback from patients who had SSD during a pretest of the case report form (CRF) and used it to improve the final design of the CRF. We carefully assessed the burden of the trial interventions on patients. We intend to disseminate the main results to the trial participants via email. The study outcomes will be disseminated in conference reports and academic publications.

Discussion:

Point 14: Several points mentioned above do also refer to the discussion section.

Response:

We have revised the above-mentioned points item by item.

Point 15: Figure 1 in my file is a bit blurry and has quite small fond.

Response:

We revised our wording in Figure 1 and adjusted the clarity of the font. We hope the reviewer is satisfied with this version.

Point 16: Several points mentioned above, and especially a clearer more explicit statement about the study rational might also needed to be adapted in the Strengths and limitations statement (P4)

Response:

In the current version, we added the benefit of providing a convenient tool to non-psychological clinicians in general hospitals. First, we introduce a tool to facilitate daily clinical work. The tool provides clinicians with an easy-to-use questionnaire that can be completed quickly and combines both somatic and psychological features to improve physicians' comfort level in screening suspected SSD patients and referring them to specific doctors. Second, our previous study has shown the reliability and factorial validity of the SSS-Ch by utilizing an early version of it. The current study further modifies the SSS-Ch based on the DSM-5 and, for the first time, evaluates its clinical utility. Additionally, we acknowledge our limitations, including the population of our study. A potential limitation of this study is that it represents the efficacy of the SSS-Ch only in patients without organic disease. Therefore, the further application of the SSS-Ch to patients with specific diseases should be separately investigated. Since only patients without a positive physical examination will be referred to the special clinic, a referral bias exists due to the nature of our clinic. Moreover, the epidemiology of health care facilities is different from that of general hospitals; therefore, the diagnostic accuracy in a health care sample needs additional investigation.

Reviewer: 2

This is a very interesting study design and in my opinion the attempt to create a method allowing to assess somatization is really needed. However, I have many major concerns related to the manuscript itself as well as the way in which the study design is presented.

Point 1: First of all, in many places English is very poor and it makes the reader confused, since he/she is not able to recognize/understand the authors' aims, intentions, and way of understanding the problem..... For example in manuscript title it is said: "severity evaluation of a somatic symptom questionnaire" – it is not a "severity of questionnaire" assessed, but "severity of symptoms". Many information is not clear and not clearly specified/described or described differently in different sections of the manuscript – these make the study design not clear for the readers.

Response:

Thank you for your suggestions. We have sent our manuscript to American Journal Experts for dedicated English language editing. We hope the current version clearly expresses the study idea and details.

AMERICAN JOURNAL EXPERTS

EDITORIAL CERTIFICATE

This document certifies that the manuscript listed below was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts.

Manuscript title:
Research Protocol for a Diagnostic Study: Identify and Measure Severity of Somatic Symptom Disorder using Self-reported Somatic Symptom Scale-China (SSS-Ch)

Authors:
Meng Jiang

Date Issued:
November 12, 2018

Certificate Verification Key:
001A-826B-BFCE-1A76-3A1A

This certificate may be verified at www.aje.com/certificate. This document certifies that the manuscript listed above was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts. Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final AJE edited version, please visit our verification page. If you have any questions or concerns about this edited document, please contact American Journal Experts at support@aje.com.

American Journal Experts provides a range of editing, translation and manuscript services for researchers and publishers around the world. Our top-quality PhD editors are all native English speakers from America's top universities. Our editors come from nearly every research field and possess the highest qualifications to edit research manuscripts written by non-native English speakers. For more information about our company, services and partner discounts, please visit www.aje.com.

Additionally, we re-arranged our manuscript to make the points more clearly; for example, we clearly stated our reference standard and placed it in a separate paragraph. We made a more explicit statement about the study rationale. In addition, we provided more detailed information about our previous research using the SSS-Ch.

Point 2: The other concern is that it is proved that psychotherapy is an effective method of treatment for somatization. Would it be considered? Why the only intervention planned is biological treatment with the use of drugs?

Response:

Thank you for raising this issue. As the reviewer said, we agree that psychotherapy is effective. The reason we do not include it is due to the nature of Chinese culture. In our Introduction, paragraph 3, we stated our current medical situation: Differences in medical care across cultures affect the management of these somatic symptoms. Individuals in China usually refuse to receive psychological counselling. Thus, in general medical hospitals, non-psychiatric physicians must face more patients with psychological symptoms. In Methods – Medication: Since patients in China usually refuse to accept psychotherapy, medications will be prescribed according to the physician's evaluation.

Point 3: The authors use the term "gold standard" for physicians' diagnosis. I have doubts if it can be called with this term… The whole manuscript should be checked according to SPIRIT checklist for study protocols (including interventional studies). Not every required points are met.

Response:

We agree that the term “gold standard” is not proper. Instead, we use the term “reference standard”, as other papers have done.

As for the nature of the study, our study is composed of 2 stages (Figure 2). First is the diagnostic performance of our SSS-Ch questionnaire for SSD. The second stage is an exploratory follow-up that uses the SSS-Ch questionnaire as a tool to monitor treatment efficacy. The SSD diagnostic performance is our main emphasis, which aims to evaluate whether the results of the SSS-Ch questionnaire are consistent with physician assessment. The follow-up stage is exploratory and relatively less emphasized in this study. On this basis, as the editor suggested, we used the STARD checklist to standardize our paper. Please find the STARD checklist inserted below.

Section & Topic	No	Item
TITLE OR ABSTRACT		
Page 1, line 1-2	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
Page 2, line 2-18 for protocol article	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
Page 4-5, line 2-21	3	Scientific and clinical background, including the intended use and clinical role of the index test
Page 5, line 22-page 6, line3; Page 6,line 17-page 7, line 3	4	Study objectives and hypotheses
METHODS		
Study design Page 8, line 3; Page 14,line 13-22	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants Page 10,line 1-11	6	Eligibility criteria
Page 9,line 12-16	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
Page 9, line 12-21	8	Where and when potentially eligible participants were identified (setting, location and dates)
Page 9, line 12	9	Whether participants formed a consecutive, random or convenience series
Test methods Page 8,line 12-page 9, line 6; Figure 1	10a	Index test, in sufficient detail to allow replication
Page 10,line 12;Suppl Fig1	10b	Reference standard, in sufficient detail to allow replication
Page 4,line 23-page 5, line 5	11	Rationale for choosing the reference standard (if alternatives exist)
Page 8,line 19-page 9,line 6	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
Suppl Fig1	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
Page 9, line 18-21	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test

Page 9, line 16-18; Page 12, line 1-3	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis Page12, line 15-21	14	Methods for estimating or comparing measures of diagnostic accuracy
Page10, line 19	15	How indeterminate index test or reference standard results were handled
Page 14,line 10-12	16	How missing data on the index test and reference standard were handled
Page12, line 15-21	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
Page 13, line 3-13	18	Intended sample size and how it was determined
RESULTS		
Participants	19	Flow of participants, using a diagram
NA	20	Baseline demographic and clinical characteristics of participants
NA	21a	Distribution of severity of disease in those with the target condition
NA	21b	Distribution of alternative diagnoses in those without the target condition
NA	22	Time interval and any clinical interventions between index test and reference standard
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
NA	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
NA	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
Page 17,line 1-16	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
Page 16,line 2-9;line 18-25; Page 17 line 17-22	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
Page 2,line 24;Page 18,line 23-24 ;	28	Registration number and name of registry
NA	29	Where the full study protocol can be accessed
Page 18,line 15-22	30	Sources of funding and other support; role of funders

Point 4: Please find some detailed comments:

Title: According to the SPIRIT checklist title: “Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym” – I think the current title does not meet these criteria. See also the above comment related to title.

Response:

The STARD checklist provides the following description of the title: “Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC).” Our current title is “Research Protocol for a Diagnostic Study: Identifying and Measuring the Severity of Somatic Symptom Disorder Using the Self-reported Somatic Symptom Scale-China (SSS-Ch)”.

Abstract:

Point 5: Abstract needs improvement. It is not prepared according to BMJ Open guidelines for abstract, e.g.:, it should be section Methods and analysis not Methods/Design.

Response:

We have re-arranged our Abstract according to the BMJ Open guidelines and combined Methods and Analysis:

Methods and Analysis At least patients without organic disease but presenting with physical discomfort will be recruited at a general hospital. Each patient will undergo a DSM-5-guided physician diagnosis, including disease identification and severity assessment, as a reference standard. This research will utilize the SSS-Ch to evaluate its diagnostic performance in SSD compared to that of the Patient Health Questionnaire-15 (PHQ-15) and other SSD-related questionnaires. Statistical tests for the area under the curve (AUC) of the receiver operating curve (ROC) and the Spearman's correlation between two groups will be used to compare the accuracy of the SSD identification and severity assessment. In addition to this standard diagnostic study, we will conduct follow-up investigations to explore the characteristics of the SSS-Ch in monitoring treatment effects.

Additional comments to abstract:

Point 6: PHQ-15, PHQ-9 GAD-7 are not "checklists" – they are questionnaires.

Response:

All the uses of "checklists" were replaced by "questionnaires".

Point 7: Severity assessment accuracy of SSD – does it mean one of the outcomes would be assessment of the accuracy of SSS-Ch in the severity of symptoms? it is not clear…

Response:

Yes. This means assessment of the accuracy of the SSS-Ch in evaluating the severity of symptoms. We apologize for the unclear expression.

Point 8: It is said that primary care physician will diagnose patients and that the study would be conducted in three "centres". What does it mean "centres"? Are they primary care outpatients clinics? In the text we can see also "outpatients in internal medicine" and "3 sites". Please specify the setting in which the study is/would be continued.

Response:

Sorry for our English wording. The study will be conducted in a tertiary general hospital with 3 sites located in different districts in Shanghai. To make it easier for readers to understand the protocol, we do not mention the "3 centres" in our current version. All the doctors involved will be physicians and psychologists in our general hospital.

Strenghts and limitations

Point 9: Page 4 of 22, line 5 and 13. the first bullet point as well as the second are not in accord with study protocol of this type of publication. They are in fact results (point 1) and conclusions (point 2).

Response:

We removed the previous first and second points; instead, we emphasized the potential utility of the questionnaire. First, we introduce a tool to facilitate daily clinical work. The tool provides clinicians with an easy-to-use questionnaire that can be completed quickly and combines both somatic and psychological features to improve physicians' comfort level in screening suspected SSD patients and referring them to specific doctors. Second, our previous study has shown the reliability and factorial validity of the SSS-Ch by utilizing an early version of it. The current study further modifies the SSS-Ch based on the DSM-5 and, for the first time, is applied to evaluate its clinical utility. Third, patients will benefit by improving their awareness of the disease and their ability to self-monitor their symptoms.

Point10: According to Editors' instruction: "There should be no results or conclusions present in the study protocol". This section should relate to methods mainly.

Response:

The results and conclusions are not included in this paper.

Point 11: Introduction

1) It is said that to date no questionnaire evaluates somatic symptoms. However PRIME-MD includes the list of symptoms and allows the preliminary diagnosis of somatoform disorders in primary care. Please comment on this. In fact, you have cited the PRIME-MD study in your text.

2) More current literature should be cited (2016-2018).

Response:

PRIME-MD focused primarily on psychological disorders (emotional items). Our questionnaire items include both physical and psychological symptoms. An organically based evaluation is used for the first time in our questionnaire.

As the reviewer said, it is not appropriate to say that there is no questionnaire that evaluates somatic symptoms. We summarized the current publications related to questionnaires evaluating somatic symptoms in the Introduction, second paragraph: A series of studies has focused on this issue, using various self-reported questionnaires asking about either physical or psychological symptoms to screen for SSD (1-6). Laferton et al. used the Patient Health Questionnaire 15-item somatic scale (PHQ-15), the Whiteley Index-7 and the Scale for the Assessment of Illness Behaviour questionnaires to identify SSD (1). The Somatic Symptom Scale-8 and Somatic Symptom Scale-12 have been used to assess the validity and reliability of somatic symptoms and the psychological symptoms of SSD, respectively (2-5). Tu et al. have reported using the Whiteley Index-7 to screen for SSD (6).

References

1 Laferton, J. A. C., Stenzel, N. M., Rief, W., Klaus, K., Brähler, E., & Mewes, R. (2017). Screening for DSM-5 Somatic Symptom Disorder: Diagnostic Accuracy of Self-Report Measures Within a Population Sample. Psychosomatic Medicine, 79(9). http://doi.org/10.1097/PSY.0000000000000530

2 Gierk, B., Kohlmann, S., Kroenke, K., Spangenberg, L., Zenger, M., Brähler, E., & Löwe, B. (2014). The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Internal Medicine, 174(3), 399–407. http://doi.org/10.1001/jamainternmed.2013.12179

3 Gierk, B., Kohlmann, S., Toussaint, A., Wahl, I., Brünahl, C. a, Murray, A. M., & Löwe, B. (2015). Assessing somatic symptom burden: a psychometric comparison of the patient health questionnaire-15 (PHQ-15) and the somatic symptom scale-8 (SSS-8). Journal of Psychosomatic Research, 78(4), 352–5. http://doi.org/10.1016/j.jpsychores.2014.11.006

4 Toussaint, A., Murray, A. M., Voigt, K., Herzog, A., Gierk, B., Kroenke, K., … Lowe, B. (2015). Development and Validation of the Somatic Symptom Disorder-B Criteria Scale (SSD-12). Psychosomatic Medicine, 78, 5–12. http://doi.org/10.1097/PSY.0000000000000240

5 Toussaint, A., Löwe, B., Brähler, E., & Jordan, P. (2017). The Somatic Symptom Disorder - B Criteria Scale (SSD-12): Factorial structure, validity and population-based norms. Journal of Psychosomatic Research, 97(June), 9–17. http://doi.org/10.1016/j.jpsychores.2017.03.017

6 Tu, C.-Y., Liao, S.-C., Liu, C.-Y., Chen, T.-T., Chen, I.-M., Lin, K.-F., & Huang, W.-L. (2016). Application of the Chinese Version of the Whiteley Index-7 for Detecting DSM-5 Somatic Symptom and Related Disorders. Psychosomatics, 57(3), 283–291. http://doi.org/10.1016/j.psym.2015.12.010

Study objectives

Point 12: Page 7 of 22, line 42: in which circumstances SSD is accompanied by anxiety and depression. What “circumstances” do you mean? Would they be some correlates? Which variables? Clinical state? Please specify.

Response:

We apologize for our wording. We mean: What clinic features of SSD patients would be more likely to be accompanied by anxiety and depression.

Study overview (page 8 of 22)

Point 13: “Prospective interventional diagnostic design conducted in internal medicine department in 3 sites of tertiary hospitals”. Confusion: what is the study specification? I understand that is it prospective and interventional, however what does mean “diagnostic”?

Response:

Our study is composed of 2 stages (Figure 2). The first stage is the performance of our SSS-Ch questionnaire in diagnosing SSD. The second stage is an exploratory follow-up that uses the SSS-Ch questionnaire as a tool to monitor treatment efficacy. The diagnostic performance of our SSS-Ch is our main emphasis, which aims to evaluate whether the results of the SSS-Ch questionnaire are consistent with the physician assessment. The follow-up stage is exploratory and is relatively less emphasized in this study; thus, it would be assessed only in SSD patients. We added these statements to the “Methods – Study Design” section to make the nature of our study more understandable.

Point 14: This is not clear who would evaluate and diagnose the patient because in abstract it is primary care physician and here we see “internal medicine department”. Is this department outpatient clinic? Does it employ primary care physicians? Again, please specify.

Response:

Physician judgement is set as the reference standard. The physician team is composed of both general hospital “specified physicians” (that is, physicians qualified as national psychological counsellors) and psychologists. As we mentioned above, all the doctors involved will be physicians and psychologists in our hospital.

Study design:

Point 15: page 9 of 22: “corresponding examinations” – which examination will be performed? How the “systematic disease” would be excluded?

Response:

The selected corresponding examinations are based on clinical evaluation. For example, a patient with chest pain would receive an EKG, echocardiography, a treadmill test or coronary angiography to exclude coronary artery disease.

Point 16: Inclusion criteria: patients who agree to complete checklists. What are checklists? Are they PHQ questionnaires? The word checklist should not be used because these questionnaires are standardized methods.

Response:

We changed the word “checklists” to “questionnaires”.

Intervention:

Point 17: In abstract only SSRI and SNRI are mentioned. Here we can see also SARI. Please specify treatment agents.

Response:

Serotonin antagonist and reuptake inhibitor (SARI) has been added to the Abstract.

Point 18: Page 10 of 22 line 41: “SF-20 survey will be conducted as the healthy reference to evaluate patient status”. SF-20 is used to evaluate self-reported quality of life but not “patient status”…again be specific what would be the aim of using SF-20 ? Would SF-20 be given at baseline assessment?

Response:

As we mentioned above, the diagnostic performance of our SSS-Ch is our main aim. In this part, we do not use the SF-20. In an exploratory follow-up stage, we intend to explore whether the SSS-Ch questionnaire can serve as a tool to monitor treatment efficacy. At this point, we use the SF-20 as one

of the indicators of therapeutic effect. We have changed the phrase “patient status” to “self-reported quality of life”.

Follow-up

Point 19: Page 10 of 22, Line 35: “Face to face …for patients taking medicines”. Only these patients would be reassessed? Why patients not taking drugs would not be assessed twice?

Response:

Since we set the physician assessment as the reference standard and aim to develop a questionnaire that facilitates clinical work, once the physician considers that a patient does not fit the SSD diagnosis, we will not assess the patient again.

References:

Point 20: The references do not include up-to-date literature. We can see articles from 1994, 1996, 1999, 2001 and the newest from 2014. Please include current literature. Is any of the references systematic review?

Response:

We have updated the referenced literature with current publications (from 2014 to 2017) and added them to the Introduction and Methods sections. The following references have been added.

References

1 Laferton, J. A. C., Stenzel, N. M., Rief, W., Klaus, K., Brähler, E., & Mewes, R. (2017). Screening for DSM-5 Somatic Symptom Disorder: Diagnostic Accuracy of Self-Report Measures Within a Population Sample. Psychosomatic Medicine, 79(9). http://doi.org/10.1097/PSY.0000000000000530

2 Gierk, B., Kohlmann, S., Kroenke, K., Spangenberg, L., Zenger, M., Brähler, E., & Löwe, B. (2014). The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Internal Medicine, 174(3), 399–407. http://doi.org/10.1001/jamainternmed.2013.12179

3 Gierk, B., Kohlmann, S., Toussaint, A., Wahl, I., Brünahl, C. a, Murray, A. M., & Löwe, B. (2015). Assessing somatic symptom burden: a psychometric comparison of the patient health questionnaire-15 (PHQ-15) and the somatic symptom scale-8 (SSS-8). Journal of Psychosomatic Research, 78(4), 352–5. http://doi.org/10.1016/j.jpsychores.2014.11.006

4 Toussaint, A., Murray, A. M., Voigt, K., Herzog, A., Gierk, B., Kroenke, K., … Lowe, B. (2015). Development and Validation of the Somatic Symptom Disorder-B Criteria Scale (SSD-12). Psychosomatic Medicine, 78, 5–12. http://doi.org/10.1097/PSY.0000000000000240

5 Toussaint, A., Löwe, B., Brähler, E., & Jordan, P. (2017). The Somatic Symptom Disorder - B Criteria Scale (SSD-12): Factorial structure, validity and population-based norms. Journal of Psychosomatic Research, 97(June), 9–17. http://doi.org/10.1016/j.jpsychores.2017.03.017

6 Tu, C.-Y., Liao, S.-C., Liu, C.-Y., Chen, T.-T., Chen, I.-M., Lin, K.-F., & Huang, W.-L. (2016). Application of the Chinese Version of the Whiteley Index-7 for Detecting DSM-5 Somatic Symptom and Related Disorders. Psychosomatics, 57(3), 283–291. http://doi.org/10.1016/j.psym.2015.12.010

7 Axelsson E, Andersson E, Ljotsson B, et al. The health preoccupation diagnostic interview: inter-rater reliability of a structured interview for diagnostic assessment of DSM-5 somatic symptom disorder and illness anxiety disorder. Cogn Behav Ther 2016;45(4):259-69. doi: 10.1080/16506073.2016.1161663

Statistical analysis

Point 21: In my opinion the description of statistical methods should be more detailed and contain specific tests which are to be used. For example: Which test will be used to check the normality of the variables? In some cases median is required not mean, according to the results of normality checking. Thus I think professional statistical review would be demanded.

Response:

Professional statisticians have reviewed our statistical plan. After discussion, we decided not to use normality checking for the questionnaire score, as its value is discrete. Instead, we will report it using median (P25, P75) and revised our article:

We will compute the median (P25, P75) scores for each questionnaire…

For our primary analysis, we specified the statistical tests used and revised our article as follows:

The primary analysis of the diagnostic performance will consist of two comparisons using Bonferroni correction:

(1) the non-inferior comparison of the SSS-Ch with the PHQ-15 in SSD diagnostic accuracy as measured by the AUC of the ROC with Δ=0.05, α=0.025 in the whole study population using Delong's method (8) and

(2) the non-inferior comparison of the SSS-Ch with the PHQ-15 in SSD severity assessment measured by Spearman's correlation with Δ=0.1, α=0.025 in the population with a confirmed SSD diagnosis using Fisher's Z test.

Reference:

8 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845.

Point 22: Page 12 of 22 line 12: not only number but also percentage of patients? In all the questionnaires used? Because you mentioned "questionnaire"?

Response:

We agree. We apologize for the unclear expression. We have revised our statistical analysis section as follows:

We will compute the median (P25, P75) scores for each questionnaire and the number and percentage of patients (%) in each diagnostic category as descriptive statistics.

Point 23: As far as I know Spearman's correlation can be used in case of normal distribution of variable. What if the distribution would not be normal… ?

Response:

Spearman's correlation is a non-parametric statistic that can be used when the distribution is not normal, in contrast with Pearson's correlation.

Point 24: The description of statistical analysis should be closely related to the study objectives described earlier.

Response:

We agree, and we have reorganized the statistical analysis section to be consistent with our study objectives.

Point 25: Page 12 of 22, line 29: AUC, ROC - abbreviations, please give full name of the methods.

Response:

We added these full names at their first appearance in our article.

Point 26: Patient and public involvement statement

Outcome measures informed by patients' priorities, experience and preferences" – what do you mean? It is not clear…

Response:

We have re-written the patient and public involvement statement to make the information clear. Patients were involved at the design stage of the trial, including clarifying the understandability of the SSS-Ch questionnaire and discussing the length of the consultation time, the manner of notification of the disease condition, the follow-up method, and the dissemination of the results. Before the formal recruitment started, we received feedback from patients who had SSD during a pretest of the case report form (CRF) and used it to improve the final design of the CRF. We carefully assessed the burden of the trial interventions on patients. We intend to disseminate the main results to the trial participants via email. The study outcomes will be disseminated in conference reports and academic publications.

Point 27: Special clinic prepared, patient privacy protection – again what do you mean? It is not clear…

Response:

We removed that sentence to avoid confusion. We meant to express that the patient will be transferred to the specialist clinic, and the patient can choose to communicate with a physician with or without being accompanied.

Thank you for an opportunity to review this manuscript.

Reviewer: 3

The authors have written a protocol for a diagnostic study to test the performance of a scale to measure somatic symptoms in a Chinese population.

Please note that my review is restricted to the statistical aspects of the paper as I am not a clinician.

There were several things I liked about this paper - for example, the presence of a sample size calculation, the blinding of the physician (gold standard) to the results of the SSD scale - and I hope the authors feel that my comments will help to improve it further.

In no particular order:

We appreciate the comments. The reviewer provides good suggestions.

Point 1: The authors have not mentioned the STARD statement (http://www.equator-network.org/reporting-guidelines/stard/). Although there is not (yet) a protocol version of the STARD statement, I would like the authors to say that they will report their study according to STARD.

Response:

We have added the statement and will report according to STARD.

Statistical analysis

We will report our results according to STARD.

Section & Topic	No	Item
TITLE OR ABSTRACT		
Page 1, line 1-2	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
Page 2, line 2-18 for protocol article	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
Page 4-5, line 2-21	3	Scientific and clinical background, including the intended use and clinical role of the index test

Page 5, line 22-page 6, line3; Page 6,line 17-page 7, line 3	4	Study objectives and hypotheses
METHODS		
Study design Page 8, line 3; Page 14,line 13-22	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants Page 10,line 1-11	6	Eligibility criteria
Page 9,line 12-16	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
Page 9, line 12-21	8	Where and when potentially eligible participants were identified (setting, location and dates)
Page 9, line 12	9	Whether participants formed a consecutive, random or convenience series
Test methods Page 8,line 12-page 9, line 6; Figure 1	10a	Index test, in sufficient detail to allow replication
Page 10,line 12;Suppl Fig1	10b	Reference standard, in sufficient detail to allow replication
Page 4,line 23-page 5, line 5	11	Rationale for choosing the reference standard (if alternatives exist)
Page 8,line 19-page 9,line 6	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
Suppl Fig1	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
Page 9, line 18-21	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
Page 9, line 16-18; Page 12, line 1-3	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis Page12, line 15-21	14	Methods for estimating or comparing measures of diagnostic accuracy
Page10, line 19	15	How indeterminate index test or reference standard results were handled
Page 14,line 10-12	16	How missing data on the index test and reference standard were handled
Page12, line 15-21	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
Page 13, line 3-13	18	Intended sample size and how it was determined
RESULTS		
Participants	19	Flow of participants, using a diagram
NA	20	Baseline demographic and clinical characteristics of participants
NA	21a	Distribution of severity of disease in those with the target condition
NA	21b	Distribution of alternative diagnoses in those without the target condition
NA	22	Time interval and any clinical interventions between index test and reference standard
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
NA	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
NA	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
Page 17,line 1-16	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability

Page 16,line 2-9;line 18-25; Page 17 line 17-22	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
Page 2,line 24;Page 18,line 23-24 ;	28	Registration number and name of registry
NA	29	Where the full study protocol can be accessed
Page 18,line 15-22	30	Sources of funding and other support; role of funders

Point 2: The primary outcome will be "diagnosis and severity assessment of SSD" and they plan to use Spearman's correlation coefficient to assess severity assessment between the SSD scale and the physician's assessment. I think this is less than ideal. The authors do not say which software they will be using for the analysis (this should be added), but have they considered R? Please https://web.expasy.org/pROC/ for free software to compare ROC curves ("statistical tests based on U-statistics or bootstrap"). Please also note the list of useful references at the bottom of the webpage.

Response:

We do use R (version 3.5.1) to perform the statistical analysis. We appreciate the reviewer’s suggestion of the pROC package, as it is helpful for us to test the diagnostic performance.

In our study, the diagnostic performance is two-fold: one is the accuracy of identifying the disease, for which we compare AUCs using Delong’s method (and the “pROC” R package); the other is the accuracy of severity assessment (mild, moderate, or severe), for which we compare Spearman's correlation coefficients using Fisher’s Z-test. We clarify our statistical plan in our revised protocol.

Point 3: On page 12 lines 12-15 the authors say, "We will compute the mean (SD) questionnaire scores and the number of patients (%) in each diagnostic category as descriptive statistics". As these are not interval data, the median and interquartile range would be more appropriate.

Response:

We agree and have revised our protocol as follows:

We will compute the median (P25, P75) questionnaire scores and the number of patients (%) in each diagnostic category as descriptive statistics.

Point 4: On pages 15 (line 56) - 16 (lines 3-7), the authors say, "Fortunately, each subject in our study will undergo the same set of questionnaires for the entire scale, so the missing samples who are lost to follow-up will not differ among the groups and therefore will not produce significant bias and will not affect our assessment." I disagree. Lost to follow-up is very likely to change the disease spectrum in the sample and will therefore affect the diagnostic performance of the scale and hence be a source of bias. See https://www.bmj.com/content/353/bmj.i3139. The data are unlikely to be MCAR or MAR. The authors need to add a plan to compare people who are and are not lost to follow-up and attempt to ascertain how the consequent bias has affected their results.

Response:

We agree, but it is impractical for us to assess the patients who are lost to follow-up.

Our analysis of the follow-up data is exploratory. We add this limitation in our revised article: The potential of monitoring the treatment effect will be affected by loss to follow-up bias due to the unpredictable pattern of loss to follow-up.

Point 5: The authors propose to use a "state of the art" method to impute data (page 12 lines 55-56) and cite reference 16 in support. I had a look at reference 16, and it proposes hot deck imputation for people who use SPSS. Hot deck imputation is not "state of the art", and is not preferable to multiple imputation. Reference 16 simply says that hot deck imputation (which is a type of single imputation) can be better than listwise deletion but is not as sophisticated as some other methods, such as multiple imputation. At the time of writing reference 16 (2011), multiple imputation was not available in SPSS. Please do not use hot deck imputation. Please consider if the assumptions of imputation are met (MAR), and if they are, use multiple imputation. The authors have not said which software they are using for the analysis, but I can recommend R and its packages for multiple imputation, please see https://www.analyticsvidhya.com/blog/2016/03/tutorial-powerful-packages-imputing-missing-values/. R is freely available from https://cran.r-project.org/.

Response:

We use R, and we will use multiple imputation instead of hot deck imputation. We have revised our protocol.

Point 6: On page 7, lines 39-40, the authors says that they "intend to evaluate whether the current cut-off value needs to be optimized". How? They need to read this: http://clinchem.aaccjnls.org/content/54/4/729 and refrain from data-driven methods to derive cut-offs. My preference would be to pre-specify cut-offs and assess their performance (sensitivity, specificity with confidence intervals).

Response:

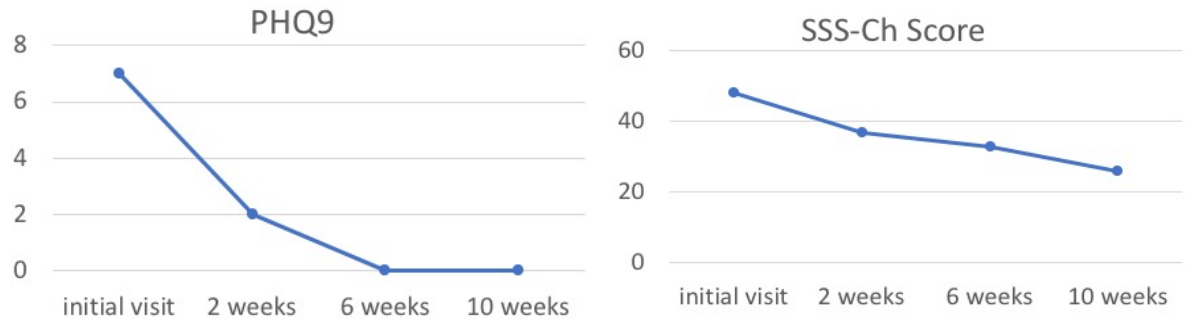
We realized this problem, and our main analysis will use the pre-specified cut-offs from our previous study. As a supplementary analysis, we will also validate this cut-off by comparing it with an optimized cut-off, but we will not draw statistical conclusions regarding the optimal cut-off.

Point 7: The authors intend to see whether the SSD scale is "effective in monitoring treatment efficacy". I'm not sure how they are going to do this. They will ask patients to redo the questionnaire, but how this equates to monitoring treatment efficacy I do not know. Are they assuming that if people's scores become better over the 2, 6, and 10 week follow-up assessment that it means the scale can be used for monitoring? Does that mean they are also assuming that the SSRIs/SNRIs work for SSD? Supposing there is no change, does that mean the SSD scale is not good for monitoring or that the patient is unresponsive to drug therapy? What about regression to the mean? To be sure that any changes seen in these patients is a true change, they will need to get the patients reassessed by a physician.

Response:

The SSRI/SNRI works on both SSD and psychological symptoms (anxiety or depression). Thus, in the real world, most patient feedback positively affects medications. We will observe how the disease behaves over a period of time using continuous changes in the scale of the SSS-Ch and other questionnaires.

Below is a typical example: the PHQ-9 drops dramatically, and can no longer monitor the status change of the patient due to plateau effect; however, we see a stepwise drop of the SSS-Ch scale, which enables us to clearly monitor the patient's change in terms of physical or psychological status.



Point 8: How will the three centres be accounted for in the analysis?

Response:

We apologize for our English wording. The study will be conducted in a tertiary general hospital with 3 sites located in different districts in Shanghai. To make it easier for readers to understand the protocol, we do not mention the 3 centres in our current version. All the doctors involved will be physicians and psychologists in our general hospital.

Point 9: Will the patients complete the SSS-Ch questionnaire on the same day as being assessed by the physician?

Response:

Yes, they will.

Point 10: Will they be using logistic regression to adjust for age and sex (page 12 lines 55-56)? Please give details.

Response:

We will not use logistic regression. We apologize for the misleading terminology "sensitivity analysis" here, we intend to check the heterogeneity of our results in different subpopulations; we rephrased it as "subgroup analysis according to gender and age" in our revised article:

Subgroup analysis according to gender and age will also be conducted.

VERSION 2 – REVIEW

REVIEWER	Dr. Johannes Laferton Philips-Universtiy of Marburg Clinical Psychology
REVIEW RETURNED	22-Nov-2018

GENERAL COMMENTS	Peer Review – BMJ Open – 2018-024290.R1 Title: Somatic Symptom Scale-China (SSS-Ch) study: protocol for measurement and severity evaluation of a self-report version of a somatic symptom questionnaire in a general hospital in China The reviewer thanks the authors for extensive work in revising the manuscript. Several points have been addressed satisfactory. However, two main concerns remain: 1.) it is still not entirely clear how the reference standard will be performed (the most crucial to this study). 2.) The introduction now features up to date research, however, it is still not clear how the actual research questions are derived from these studies. Abstract: Point 1: The aims continue to state, that there are no measures that account for psychological symptoms of SSD. Which is not true as I mentioned in my last peer review. Point 2: Strengths and limitations: Improving the 'physicians comfort level': I am concerned that the current phrasing sends a messages that could be mistaken. Our primary goal as clinicians and researchers, should always be the patient (and their health and comfort level via accurate and at best brief diagnostic tools). Point 3: P4 and 5: DSM-5 is described two times in the introduction which is partially redundant and takes a lot of space. Moreover, it is not the DSM-5 that its used for diagnosis. The DSM-5 just lists the criteria. Diagnosis has to be done following these criteria using structure or unstructured interviews etc… P4L6: do the authors have a reference for the tendency of Chinese people refusing psychotherapy? P5L14-P6L3: The authors now cite the newest screening studies but do not elaborate how these studies relate to their new approach. I.e. they state 'to provide a measure for the more true clinical picture…' but do not argue how other measures fall short of this claim. In the end, it would really help the reader (and reviewer) very much, to get a precise argument of how and why the SSS-Ch now incorporates general anxiety and depression items and why the authors think that would improve diagnosis of SSD? P5L5ff: This is not clear to me? How does the SSS-Ch introduce illness anxiety disorder? What do the authors mean by 'organ based evaluation' Methods: P8L12: 'The SSS-Ch is a somatic symptom scale…' Yet it includes anxiety and depression symptoms? Doesn't that make it a somatic and psychological symptom scale?

	P9L1: Severity categories: The authors list four severity categories. They argue that these are based on the empirical cut off (30). I do not fully understand how 4 categories are based on one cut off score? P10L12ff: Was the physician interview standardized? If yes, that would ad credibility to the reference standard and hence to the paper. If not that is a major limitation to the validity of the reference standard? The authors do cite the Axelsson study. In that study patients were interviewed by a standard procedure (MINI). If this was not the case, the procedure used here- to me – does not appear similar to Axelson. If so, the reference should not be used to justify the current reference standard. Moreover, there is no mentioning of how the reference standard assesses severity. However, that would be the criterion on goal 2. Figure 1: 'You MUST fill all item' How did that coincide with the voluntary nature of the trial, if the participants are 'forced' to fill in all items even if, i.e. the would be ashamed by filling in certain items?

REVIEWER	Katarzyna Nowicka-Sauer Medical University of Gdańsk
REVIEW RETURNED	07-Jan-2019

GENERAL COMMENTS	Dear Authors, the protocol is interesting and such a study is worth to be done. However, many statements are unclear for the reader and still the protocol is not clear. In my opinion clear plan with particular detailed description of examinations and questionnaires should be provided. Second, as it is mentioned in instructions for authors, SPIRIT checklist is appropriate for study protocols instead of STARD. Please check instructions for authors and provide the proper checklist. Moreover, in my opinion professional English editing is required. There are many places in which a reader can be confused, because it is not clear what authors wanted to express or inappropriate English language is used. Please find my detailed comments below: Abstract – two different versions of abstract are available. Strenghts and limitations Page 3, Line 5: SSD – abbreviation without explanation; "specific doctors" – what do you mean ? what kind of doctors? Line 12. 4. "further application of the SSS-Ch in patients with specific diseases should be separately investigated" – language: further application should be performed or further research on...is required... Line 16. "our clinic" – what kind of clinic? Introduction, p. 4. Line 5 "Diagnosis of SSD emphasizes the existence of positive symptoms and signs" – what do you mean by "positive symptoms"? Line 14: English: Patients with somatization had approximately twice as much outpatient and inpatient medical care utilization and annual medical care costs as patients without somatization. Line 19: English: Hence, it is highly important that physicians be prepared to identify SSD

	P. 5; Line 3: “and they emphasize that it is important to evaluate patients in terms of their psychology” English: what do you mean by “evaluate their psychology”? It is not clear statement. Psychology cannot be evaluated... Line 4 “They also incorporate illness anxiety disorder”. What or who are they?... Line 5 “Differences in medical care across cultures affect the management of these somatic symptoms”. Which symptoms ? do you mean symptoms related to SSD ? Line 19: . “The Somatic Symptom Scale-8 and Somatic Symptom Scale-12 have been used to assess the validity and reliability of somatic symptoms and the psychological symptoms of SSD.“ Again, it is not clear. If we assess the reliability and validity we assess the reliability and validity of the methods, not symptoms!!! Page 5, line 2: English: “ to provide a tool for clinicians” it should be to provide clinicians with... Line 6: “For the first time, an organ-based evaluation is used” what do you mean? That it was done for the first time in the literature? Or that first stage of your study was orga-based evaluation...? Line 7: English: “The questionnaire is....can be entirely self-administered by the patient”. Rather : the questionnaire is self-administered. Line 12: English, not clear: The correlation coefficient between each dimension and the total was between 0.76 and 0.88, and the correlation coefficient within dimensions was 0.56-0.70.” What do you mean by “TOTAL”? total score? In fact I don’t understand which correlations you described.. Line 13: Items in the scale assess somatic symptoms (50%, 10/20 items), anxiety (20%, 4/20 items), depression (20%, 4/20 items), and anxiety and depression (10%, 2/20 items). You are describing the details of method and you do the same later in on page 8 – I think this description is redundant and should be deleted from Introduction, Line 18: PHQ-15 can be assessed for assessing SSD, however, without data on clinical state of patients, diagnosed diseases we cannot diagnose somatisation. Line 21: “PHQ-15 for assessing severity” – please provide details – how will you assess severity of SSD using PHQ-15 ? Page 8, line 8: English: “Particular attention will be paid to the appropriate storage of this study”. Storage of data, not storage of study. Page 8, line 10 EMA – abbreviation without explanation. Page 9: Line 9: “The first stage is a prospective diagnostic” – the first stage cannot be prospective. Page 9, line: 13: what do you mean by “corresponding examination? What examination? Who will do it ? what is the aim of this “corresponding examination? Page 9, line 14: “Patients with no organic disease that can account for their discomfort will be considered to have a probability of somatic disorder.” Do you mean “psychosomatic disorder? Or somatic disorder? It should be “to have a probable....disorder” not to have a probability disorder”. Page 9; line 16: “They will successively fill out the SSS-Ch”: what do you maen using word “successively” ? Line 17: and other self-reported instruments for the sake of validity estimation – what do you mean by using words: sake validity ? Page 9; line 22: Why only medication is considered to be received by patients as treatment method? Wouldn’t you refer patients to counselling or psychotherapy? Why ? in fact not every patient with SSD demands medications...

	Page 9, line: 24: will SF-20 be administered to patients only once, at follow-up ? If so, how would it be possible to assess therapeutic effect or change after treatment without preliminary results of quality of life measurement ? Page 10; line 7: what do you mean by “serious mental disorder?” Page 10. Line 13: “As in Axelsson et al6, judgement by a physician is set as the reference standard to test consistency” – English – it is not clear. Page 10. Line 19:English: “When there is diagnostic uncertainty, the senior physician will be consulted with”. The patient would be consulted? Or the diagnosis would be consulted with senior physician? It is not clear... Page 10. Line 20: “Assessing capacity and obtaining informed consent” what do you mean by assessing capacity? Line 21: “Informed consent will be sought by a trained researcher” what do you mean by “will be sought”? The trained researched will give the patients Informed consent...? it is not clear. Page 11; line 3 Blinding: We will also ensure that the patients are comfortable – this fact has nothing common with blinding.... “Then, an initial consultation will be blindly conducted”...what do you mean? If you use term “initial consultation” use it across the manuscript. The names of study stages should be determined at the very beginning and these terms should be used through the whole paper. Using different terms makes the plan unclear and confusing. You did not mention such a stage named “initial consultation” earlier in your manuscript. Page 11, line 13: “The patients will communicate with the doctor throughout the diagnosis and treatment” – what do you mean? That patients would have the possibility to contact with their doctor? We cannot assume that patients WILL communicate. Are they obliged to communicate? Page 23: Line 2: “PHQ-9 and GAD-7” abbreviations without explanation – it is done in line 14, should be here. What is the purpose of using them ? “An SF-20” – grammar: – the SF-20 – is was mentioned earlier. Page 12, Line 9: “The criterion validity will be calculated by the correlations of the diagnostic results and the severity assessments of somatic symptoms between the SSS-Ch and the reference standard.” Rather: The criterion validity will be determined by assessment of... What do you mean by “diagnostic results” and “severity assessments” ? severity of symptoms not severity of assessments. Please specify again what is “reference standard”. Page 13. Line 6: “special clinics” – what kind of clinic ? please specify across the whole paper Line 18: English: “The criterion validity will be measured by the kappa coefficient..” Page 14, line 14: It is not clear:... Patients were involved at the design stage of the trial, including clarifying the understandability of the SSS-Ch questionnaire and discussing the length of the consultation time.. Page 15: English: In November 2018, patient recruitment was not completed. Page 16: ;ine 8: “Moreover, we will examine the characteristics of the SSS-Ch compared with other somatic symptom questionnaires.” Which questionnaires? Line 12: Somatic symptoms may interact with mental items and mental symptoms – items interact with symptoms? What do you mean?

	Page 17, line 4: Moreover, the epidemiology of health care facilities is different from that of general hospitals – what do you mean ? Page 17, Line 14: “each subject in our study will undergo the same set of questionnaires for the entire scale” – it is not clear. Earlier you mentioned that at the first stage patients will undergo PHQ-15 and SSS-CH and other questionnaires (like SF-20, PHQ-9, GAD-7) only at follow-up. Please provide clear plan, on which stages which assessments will be done. Unfortunately, it is not clear even after reading all the manuscript.... Thank you for such an interesting idea of study.

REVIEWER	Francesca Chappell University of Edinburgh, UK
REVIEW RETURNED	12-Dec-2018

GENERAL COMMENTS	I am glad to see that the authors have made changes to the paper which I feel have improved it. However, I do still have some reservations. Please see below. 1. The authors still intend to use Spearman's correlation coefficient to assess the relationship between the SSD scale and the physician's assessment. This is flawed. They say, "The accuracy of the severity assessment of a 20 questionnaire is measured by the Spearman correlation between the questionnaire score and the physician’s severity assessment." No it is not. Please look at this website (which was also published as a paper in the Lancet): https://www-users.york.ac.uk/~mb55/meas/ba.htm about comparing ways of measuring the same thing. Please pay particular attention to the section headed, "Inappropriate use of the correlation coefficient". 2. , "Fortunately, each subject in our study will undergo the same set of questionnaires for the entire scale, so the missing samples who are lost to follow-up will not differ among the groups and therefore will not produce significant bias and will not affect our assessment." I commented on this statement at the first peer review and I still disagree. How can they say, that missing data will not differ among groups? The authors themselves say "The potential of monitoring the treatment effect will be affected by loss to follow-up bias due to the unpredictable pattern of loss to follow-up." I agree with this latter statement (point 6, Strengths and Limitations). 3. It's good to see that they have changed the imputation method from hot deck to multiple imputation. However, I'd also like to see a statement about only conducting multiple imputation if the data are MAR or MCAR. 4. Re the "monitoring treatment efficacy in patients is measured by correlation with the SF-20 during follow-up visits" and the mention of subgroup analyses according to gender and age. I don't agree with the use of correlation to measure treatment efficacy. There are statistical models that can handle repeated follow-up data where not all patients get follow and those who do have follow-up might not have it at the same time. Treatment efficacy could be easily adjusted for age and sex in these models. Also, I still not quite sure what the authors intend to do for the analysis. Please read this paper (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2971698/) which covers repeated measures and missing data.

VERSION 2 – AUTHOR RESPONSE

Detailed Responses to Reviewer #1

Reviewer Name: Dr. Johannes Laferton

Institution and Country: Philips-Universtiy of Marburg

Clinical Psychology

General Comments:

Peer Review – BMJ Open – 2018-024290.R1

The reviewer thanks the authors for extensive work in revising the manuscript. Several points have been addressed satisfactory. However, two main concerns remain: 1.) it is still not entirely clear how the reference standard will be performed (the most crucial to this study). 2.) The introduction now features up to date research, however, it is still not clear how the actual research questions are derived from these studies.

Response:

Thank you for your expert evaluation and positive comments on our manuscript. We have made every possible effort to address the concerns (especially the 2 main concerns that the reviewer addressed), and the changes are highlighted in yellow in the revised manuscript. We feel these revisions have significantly strengthened out manuscript. Detailed responses are below.

Abstract:

The aims continue to state, that there are no measures that account for psychological symptoms of SSD. Which is not true as I mentioned in my last peer review.

Response:

We agree, and we searched the literature for studies you mentioned. We reworded the sentence in the Abstract to read: “Self-report questionnaires that combine both somatic symptoms and psychological characteristics are useful in screening for SSD.”

Strengths and limitations: Improving the ‘physicians comfort level’: I am concerned that the current phrasing sends a message that could be mistaken. Our primary goal as clinicians and researchers, should always be the patient (and their health and comfort level via accurate and at best brief diagnostic tools).

Response:

We agree and have carefully revised our manuscript accordingly. The Strengths and limitations Section now reads as follows: “First, we introduce a tool to benefit patients and to facilitate daily clinical work. The primary goal is to screen suspected somatic symptom disorder（SSD） patients via accurate and brief diagnostic tools. Additionally, the tool provides clinicians with an easy-to-use questionnaire that can be completed quickly and combines both somatic and psychological features.”

P4 and 5: DSM-5 is described two times in the introduction which is partially redundant and takes a lot of space. Moreover, it is not the DSM-5 that is used for diagnosis. The DSM-5 just lists the criteria. Diagnosis has to be done following these criteria using structure or unstructured interviews etc…

Response:

We thank the Reviewer for the constructive suggestions and expert evaluation of our manuscript. We have followed your suggestions to remove the redundant phrases, particularly on DSM-5, and restructure the Introduction section to clarify how our research questions are derived from the referenced studies. “Recent studies, including one by Laferton et al., have also indicated that the combination of self-report measures could increase diagnostic quality in clinical practice[7]. The DSM-5, however, is clinically difficult to follow because it depends on qualified and experienced physicians conducting an interview[8].”… “It is more favourable to have a tool to screen suspected SSD patients via accurate and brief diagnostic questionnaires and to facilitate daily clinical work. A series of studies has focused on this issue; the Patient Health Questionnaire-15 (PHQ-15) and the Somatic Symptom Scale-8 are screening tools for SSD[9 10]; however, these types of self-report questionnaires do not incorporate psychological features. The Whiteley Index-7 focuses on health anxiety[11], the Scale for the Assessment of Illness Behavior questionnaires focuses on excessive illness behaviour, and the Somatic Symptom Scale-12 assesses psychological features[12 13]. The latter three questionnaires focus less on physical features. Based on published studies, we aim to develop a self-administered questionnaire to provide a comprehensive reflection of both somatic and psychological features.”…“ It is designed to be used in general medical facilities and to provide clinicians with an easy-to-use questionnaire to detect both somatic and psychological features” We explained how the structure interview based on DSM-5 will be performed in the Methods section.

References:

1 Axelsson E, Andersson E, Ljotsson B, et al. The health preoccupation diagnostic interview: inter-rater reliability of a structured interview for diagnostic assessment of DSM-5 somatic symptom disorder and illness anxiety disorder. Cogn Behav Ther 2016;45(4):259-69. doi: 10.1080/16506073.2016.1161663

2 Gierk B, Kohlmann S, Kroenke K, et al. The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Intern Med 2014;174(3):399-407. doi: 10.1001/jamainternmed.2013.12179

3 Gierk B, Kohlmann S, Toussaint A, et al. Assessing somatic symptom burden: a psychometric comparison of the patient health questionnaire-15 (PHQ-15) and the somatic symptom scale-8 (SSS-8). J Psychosom Res 2015;78(4):352-5. doi: 10.1016/j.jpsychores.2014.11.006

4 Tu CY, Liao SC, Liu CY, et al. Application of the Chinese Version of the Whiteley Index-7 for Detecting DSM-5 Somatic Symptom and Related Disorders. Psychosomatics 2016;57(3):283-91. doi: 10.1016/j.psym.2015.12.010

5 Toussaint A, Murray AM, Voigt K, et al. Development and Validation of the Somatic Symptom Disorder-B Criteria Scale (SSD-12). Psychosom Med 2016;78(1):5-12. doi: 10.1097/PSY.0000000000000240

6 Toussaint A, Lowe B, Brahler E, et al. The Somatic Symptom Disorder - B Criteria Scale (SSD-12): Factorial structure, validity and population-based norms. J Psychosom Res 2017;97:9-17. doi: 10.1016/j.jpsychores.2017.03.017

P4L6: do the authors have a reference for the tendency of Chinese people refusing psychotherapy?

Response:

We have put the references below and inserted them into the manuscript.

References:

1. Liang D, Mays VM, Hwang WC. Integrated mental health services in China: challenges and planning for the future. Health Policy Plan 2018;33(1):107-22. doi: 10.1093/heapol/czx137

2. Leong FT, Lau AS. Barriers to providing effective mental health services to Asian Americans. Ment Health Serv Res 2001;3(4):201-14.

P5L14-P6L3: The authors now cite the newest screening studies but do not elaborate how these studies relate to their new approach. I.e. they state ‘to provide a measure for the more true clinical picture…’ but do not argue how other measures fall short of this claim. In the end, it would really help the reader (and reviewer) very much, to get a precise argument of how and why the SSS-Ch now incorporates general anxiety and depression items and why the authors think that would improve diagnosis of SSD?

Response:

We thank the Reviewer for the constructive suggestions and followed your suggestions to explain why the SSS-CN questionnaire followed the request of DSM -5 and would help the clinician to improve the diagnosis of SSD. The Introduction Section now reads as follows: “The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) aims to identify SSD patients and assessing the severity of the disorder[2]. It agreed that the SSD companioned anxiety and depressive disorder (in approximately 57.7% of SSD patients)[1] adds severity and complexity to the somatic components. It emphasizes that it is important to evaluate patients in terms of their psychological situation, behaviour and physical condition altogether and then treat the patients according to the severity of the disorder. It also emphasizes the evaluation in subjects who have excessive concerns about health issues. Recent studies, including one by Laferton et al., have also indicated that the combination of self-report measures could increase diagnostic quality in clinical practice[7]. The DSM-5, however, is clinically difficult to follow because it depends on qualified and experienced physicians conducting an interview[8], which makes clinicians in the general hospital feel less confident in dealing with such individuals. On the other hand, individuals in China usually refuse to receive psychological counselling. Thus, in general medical hospitals, non-psychiatric physicians must face more patients with psychological symptoms. It is more favourable to have a tool to screen suspected SSD patients via accurate and brief diagnostic questionnaires and to facilitate daily clinical work.” We also explained how the SSS-CN related to the referenced studies and how the SSS-CN incorporated general anxiety and depression items. The Introduction Section now reads as follows: “A series of studies has focused on this issue; the Patient Health Questionnaire-15 (PHQ-15) and the Somatic

Symptom Scale-8 are screening tools for SSD[9 10]; however, these types of self-report questionnaires do not incorporate psychological features. The Whiteley Index-7 focuses on health anxiety[11], the Scale for the Assessment of Illness Behavior questionnaires focuses on excessive illness behaviour, and the Somatic Symptom Scale-12 assesses psychological features[12 13]. The latter three questionnaires focus less on physical features.

Based on published studies, we aim to develop a self-administered questionnaire to provide a comprehensive reflection of both somatic and psychological features. The Somatic Symptom Scale-China (SSS-CN) questionnaire was developed based on the DSM-5. Psychology and behaviour items are interleaved with somatic symptoms. It incorporates affective, cognitive, and behavioural components. It is designed to be used in general medical facilities and to provide clinicians with an easy-to-use questionnaire to detect both somatic and psychological features in a time-saving way."

References:

1 Axelsson E, Andersson E, Ljotsson B, et al. The health preoccupation diagnostic interview: inter-rater reliability of a structured interview for diagnostic assessment of DSM-5 somatic symptom disorder and illness anxiety disorder. Cogn Behav Ther 2016;45(4):259-69. doi: 10.1080/16506073.2016.1161663

2 Gierk B, Kohlmann S, Kroenke K, et al. The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Intern Med 2014;174(3):399-407. doi: 10.1001/jamainternmed.2013.12179

3 Gierk B, Kohlmann S, Toussaint A, et al. Assessing somatic symptom burden: a psychometric comparison of the patient health questionnaire-15 (PHQ-15) and the somatic symptom scale-8 (SSS-8). J Psychosom Res 2015;78(4):352-5. doi: 10.1016/j.jpsychores.2014.11.006

4 Tu CY, Liao SC, Liu CY, et al. Application of the Chinese Version of the Whiteley Index-7 for Detecting DSM-5 Somatic Symptom and Related Disorders. Psychosomatics 2016;57(3):283-91. doi: 10.1016/j.psym.2015.12.010

5 Toussaint A, Murray AM, Voigt K, et al. Development and Validation of the Somatic Symptom Disorder-B Criteria Scale (SSD-12). Psychosom Med 2016;78(1):5-12. doi: 10.1097/PSY.0000000000000240

6 Toussaint A, Lowe B, Brahler E, et al. The Somatic Symptom Disorder - B Criteria Scale (SSD-12): Factorial structure, validity and population-based norms. J Psychosom Res 2017;97:9-17. doi: 10.1016/j.jpsychores.2017.03.017

P5L5ff: This is not clear to me? How does the SSS-Ch introduce illness anxiety disorder? What do the authors mean by 'organ based evaluation'?

Response:

Item 17 in the SSS-CN questionnaire (Excess concerns about health issues, excessive worry that you or family members are ill}) focuses on the illness anxiety disorder. The organ-based evaluation means that each item (1, 5, 9, 10, 12, 13, 16, and 18-20) represents discomfort from a body system such as the neuro-system in item 1, cardiovascular system in item 5, gastrointestinal system in item 9, etc.)

Methods:

P8L12: ‘The SSS-Ch is a somatic symptom scale…’ Yet it includes anxiety and depression symptoms? Doesn’t that make it a somatic and psychological symptom scale?

Response:

We agree, and this has been done.

P9L1: Severity categories: The authors list four severity categories. They argue that these are based on the empirical cut off (30). I do not fully understand how 4 categories are based on one cut off score?

Response:

We have 3 cut offs to define the patient status from normal to severe condition, which are 30,40,60. However, as the reviewer noted, only one cut-off score (30) was verified by our pilot study. Thus, we need to test the rest of the performance of cut offs in the present study. We mentioned our intention in Outcomes (second paragraph) of Methods: “The criterion validity will be determined by assessment of the presence and severity of SSD between the reference standard (physician assessment based on structure interview) and the SSS-CN questionnaire.”, and in Diagnostic performance (first paragraph) of Methods: “The accuracy of the severity assessment of a questionnaire is measured by the VUS (volume under the surface), which is a multi-class generalization of AUC of ROC between the questionnaire score and the physician’s severity assessment[18]..”

P10L12: Was the physician interview standardized? If yes, that would add credibility to the reference standard and hence to the paper. If not that is a major limitation to the validity of the reference standard? The authors do cite the Axelsson study. In that study patients were interviewed by a standard procedure (MINI). If this was not the case, the procedure used here- to me – does not appear similar to Axelson. If so, the reference should not be used to justify the current reference standard. Moreover, there is no mentioning of how the reference standard assesses severity. However, that would be the criterion on goal 2.

Response:

Yes, patients were interviewed by a standard procedure. A structured clinical interview (SCID-5-CV) according to the corresponding DSM-5 criterion was used. The interview questions include modules from somatic symptoms and related disorders to depression disorder, anxiety disorder, obsessive-compulsive related disorder and sleep-wake disorders. The test time is approximately 30-45 minutes. The physician further assesses the severity based on the number of symptoms specified in excessive thoughts, feeling, or behaviours related to the somatic symptoms or associated health concerns (mild-1 symptom; moderate-two or more of the symptoms; severe-two or more of the symptoms plus multiple somatic complaints). We have followed the reviewer’s suggestion to add the description in the Reference standard Section.

Figure 1: ‘You MUST fill all item’ How did that coincide with the voluntary nature of the trial, if the participants are ‘forced’ to fill in all items even if, i.e. the would be ashamed by filling in certain items?

Response:

We apologize for the miswording when translating the Chinese SSS-CN version into English. According to the inclusion criteria, the subject must consent to participate first, and sample questionnaires will be given to inform the process. When filling in the questionnaires, all the items are required to be filled. We changed the word from MUST to REQUIRED (Figure 1) to make our idea clearly.

Detailed Responses to Reviewer #2

Reviewer Name: Katarzyna Nowicka-Sauer

Institution and Country: Medical University of Gdańsk

General Comments:

Peer Review – BMJ Open – 2018-024290. R1

Dear Authors, the protocol is interesting and such a study is worth to be done. However, many statements are unclear for the reader and still the protocol is not clear. In my opinion clear plan with particular detailed description of examinations and questionnaires should be provided.

Second, as it is mentioned in instructions for authors, SPIRIT checklist is appropriate for study protocols instead of STARD. Please check instructions for authors and provide the proper checklist.

Moreover, in my opinion professional English editing is required. There are many places in which a reader can be confused, because it is not clear what authors wanted to express or inappropriate English language is used.

Response:

We sincerely thank the Reviewer for the constructive suggestions and help in improving our manuscript. We are particularly grateful to the reviewer's positive comment that "the protocol is interesting and such a study is worth to be done." We have made every possible effort to address the concerns, and we highlighted the relevant changes in the revised manuscript.

Second, as for the type of checklist, because our study is not a randomized controlled trial, many of the SPIRIT checklist items do not apply to our study design. The editorial team has requested the STARD checklist for the reporting of diagnostic accuracy studies.

Moreover, the English editing has been completed by American Journal Experts.

Detailed responses are below.

Abstract – two different versions of abstract are available.

Response:

This has been corrected.

Strenghts and limitations

Page 3, Line 5: SSD – abbreviation without explanation; “specific doctors” – what do you mean? what kind of doctors?

Response:

Explanation of SSD has been added. Specific doctors mean that the physicians qualified as national psychological counsellors or psychologists. We explain this in the Reference standard of Participants and Procedure.

Line 12. 4. “further application of the SSS-Ch in patients with specific diseases should be separately investigated” – language: further application should be performed or further research on...is required...

Response:

We now revised the sentence to read: “Further research on the application of SSS-CN in patients with both SSD and diagnosed medical disorders is required.”

Line 16. “our clinic” – what kind of clinic?

Response:

The clinic here refers to the place where qualified physicians (physicians qualified as national psychological counsellors and psychologists) practice medicine.

We added the explanation in the Sample Size Calculation section.

Introduction, p. 4. Line 5 “Diagnosis of SSD emphasizes the existence of positive symptoms and signs” – what do you mean by “positive symptoms”?

Response:

We revised the sentence to read: “The diagnosis of SSD emphasizes the existence of symptoms and signs”.

Line 14: English: Patients with somatization had approximately twice as much outpatient and inpatient medical care utilization and annual medical care costs as patients without somatization.

Response:

We revised the sentence to read: “Patients with somatization had approximately twice as much cost as patients without somatization on medical care utilization and annual medical care.”

Line 19: English: Hence, it is highly important that physicians be prepared to identify SSD

Response:

We revised the sentence to read: “Hence, it is highly important that physicians are prepared to identify SSD…”

P. 5; Line 3: “and they emphasize that it is important to evaluate patients in terms of their psychology” English: what do you mean by “evaluate their psychology”? It is not clear statement. Psychology cannot be evaluated...

Response:

We revised the sentence to[read: “It emphasizes that it is important to evaluate patients in terms of their psychological situation.”

Line 4 “They also incorporate illness anxiety disorder”. What or who are they?...

Response:

It refers to the the DSM-5 criteria.

Line 5 “Differences in medical care across cultures affect the management of these somatic symptoms”. Which symptoms? do you mean symptoms related to SSD?

Response:

Yes, we mean symptoms related to SSD. Due to space limitations, we removed that sentence.

Line 19: “The Somatic Symptom Scale-8 and Somatic Symptom Scale-12 have been used to assess the validity and reliability of somatic symptoms and the psychological symptoms of SSD. “Again, it is not clear. If we assess the reliability and validity we assess the reliability and validity of the methods, not symptoms!!!

Response:

These revisions have been made.

Page 5, line 2: English: “to provide a tool for clinicians” it should be to provide clinicians with...

Response:

This has been revised to read: “to provide clinicians with an easy-to-use questionnaire to detect both somatic and psychological features in a time-saving way.”

Line 6: “For the first time, an organ-based evaluation is used” what do you mean? That it was done for the first time in the literature? Or that first stage of your study was orga-based evaluation...?

Response:

We removed the above sentence and accordingly reworded the sentence in the second paragraph of Description of the SSS-CN and Assessment of Severity: “The SSS-CN assesses 10 somatic clusters that account for 50% of the physical complaints (1 item per body system)”.

Line 7: English: “The questionnaire is....can be entirely self-administered by the patient”. Rather : the questionnaire is self-administered.

Response:

We moved the sentence into Description of the SSS-CN and Assessment of Severity section and have reworded the sentence as follows: “The questionnaire is self-administered with an abbreviated 20-item measure.”

Line 12: English, not clear: The correlation coefficient between each dimension and the total was between 0.76 and 0.88, and the correlation coefficient within dimensions was 0.56-0.70.” What do you mean by “TOTAL”? total score? In fact I don’t understand which correlations you described.

Response:

We revised the description as follows: “The correlation coefficient between each dimension and the total was between 0.76 and 0.88, and the correlation coefficient within dimensions was 0.56-0.70.” Please find the details below: S: somatic symptoms score; A: anxiety score; D: depression score; AD: anxiety and depression score.

表 2　总分与因子、因子与因子间的相关系数（r 值）

因子	总分	S 分	A 分	D 分
S 分	0.88[a]			
A 分	0.84[a]	0.58[a]		
D 分	0.82[a]	0.56[a]	0.70[a]	
AD 分	0.76[a]	0.61[a]	0.56[a]	0.60[a]

注：[a]$P<0.001$

Line 13: Items in the scale assess somatic symptoms (50%, 10/20 items), anxiety (20%, 4/20 items), depression (20%, 4/20 items), and anxiety and depression (10%, 2/20 items). You are describing the

details of method and you do the same later in on page 8 – I think this description is redundant and should be deleted from Introduction,

Response:

We agree, and this revision has been made.

Line 18: PHQ-15 can be assessed for assessing SSD, however, without data on clinical state of patients, diagnosed diseases we cannot diagnose somatisation.

Response:

In our study, patients without systemic disease that can account for their physical discomfort are going to be included. Relevant examinations will be performed. For example, a patient with chest pain would be recommended by a physician to receive an EKG, echocardiography, a treadmill test or coronary angiography to exclude cardiovascular disease. Patients with no organic disease that can account for their discomfort will be considered to have a probable psychosomatic disorder.

Line 21: “PHQ-15 for assessing severity” – please provide details – how will you assess severity of SSD using PHQ-15?

Response:

We provided details on how we assessed the severity of SSD using PHQ-15 in Statistical Analysis. “Severity of PHQ-15 were based on scores, normal (score 0–4), low (score 5–9), medium (score 10–14), and high (score 15–30).” (Kurt Kroenke et al, Psychosomatic Medicine 64:258–266 (2002)

Page 8, line 8: English: “Particular attention will be paid to the appropriate storage of this study”. Storage of the data, not storage of study.

Response:

This revision has been made.

Page 8, line 10 EMA – abbreviation without explanation.

Response:

This revision has been made.

Page 9: Line 9: “The first stage is a prospective diagnostic” – the first stage cannot be prospective.

Response:

If I understand correctly, what you mean is that the sentence has grammer error. We rewrote the sentence as follows: “The first stage is a prospective diagnostic stage to test the diagnostic performance of the SSS-CN questionnaire.”

Page 9, line: 13: what do you mean by “corresponding examination? What examination? Who will do it? what is the aim of this “corresponding examination?

Response:

The aim of the corresponding examination is to exclude organic disease. For example, a patient with chest pain would be recommended by a physician to receive an EKG, echocardiography, a treadmill test or coronary angiography to exclude cardiovascular disease. Patients with no organic disease that can account for their discomfort will be considered to have a probable psychosomatic disorder. These patients will then be transferred to a specialist clinic for the treatment of suspected SSD. We added this description to the Study Design.

Page 9, line 14: “Patients with no organic disease that can account for their discomfort will be considered to have a probability of somatic disorder.” Do you mean “psychosomatic disorder? Or somatic disorder? It should be “to have a probable....disorder” not to have a probability disorder”.

Response:

We have rewritten the sentence as read: “Patients with no organic disease that can account for their discomfort will be considered to have a probable psychosomatic disorder.”

Page 9; line 16: “They will successively fill out the SSS-Ch”: what do you mean using word “successively”?

Response:

We removed the word “successively”; it does not mean anything here.

Line 17: and other self-reported instruments for the sake of validity estimation – what do you mean by using words: sake validity?

Response:

We use other self-reported instruments, including PHQ15, Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder Scale-7 (GAD-7) and SF-20, to verify the structural validity of SSS-CN. To make the protocol more readable, we rephrase our sentence as follows: " we use other self-reported instruments, including PHQ15, Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder Scale-7 (GAD-7) and SF-20, to verify the structural validity of SSS-CN.”

Page 9; line 22: Why only medication is considered to be received by patients as treatment method? Wouldn’t you refer patients to counselling or psychotherapy? Why? in fact not every patient with SSD demands medications.

Response:

We noticed that there are societal and cultural barrier culture differences in response to psychotherapy between Asians and patients from other origins. The Chinese World Mental Health

Survey (2001–02) conducted in Beijing and Shanghai found that only 3.4% of respondents with a psychiatric disorder sought professional help during the previous 12 months (1). Similarly, in a large epidemiologic study conducted in four provinces [Gansu, Qinghai, Shandong and Zhejiang (2001–05)—63004 participants aged 18 years or older in 96 urban neighbourhoods and 267 rural villages], only 8% of individuals with mental disorders sought professional help within the general healthcare setting, and only 5% sought help from mental health professionals (mainly hospital-based psychiatrists) (2). Second, in China, most psychiatric care is provided in specialty mental health hospitals that are not co-located within general hospitals, which increases stigma and results in access barriers. Third, Chinese and Asian Americans are likely drop out and pre-mature termination from psychotherapy service (3). Fourth, another challenging barrier is the shortage of psychiatrists, psychiatric nurses, and counselling and clinical psychologists to provide psychotherapy (4). Specifically, China had only 1.49 psychiatrists per 100 000 people, while on average, middle- and high-income countries worldwide have 2.03 psychiatrists per 100 000 (5). Finally, insurance currently pays for treatment with medication but typically does not support psychotherapy, community recovery services, or preventive care.

References:

1 Shen Y, Zhang M, Huang Y. et al. 2006. Twelve-month prevalence, severity, and unmet need for treatment of mental disorders in metropolitan China. Psychological Medicine 36: 257–68.

2 Phillips M, Zhang J, Shi Q. et al. 2009. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. The Lancet 373: 2041–53.

3 Leong FT, Lau AS. Barriers to providing effective mental health services to Asian Americans. Ment Health Serv Res 2001;3(4):201-14.

4 Liu C, Chen L, Xie B. et al. 2013. Number and characteristics of medical professionals working in Chinese mental health facilities. Shanghai Archives of Psychiatry 25: 277–86.

5 The State Council. 2015a. National Mental Health Working Plan (2015–2020) http://www.gov.cn/zhengce/content/2015-06/18/content_9860.htm, accessed 1August 2016.

Page 9, line: 24: will SF-20 be administered to patients only once, at follow-up? If so, how would it be possible to assess therapeutic effect or change after treatment without preliminary results of quality of life measurement?

Response:

We thank the Reviewer for pointing out this. We noticed this issue and added SF-20 at the first-time consultation. Please find it in our study flow (Figure 2). Clinically, we had the SF-20 survey as the first-time consultation for our recruited subjects.

Page 10; line 7: what do you mean by “serious mental disorder?”

Response:

We mean mental disorder.

Page 10. Line 13: “As in Axelsson et al6, judgement by a physician is set as the reference standard to test consistency” – English – it is not clear.

Response:

We have rewritten this part to introduce our structure interview and the reference standard.

Page 10. Line 19: English: “When there is diagnostic uncertainty, the senior physician will be consulted with”. The patient would be consulted? Or the diagnosis would be consulted with senior physician? It is not clear...

Response:

We rewrote the sentence as follows: “When there is diagnostic uncertainty, the patient will be referred to the senior physician to obtain a diagnosis.”

Page 10. Line 20: “Assessing capacity and obtaining informed consent” what do you mean by assessing capacity?

Response:

We removed the phrase “assessing capacity” to avoid confusion because we do not perform any assessment.

Line 21: “Informed consent will be sought by a trained researcher” what do you mean by “will be sought”? The trained researched will give the patients Informed consent...? it is not clear.

Response:

Yes, we agree and changed the word to “give”.

Page 11; line 3 Blinding: We will also ensure that the patients are comfortable – this fact has nothing common with blinding.... “Then, an initial consultation will be blindly conducted”...what do you mean? If you use term “initial consultation” use it across the manuscript. The names of study stages should be determined at the very beginning and these terms should be used through the whole paper. Using different terms makes the plan unclear and confusing. You did not mention such a stage named “initial consultation” earlier in your manuscript.

Response:

We agree and have removed the sentence “We will also ensure that the patients are comfortable”. Additionally, we thank you for your expert evaluation and positive comments and accordingly define the initial consultation in the Study Design section.

Page 11, line 13: “The patients will communicate with the doctor throughout the diagnosis and treatment” – what do you mean? That patients would have the possibility to contact with their doctor? We cannot assume that patients WILL communicate. Are they obliged to communicate?

Response:

Sorry for the confusion. We mean “During the follow-up consultations, the patients will be allowed to communicate with the doctor throughout the diagnosis and treatment.”

Page 12: Line 2: “PHQ-9 and GAD-7” abbreviations without explanation – it is done in line 14, should be here. What is the purpose of using them? “An SF-20” – grammar: – the SF-20 – it was mentioned earlier.

Response:

Abbreviations have been placed when they were first used and grammar has been corrected. The utility of these questionnaires was described in the Outcome Measures section. The construct validity will be tested by confirmatory factor analysis, comparing the corresponding factors of SSS-CN with the PHQ-15, PHQ-9 and GAD-7.

Page 12, Line 9: “The criterion validity will be calculated by the correlations of the diagnostic results and the severity assessments of somatic symptoms between the SSS-Ch and the reference standard.” Rather: The criterion validity will be determined by assessment of... What do you mean by “diagnostic results” and “severity assessments”? severity of symptoms not severity of assessments. Please specify again what is “reference standard”.

Response:

We rewrote the sentence as follows: “The criterion validity will be determined by assessment of the presence and severity of SSD between the reference standard (physician assessment based on structure interview) and the SSS-CN questionnaire.”

Page 13. Line 6: “special clinics” – what kind of clinic? please specify across the whole paper

Response:

We defined the special clinics as where physicians qualified as national psychological counsellors and psychologists practice medicine. We put the explanation after the term “special clinics”.

Line 18: English: “The criterion validity will be measured by the kappa coefficient.”

Response:

We rewrote the sentence as follows: “The criterion validity will be measured by the kappa coefficient between the questionnaire score and the physician assessment.”

Page 14, line 14: It is not clear:... Patients were involved at the design stage of the trial, including clarifying the understandability of the SSS-Ch questionnaire and discussing the length of the consultation time.

Response:

We rewrote this sentence as follows: “Patients were involved at the design stage of the trial, including ensuring that the content of the SSS-CN questionnaire can be understood, the length of the consultation time, the manner of notification of the disease condition, the follow-up method, and the dissemination of the results are acceptable.”

Page 15: English: In November 2018, patient recruitment was not completed.

Response:

This revision has been made.

Page 16: line 8: “Moreover, we will examine the characteristics of the SSS-Ch compared with other somatic symptom questionnaires.” Which questionnaires?

Response:

This revision has been made. “Moreover, we will examine the characteristics of the SSS-CN compared with other somatic symptom questionnaire (PHQ15).”

Line 12: Somatic symptoms may interact with mental items and mental symptoms – items interact with symptoms? What do you mean?

Response:

We rewrote this to read: “Somatic and mental symptoms may interact,…”

Page 17, line 4: Moreover, the epidemiology of health care facilities is different from that of general hospitals – what do you mean?

Response:

We rewrote this to read: “the epidemiology of primary health care facilities is different from the epidemiology of general hospitals.”

Page 17, Line 14: “each subject in our study will undergo the same set of questionnaires for the entire scale” – it is not clear. Earlier you mentioned that at the first stage patients will undergo PHQ-15 and SSS-CH and other questionnaires (like SF-20, PHQ-9, GAD-7) only at follow-up. Please provide clear plan, on which stages which assessments will be done. Unfortunately, it is not clear even after reading all the manuscript....

Response:

Sorry for the confusion. The subject will complete 5 questionnaires (SSS-CN, PHQ15, PHQ-9, GAD-7 and SF-20) both at the initial consultation and at the week 10 follow-up. At week 2, 6 follow-up, 4 questionnaires will be given (SSS-CN, PHQ15, PHQ-9, GAD-7). We showed the details in Figure 2 (study flow) and described in Study Design and Follow-up sections.

Detailed Responses to Reviewer #3

Reviewer Name: Francesca Chappell

Institution and Country: University of Edinburgh, UK

General Comments:

Peer Review – BMJ Open – 2018-024290. R1

I am glad to see that the authors have made changes to the paper which I feel have improved it. However, I do still have some reservations. Please see below.

Response:

We sincerely thank the Reviewer for the constructive suggestions and help in improving our manuscript. We have made every possible effort to address the concerns and highlight the changes in Yellow in the revision. We feel that these revisions significantly strengthened our manuscript.

Detailed responses are below.

The authors still intend to use Spearman's correlation coefficient to assess the relationship between the SSD scale and the physician's assessment. This is flawed. They say, "The accuracy of the severity assessment of a 20 questionnaire is measured by the Spearman correlation between the questionnaire score and the physician’s severity assessment." No, it is not. Please look at this website (which was also published as a paper in the Lancet): https://www-users.york.ac.uk/~mb55/meas/ba.htm about comparing ways of measuring the same thing. Please pay particular attention to the section headed, "Inappropriate use of the correlation coefficient".

Response:

Thank you for the suggestion and reference. Our team discussed the choice of this measurement. In our study, we plan to measure the diagnostic performance of a quantitative marker (SSS-CH or PHQ-15 score) in classifying an ordered three-class health condition (mild, moderate, severe). In the literature, the statistical kappa is suggested in similar situations. However, applying kappa requires confirmed cut-offs to discretize the questionnaire scores into three severity levels. In the population of our study, the proper cut-offs of both SSS-Ch and PHQ15 for severity assessment remain uncertain to some extent. We worry that the cut-offs may lead to an improper comparison of diagnostic performance.

We find in reference (1) that the VUS (volume under the surface) of ordered three-class ROC serves as a measure of diagnostic performance that may be applied to our study and is independent of cut-

offs. VUS is a multi-class generalization of AUC (area under the curve) of ROC. Therefore, we decide to apply VUS according to reference (1) and modify our protocol in the following sections accordingly:

In the section of Outcome measures/ diagnostic performance:

“The accuracy of the severity assessment of a questionnaire is measured by the Spearman correlation between the questionnaire score and the physician’s severity assessment.”

We change into the following:

“The accuracy of the severity assessment of a questionnaire is measured by the VUS (volume under the surface), which is a multi-class generalization of AUC (area under the curve) of ROC between the questionnaire score and the physician’s severity assessment [(18)].”

In the section of sample size calculation:

“… and Spearman’s correlation of the PHQ-15 score with the physician's diagnosis was 0.77 (95% CI: 0.43, 0.92). ”

We change into

“… and the VUS of multi-class ROC for PHQ-15 score with respect to the severity assessment was 0.7.”

In the section of sample size calculation:

“the sample size for severity assessment was 567.”

We change into

“The sample size for severity assessment was 517.”

In the section of statistical analysis

“(2) the non-inferior comparison of the SSS-Ch with the PHQ-15 in SSD severity assessment measured by Spearman’s correlation with Δ=0.1, α=0.025 in the population with a confirmed SSD diagnosis using Fisher’s Z test.”

We change into the following:

“(2) the non-inferior comparison of the SSS-Ch with the PHQ-15 in SSD severity assessment measured by VUS with Δ=0.1, α=0.025 in the population with a confirmed SSD diagnosis using Z test [(18)].”

Reference:

1 Christos T. Nakas, Constantin T. Yiannoutsos. Ordered multiple-class ROC analysis with continuous measurements. Statist. Med. 2004; 23:3437–3449.

In the section of Outcome measures/ diagnostic performance:

"Fortunately, each subject in our study will undergo the same set of questionnaires for the entire scale, so the missing samples who are lost to follow-up will not differ among the groups and therefore will not produce significant bias and will not affect our assessment." I commented on this statement at the first peer review and I still disagree. How can they say, that missing data will not differ among

groups? The authors themselves say "The potential of monitoring the treatment effect will be affected by loss to follow-up bias due to the unpredictable pattern of loss to follow-up." I agree with this latter statement (point 6, Strengths and Limitations).

Response:

We agree with the reviewer. We removed this flawed statement mentioned above and changed the sentence with: “Because of the difficulty of compliance, only a small fraction (16% by estimation) of patients in study would be involved in the follow up, and the result of monitoring the treatment effect may be affected by loss to follow-up bias.”

It's good to see that they have changed the imputation method from hot deck to multiple imputation. However, I'd also like to see a statement about only conducting multiple imputation if the data are MAR or MCAR.

Response:

We agree that missing value imputation should only be used when MAR or MCAR. As this assumption cannot be statistically tested, we are not sure what is the practical way of assuring it by design. We modified the statement as follows: “Missing values will be imputed with multiple imputation under the assumption of MAR.”

Re the "monitoring treatment efficacy in patients is measured by correlation with the SF-20 during follow-up visits" and the mention of subgroup analyses according to gender and age. I don't agree with the use of correlation to measure treatment efficacy. There are statistical models that can handle repeated follow-up data where not all patients get follow and those who do have follow-up might not have it at the same time. Treatment efficacy could be easily adjusted for age and sex in these models. Also, I still not quite sure what the authors intend to do for the analysis. Please read this paper (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2971698/) which covers repeated measures and missing data.

Response:

We sincerely thank the reviewer for the reference (repeated measure and missing data).

Please allow us to explain more clearly.

We measure the correlation between SSS-CN and SF-20, not because we need a technique to overcome repeated measures or missing values, but because we consider Quality of Life (by SF-20) to be the clinical endpoint related to the treatment.

Most of the SSD questionnaires, such as PHQ-15, are designed for diagnosis. The scores reflect the severity of disease at baseline, but the changes of scores do not necessarily reflect the clinical benefit of the treatment. Actually, we have already observed in clinical practice that the health conditions of the patients may improve while the PHQ-15 scores remain unchanged.

In this study, we want to investigate whether SSS-CN performs better in treatment effect monitoring, i.e., a convenient surrogate endpoint with respect to the clinical endpoint (SF-20).

VERSION 3 - REVIEW

REVIEWER	Dr. Johannes Laferton Philipps-University of Marburg, Germany
REVIEW RETURNED	12-Mar-2019

GENERAL COMMENTS	The manuscript improved after the revision. However, it still lacks precision at several points, which might mislead the reader reading specific aims and methods. General: It would save reviewer time, if the authors could state page and lines of the changes they made due to the revision. Point 1 P7L19-23: '… account for 50% of the physical complaints … account for the remaining 50% … This could be misleading: account for could imply all possible symptoms a patient might have in that aspect is measured by the scale. I think the authors mean '50% of the items ask about …' Point 2: Regarding my comment from the previous review: 'P9L1: Severity categories: The authors list four severity categories. They argue that these are based on the empirical cut off (30). I do not fully understand how 4 categories are based on one cut off score?' Please do explicitly state that these cut-offs are your hypothesis and not yet based on previous research. Point 3: Desing Checklist: Some items of the checklist can not be found as stated by the authors (lines and pages). E.g. Item13 a und b: Not found were dedicate by the authors (P12 L1-3 P9L16-18, P9L18-22). Please do make sure that your all points are correct before re-submitting your manuscript. Point 4 P29L7: Old comment: do the authors have a reference for the tendency of Chinese people refusing psychotherapy? Author Response: We have put the references below and inserted them into the manuscript. The references are NOT cited, where the claim is made. Point 5:P29L17-18: 'Based on published studies, we aim to develop a self-administered questionnaire to provide a comprehensive reflection of both somatic and psychological features. ' I think the purpose of the SSS-Ch has gotten clearer due to the revisions. However, the statement above might still be misleading. As the authors now state in the introduction, the SSD frequently is accompanied by anxiety and or depression. Therefore, they propose to assess general anxiety and depression alongside somatic symptoms (at least that is what I think the authors want to achieve wit the SSS-CH). Most of the items of the SSS-Ch are general anxiety and depression symptoms and not specific psychological features defining the SSD. Only item #17 appears to be a SSD psychological symptom. Hence, it would be more explicit to state something like '… we aim to develop a comprehensive questionnaire assessing somatic symptoms of SSD and general anxiety and depression symptoms and illness anxiety (1 item)'

REVIEWER	Francesca Chappell University of Edinburgh, United Kingdom
REVIEW RETURNED	27-Mar-2019

GENERAL COMMENTS	This paper was a pleasure to read - clear, careful, and of clinical importance, I wish the authors luck in completing the study. I have no further comments to make regarding statistics or methodology.

VERSION 3 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Detailed Responses to Reviewer #1

Reviewer: 1

Reviewer Name: Dr. Johannes Laferton

Institution and Country: Philipps-University of Marburg, Germany

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

The manuscript improved after the revision. However, it still lacks precision at several points, which might mislead the reader reading specific aims and methods.

General comment:

It would save reviewer time, if the authors could state page and lines of the changes they made due to the revision.

Response:

Thank you for your suggestions. This has been done.

Point 1 P7L19-23: ‘… account for 50% of the physical complaints … account for the remaining 50% …

This could be misleading: account for could imply all possible symptoms a patient might have in that aspect is measured by the scale. I think the authors mean ‘50% of the items ask about …’

Response:

Thank you for your suggestions. As you suggested, changes have been made in Page 7, line 19-22. “Half of the items ask about physical complaints (1 item per body system, items 1, 5, 9, 10, 12, 13, 16, and 18-20). The remaining items ask about anxiety and depression (anxiety items 6, 14, 15, and 17; depression items 3, 4, 7, and 11; and anxiety and depression items 2 and 8).”

Point 2: Regarding my comment from the previous review: ‘P9L1: Severity categories: The authors list four severity categories. They argue that these are based on the empirical cut off (30). I do not fully understand how 4 categories are based on one cut off score?’

Please do explicitly state that these cut-offs are your hypothesis and not yet based on previous research.

Response:

Thank you for your suggestions. We set 3 cutoff values (30, 40, 60) for the 4 severity categories. We followed your suggestions and clarified our statements on Page 8, line 7-10. “The selection of the cutoff value of 30 is based on the results of our previous study (it was obtained from the receiver operating curve (ROC), reaching a sensitivity of 0.97 and a specificity of 0.96)[16]. Other cutoffs (40,60) are chosen based on clinical experience rather than previous research.”

1 Qi Z, Jialiang M, Chunbo L, et al. Developing of somatic self-rating scale and its reliability and validity. Chinese Journal of Behavioral Medicine and Brain Science 2010;19(9):847-9.

Point 3: Desing Checklist: Some items of the checklist can not be found as stated by the authors (lines and pages). E.g. Item13 a und b: Not found were dedicate by the authors (P12 L1-3 P9L16-18, P9L18-22). Please do make sure that your all points are correct before re-submitting your manuscript.

Response:

We have double-checked our checklist after revision.

Point 4 P29L7: Old comment: do the authors have a reference for the tendency of Chinese people refusing psychotherapy?

Author Response:

We have put the references below and inserted them into the manuscript.

The references are NOT cited, where the claim is made.

Response:

“ref 4 and ref 8” were the references we cited. We are sorry that we did not cite the references every time we described this phenomenon. These references have now been included in the Introduction section (Page 5, line 6) and Methods section (Page 11, line 2) to help international readers understand the societal and cultural barriers to psychotherapy in China.

Point 5: P29L17-18: ‘Based on published studies, we aim to develop a self-administered questionnaire to provide a comprehensive reflection of both somatic and psychological features.’

I think the purpose of the SSS-Ch has gotten clearer due to the revisions. However, the statement above might still be misleading. As the authors now state in the introduction, the SSD frequently is accompanied by anxiety and or depression. Therefore, they propose to assess general anxiety and depression alongside somatic symptoms (at least that is what I think the authors want to achieve with the SSS-CH). Most of the items of the SSS-Ch are general anxiety and depression symptoms and not specific psychological features defining the SSD. Only item #17 appears to be a SSD psychological symptom. Hence, it would be more explicit to state something like '… we aim to develop a comprehensive questionnaire assessing somatic symptoms of SSD and general anxiety and depression symptoms and illness anxiety (1 item)'

Response:

Thank you for helping us improve the entire manuscript. We have revised the sentence and added it in Page 5, line 16-17. "we aim to develop a comprehensive questionnaire to assess somatic symptoms of SSD comorbid with anxiety and depression symptoms."

Detailed Responses to Reviewer #3

Reviewer Name: Francesca Chappell

Institution and Country: University of Edinburgh, United Kingdom

Please state any competing interests or state 'None declared': None declared.

Please leave your comments for the authors below

This paper was a pleasure to read - clear, careful, and of clinical importance, I wish the authors luck in completing the study.

I have no further comments to make regarding statistics or methodology.

VERSION 4 - REVIEW

REVIEWER	Dr. Johannes Laferton Psychologische Hochschule Berlin, Germany
REVIEW RETURNED	11-Jul-2019

GENERAL COMMENTS	The authors have appropriatly responded to all my comments.