Author’s response to reviews

Title: The theoretical and empirical basis of a BioPsychoSocial (BPS) Risk Screener for detection of older peoples’ health related needs, planning of community programs, and targeted care interventions

Authors:
Zoe Hildon (Zhildon1@jhu.edu)
Chuen Seng Tan (chuen_seng_tan@nuhs.edu.sg)
Farah Shiraz (farah_shiraz@nuhs.edu.sg)
Wai Chong Ng (nwc@tsaofoundation.org)
Xiaodong Deng (xd.deng.r@gmail.com)
Gerald Choon Huat Koh (Gerald_Koh@nuhs.edu.sg)
Kelvin Bryan Tan (kelvin_bryan_tan@moh.gov.sg)
Ian Philp (Ian.Philp@wbs.ac.uk)
Dick Wiggins (d.wiggins@ucl.ac.uk)
Su Aw (elessar.starr@gmail.com)
Treena Wu (Treenawu@tsaofoundation.org)
Hubertus Vrijhoef (bertvrijhoef@yahoo.com)

Version: 1 Date: 25 Sep 2017

Author’s response to reviews:

The theoretical and empirical basis of a BioPsychoSocial (BPS) Risk Screener for detection of older people’s health related needs, planning of community programs, and targeted care interventions

Reviewer responses:

Review 1 (statistical): Na Ling

This is an interesting paper that validated a risk stratifying tool by measuring the association of BPS scores and various health outcomes, including falls in the past 12 months, cognitive
function, number of diseases, hospitalizations, ED visits, and length of stay. The paper has practical value and I believe would contribute to service planning for vulnerable older adults.

Reviewer 2: Angela Ghesquiere

This paper examines the properties of a new composite risk assessment tool for older adults’ biopsychosocial needs. The paper has the potential to make a useful contribution to the geriatric assessment arena. The paper has many strengths, including a well-described, thorough analysis on a large sample, and tables and figures that substantiate the content. I do have a few questions and suggestions to improve the paper, however:

Reviewer 3: Tom Kingstone

Thank you for submitting this manuscript for review. I have conducted a thorough review and provide my comments for each of the sections below.

Authors response:

We thank the reviewers for their positive comments and helpful critical feedback and endeavor to address these carefully. To avoid confusing track changes, we track using this mode while excluding deletions and format changes and index the changes using comments linking to the reviewers points #1-29 below.

We have added another author to this paper; her name is Su Aw (NUS) and she helped critically appraise the literature and also helped to conceive the loads-lifts-levers model (she is co-leading the full theory paper we are preparing on this). It was an oversight to not include her from the start, as she has contributed intellectually and practically to this work. She is added to the title page and in the submission system, we will send in the signed form for adding an author from all co-authors.

Point by point response to reviewers

Abstract:

Reviewer 2

1. The abstract is difficult to follow

• We have edited the abstract to be clearer about the purpose of the work and taken out terms such as ‘frailty’ that need more space than allowed in the abstract to define. We have left in the term health ‘loads’ upfront however, defining it briefly in the abstract, as this is a core concept in this paper, i.e. what the risk screener sets out measure, p.1.
introduction/background:

Reviewer 2

2. The Introduction are difficult to follow, referring to many concepts (e.g. Third Agers and Fourth Agers, the Proactivity Model) and abbreviations (BPS resources, H-R QoL) without much, if any, explanation. There is also an overuse of abbreviations (e.g. CD courses, CM) which makes some simple concepts overly complex. The Introduction is also organized in such a way so that exactly what the authors are focusing on, and why this is important, is not entirely clear until the end. I would suggest explaining all concepts as clearly as possible upon their first mention, and reorganizing the entire introduction so that the flow of ideas is clearer.

Reviewer 3 (continuation of point 2)

I appreciate that what the authors are explaining is conceptually complex; however, the background section is too long and could be structured better. As a result, the rationale and importance of the work is lost. Also, I believe the background section would benefit from adopting a more critical gerontology perspective. The literature requires updating. For instance, the authors claim that 'successful ageing' has only recently been re-conceptualized. However, this concept has been criticized for many years and even superseded by notions of active ageing and healthy ageing - the authors cite papers that are over 10 years old. The background section is in need of amending. Overall: The authors only lightly discuss their findings in the context of existing research.

- We agree the background was dense; we have restructured and cut the review of existing measurement approaches and address this in the discussion instead, leaving more space to better unfold our theoretical and conceptual review, and our position within it. We sign-post the background section upfront, and structure it according to the core concepts that position our measure, p. 4.

- We also agree that the section on successful ageing needed more references and critical review, added in pp. 5-6. In this revised draft, we describe the literature and conceptual terms to better substantiates our position. Conceptual terms are clarified and referenced in the following sections: Defining BPS health and related risks at older ages, p. 6; vulnerability, p. 7; Loads-Levers-Lifts processes pp. 8-9. The measurement approach is then related to these sections, under: Measuring ‘loads’: a theoretically derived scoring system, p. 9.

3. The authors described adaptive capability but do not describe resilience, which is an important concept in this work.

- We thank the reviewer for drawing attention to this; the resilience tradition of research plays a big part in the way we have conceived our theoretical position. We now link the Loads-Levers-Lifts explicitly to the tradition of resilience research, and define the term, and have added classic and current references to substantiate this, p. 8.
Reviewer 2

4. While the Loads-Lever-Lifts Processes and the goal of Adaptive Capability are both explained well and connected to the screener, both the linearity of this approach and the derivation of the "some" vs. "a lot" benchmarks seems a little oversimplistic, and their validity could be better explained.

• We agree that the operationalization on the model needed more clarity with regards to the scoring system. ‘Some’ and ‘a lot’ were indeed derived simply, as above or below the median, which seems the most objective way to start to compile our score. We did not want to weight of give priority to any domain. The idea from our theoretical stance is to give equal priority to all domains and this will allow us both to see how much each domain contributes, and if dose response occurs when equally combined (by and large it did). However, much iterative testing was carried out before choosing the scoring system itself. For example, we ran sensitivity analysis on combined scoring as 0 no risk, 1-2 as low, 3-4 mid and 5-6 high compared to the final chosen combined scoring system of 0-1 as low, 2-3 mid, 4-5 high, and 6 as overwhelming. Finding a much better fit with the latter.

• The pilot administration in the community and use in programming gave insights into the lived experience of the scores. For example, indicating to care givers that those with 0 or 1 shared similar narratives, and those with a 6 score were often more in need of palliative support and were distinguishable from 4-5’s who still tended to be more open to community activities. Qualitative analysis of care-givers and older people’s views on BPS Risk is analyzed elsewhere, and will be further elaborated in the second version of this tool which used mixed methods to validate it. This forthcoming work keeps the scoring structure as we define it herein, since it remained the most robust scoring approach. Although these details are too lengthy to share in the text of this necessarily condensed introduction, the steps put into elaborating the scoring system have been clarified, p. 9.

5. While the authors do an excellent job explaining how their screener differs from QoL and bio-functional scales, and the potential implications of the screener's use, they do not clearly explain how the screener differs from other measures that also capture biopsychosocial factors, such as the SF-36?

• Like the EQ5D the SF-36 measures the bio-medical domains most emphatically, the social component is only developed in relation to ways in which bio-medical problems interfere with social participation (H-R QoL); also, these tools were not developed specifically in relation to older ages. The SF-36 has been added to the discussion, p. 20.

Reviewer 3

6. There is an over-reliance on the use of figures in the background section; could this not be described succinctly instead?
• We have moved the section reviewing other instruments to the discussion, and review these in greater depth; the pyramid figure comparing the BPS emphasis of existing measures has been cut.

• We are conscious that with all the theory in this paper the remaining background figures can help to clarify our work, especially for the less conceptually minded. Should they crowd the copy edit we suggest they can be moved to supplementary files at the editors’ discretion, as the text descriptions should stand alone. We are reluctant to cut them completely.

7. For an international audience, 'older adults' or 'older people' are more accepted terms than elder or elders.

• We have updated this terminology throughout.

8. Ln 74 Please explain what you mean by 'stages of ageing' (life stages would be more appropriate since age stages is a more contested term)

• We have removed this term and reworded, p. 3.

9. Ln 75 Is it not a case of finding the right program for the right older adults? Sorry to be pedantic.

• We fully agree, we have swapped this around, this paragraph was also moved to the discussion, p. 21.

10. Ln 96 Paragraph starting on Ln96 requires editing - the excessive use of e.g. interrupts the flow.

• We also agree with this, thanks for helping us with the flow, we have removed all the uses of e.g., p. 6.

11. Ln116 Can you provide evidence to support this statement re: frailty?

• The point about frailty being largely a bio-medical construct had been supported by references 32 to 34, p. 7.

• We argue that the construct of vulnerability, as we define it, which allows for older people to be socially and psychologically at risk, as much as bio-physically, is a less ageist and more accurate view of risk as people age. Particularly in the younger old. Catching vulnerable younger old people before they also become sick will help us to support people to age resiliently. The discourse on ageing needs to move away from models that prioritize the bio-functionality of old people, and also measure the domains accurately as possible. The Tilburg Frailty Indicator or TIF, for example has classed cognitive decline as part of their P domain, while our empirical tests did not support this classification. It showed instead that emotional health was different, although no doubt overlapping, with deteriorating cognitive health, caused by restricted blood
flow to the brain - strokes and dementia. The ‘bio’ domain was indeed very much reflective of age-related disease pathology, and the other domains clearly distinct.

Methods:

Reviewer 1

12. What language version(s) of the survey was used?

- The questionnaire was available in English, Mandarin, and Malay. Surveyors were familiar with Chinese dialects such as Hokkien and Cantonese and adapted the questions orally where necessary, using scripts. This has been added, p. 11. There are many Chinese dialects in Singapore, and older people often prefer speaking their own dialect although most will understand English and/or Mandarin. We were careful to select a survey company with experience in interviewing older people, and multi-lingual surveyors. In addition, Singapore elders speak Punjabi and Tamil, however most of these Singaporeans of Indian decent do speak good Malay or English.

13. Please indicate the scale of responses to the 35 questions used in the exploratory factor analysis (table 1a-c). Are they binary responses (yes vs no) or Likert-scale questions (3, 4, or 5 points)? In lines 253-254, the authors indicated that the questions were ordinal variables, please state clearly the scales of these variables, especially for the social domain questions.

- As the 52 items included in the factor analysis were binary, or Likert, or count responses (i.e., social domain questions which counted number of social networks), this is clarified in the text, p. 12. Therefore, we treated them as ordinal variables by assuming each item had an underlying unobserved normal variable and used polychoric measure [see refs 48, 50].

14. In lines 257-258, the authors kept the items with factor loadings of 0.5 or above, and all cross-loadings of 0.3 or above. I wonder if it’s a typo: should it be cross-loadings of 0.3 and below.

- Yes, thank you for spotting this typo, we have updated, p. 13.

15. Lines 261-269: Please define the thresholds for binary covariates: education and income. How were these variables dichotomized as good education vs. no formal education, poor vs. not poor? Please also specify the health outcomes you would like to study in this section, and specify which ones were treated as binary (how they were dichotomized) and which were count variables.

- For education, we dichotomize education into no formal education which is lowest education category and good education refers to primary education and above. For income, we dichotomize income to poor which correspond to average income is less than or equals to S$250 a month and not poor which corresponds to average income greater than this.
For binary outcome variables (Table 2a to 2c), we dichotomised as follows:

a. For self-reported health, poor or fair corresponds to “poorer health”, and good or, very good, or excellent corresponds to “better health”;

b. For self-reported falls in the last 12 months, more than one corresponds to “yes”, and none corresponds to “no”.

c. For cognitive impairment, inter-rai cognitive performance scale observed “intact” is ‘not impaired and “borderline and above” corresponds to impaired.

For the count outcome variables, (Table 2d to 2h), we used the following outcomes:

d. Total number of self-reported hospitalization in the last 6 months.

e. Total number of self-reported diseases (there were 10 in total, see detail below, response to point # 18 below).

f. Total number of hospitalization in the last 6 months from administrative data.

g. Total number of emergency department visit in the last 6 months from administrative data.

h. Total length-of-stay in the last 6 months from administrative data.

These details have been added to the methods, pp. 13-14.

16. Lines265-264, count data were explored with Poisson modeling. Did the author test the mean and variance assumption for Poisson modeling? If the assumption is violated, a negative binomial model would be more reasonable.

We thank the reviewer for highlighting that we need to include this useful detail. For the self-reported total number of hospitalization and diseases the goodness-of-fit tests suggested adequate fit and hence Poisson modeling is reasonable. For the total number of hospitalization, emergency department visit and length-of-stay from administrative data, the goodness-of-fit test suggested inadequate fit due to over-dispersion and hence Negative-binomial was used. We have added this detail in the manuscript, p. 14.

17. Lines 291-298: All test indices were reported except the Root Mean Square Error (of Approximation)

This paper is already very dense, and we would argue for keeping the focus on the objective of exploring individual BPS constructs, creating and testing the managing counts and testing the scoring system to capture risk categories.
• We would therefore emphasize that in view of our objectives our evaluation is largely exploratory. Our reported alpha-coefficients suggest that the internal consistency of the 3 domains are excellent. In addition, the percentage of variance explained (66%), following a promax rotation is very encouraging [added on p.13 & 15]. This we propose was enough to justify the further exploring of the individual domains and the results of the managing score, which as hypothesized, produced strong dose response effects.

• In sum, since the paper is largely about theory mapping to this innovation in scale development and scoring we would prefer not to report the RMSEA, or allow it to stand as a key arbiter of the value of this work. Mention of this criterion has been removed (p.13).

• In the discussion, p. 23, we have argued that given the utility of the existing scale for the purpose of risk categorizing as we have tested, in the future it would be appropriate to conduct a confirmatory FA across different settings and with comparative datasets; we conclude by saying a v2 which will build on the current work but be expanded/cohesive will be forthcoming, p. 24.

18. For the outcome of number of diseases, please indicate what diseases were included.

• Total number of self-reported diseases was added to p., these are: (1) heart disease, (2) stroke, (3) chest/lung disease, (4) cancer, (5) arthritis, (6) osteoporosis and bone fractures, (7) diabetes, (8) high blood pressure, (9) high cholesterol, and (10) doctor diagnosed depression, this is added p. 12.

Reviewer 3

19. The methods seem well-described and concise. The findings seem to be reported in full - however, please note that I am not a statistician so cannot comment on the accuracy of the statistics.

• We are glad that overall the methods read well; have added in useful details suggested by the statistical reviewers, but aimed to still keep things concise.

Results:

Reviewer 1

20. Results tables attached at the end of the manuscript have the same title. It is not quite clear to me which tables match which results. Please indicate specific outcomes in the table title.

• The tables have had the outcomes added in the top line for clarity, and their formatting has been tidied.

21. Results model section: I assume that the B, P, S domains were put into the same model, adjusted for other covariates, rather than that each was used in a separate model, adjusted for covariates. Please clarify this in the methods and results sections.
The B, P, S domains were put into the same model adjusted for other covariates. We have clarified this in the methods, p. 14, and added footnotes to all the tables.

Reviewer 2

22. The Psychological domain seems very limited compared to the Biological domain, and does not capture key psychological symptoms, such as anxiety or irritability. While I know that these items came from the EASYCare standard, the authors could have added items, and could better explain the rationale for the relatively brevity of this section.

• We agree with the reviewer that the Psychological domain is limited; a mixed methods study has been undertaken to generate new items for this and the social domain. The current study was not funded to collect additional data and was tasked with testing the theory on existing, internationally validated existing items. This initial set of analyses open the door for further funding to test the theory and elaborate the tool with a more complete, and culturally derived set of items. The limitations of the P and S domains are discussed p.22-23, and the P domain in notated to be particularly brief.

Discussion:

Reviewer 1

23. Discussion section: Can the authors comment on the phenomena that those with overwhelming problems did not show poorest performance for several outcomes.

• While it is true that the score was associated with mounting risk for all outcomes, between higher and lower risk categories in broad terms (at least across two categories out of the three) - it is also true that administrative service use data revealed an in interesting, repeated inverse association that what was expected between ‘many’ and ‘overwhelming’ problems. This could be because long to mid-term care (LMTC) services and palliative home care services were not tracked, see also the next point below, this explanation was clarified p. 19 xx.

24. Limitations: cross-sectional study design, health service utilization was self-reported instead of using claims data, thus it was subject to recall bias.

• We used both self-reported data and administrative (claims data), since our data was linked to hospital administrative data (see p. 14). However, administrative data did not include hospitalization for mental health, Institute for Mental Health (IMH) data, nor long to mid-term care (LMTC) services which we would argue, despite the chance of recall issues, makes the self-reported data more complete, see p. 21. Moreover, the data that was self-reported was also asked about over a short period of six months. These data better reflected expected associations for the managing score, while all administrative data seemed to suggest those with ‘6’pts were not using services we were tracking in favor of LMTC or palliative home care, see point #23.

• The cross-sectional nature of the study is highlighted, p. 23.
25. Please comment on the different functions of the tools (BPS domain scores vs. managing scores), and the medical context best suited for each tool.

- Thank you for asking us to elaborate this distinction, it is indeed an important point, worth being more explicit about. The overall, combined managing scores are useful to guide referrals to programmatic streams, although should be used for guidance only, and to inform a shared meaning of risk rather than as absolute indicators. These combined scores should also be useful for intervention evaluations, as they promise to be more change sensitive than traditional frailty measures, which tend to stagnate around the emphasis on the bio-functional, which may not improve even though other health domains have.
- The domain specific B, P and S managing counts will guide care planning and help the care manager decide on areas where further in-depth assessment or referrals are needed, these points are clarified at the outset of the discussion, on p. 21.

Reviewer 2

26. In the Discussion, the authors make a number of statements about the relative contributions of biological, psychological, and social factors to the health outcomes analyzed. It should be more clearly noted that these conclusions are only applicable in this particular sample, not to all older adults. And an added limitation is that the study was conducted with a convenience sample in a single neighborhood in Singapore, and cannot be generalized.

- While our data indeed was not collected across a sub-sample of neighborhoods representative of Singapore, it was a close match socio-demographically, compared to census data. We note that further studies are needed to validate this approach in other settings, p. 22 & p. 23.

Reviewer 3

27. The discussion could be strengthened by the inclusion of more literature.

- We have expanded the discussion by moving citations previously in the discussion, and added to these pp. 20-21.

28. Ln 429 the term 'health architecture' is unusual.

- We have removed this term.

29. Were older people involved in the design of this project (e.g. patient and public involvement)?

- Unfortunately, older people were not formally involved in this first phase of the project; this was due to limited resources and being tasked with adapting pre-existing tools, see point #S22. However, views and experiences of older people were informally accounted for during the pilot administration of the scoring, which led us to test several versions of the scoring system (#
4). Elder views on the BPS screener and BPS experiences have since been explored with formal qualitative analysis that will be reported elsewhere.