Peer Review File

Article information: https://dx.doi.org/10.21037/mHealth-22-56

Reviewer comments

1. Methods - Line 98 - The authors state that "we believe that continuous improvement of 6 domains of health and wellbeing are crucial to health promotion and prevention" - a cursory search by this reviewer reveals a widely varying number of (e.g. 3, 4, 5, 6 or 10) domains of health and well-being cited in literature. Could the authors provide a citation for selection of 6 domains used as the basis for this research.

R: We agree. There is no consensus regarding the topic. We have highlighted that this is our "thesis" in this pilot study.We also added references.

"In our private healthcare system (Alice Healthcare, São Paulo, SP, Brazil), we believe that continuous improvement of 6 domains of health and wellbeing are crucial to health promotion and prevention. We realize that there is no consensus in defining which are the most relevant domains that best translate health and well-being. In our thesis, we chose to test in this pilot study the following (9-21): Physical activity, sleep quality,nutrition, habits/lifestyle, mental health, quality of life"

2. Introduction - Line 88 - The authors state that "We aimed to develop a Healthcare Magenta Scorecard, mainly from the agglutination e-PROs " and (in the Introduction - Line 90) that "This study aims to describe the rationale of our healthcare scoreboard (magenta score) and some preliminary data from our initial efforts."

(a) Rather than "agglutination", did the authors mean "aggregation" of e_PROs? In aggregating multiple questionnaire instruments from disparate health and well-being domains (which have been validated in previous studies in their own right) into a new composite instrument, this reviewer believes that validation of the new resultant composite instrument is required to demonstrate that previous published validations for constituent instruments used still hold true when multiple measures are captured and combined in what is essentially a new composite instrument (e.g responding to a much longer series of questions may have an effect of validity). This reviewer believes that a validation studies are required before any inferences regarding efficacy can be asserted when standalone questionnaire measures are amalgamated into a new composite measure.

R: We agree. We changed agglutination to aggregation. We also agree that we still need to validate this new instrument. We added in the discussion section:

"Comprehensive and long e-PROs, such as magenta score may have even lower response rates(34). As such, we still need to conduct a prospective validation study to verify wheter our scorecard has acceptable psychometric properties."

(b) Did the authors consider an appropriately blinded randomised controlled study comparing groups using and not using the new scorecard to assert any benefit from compliance with use of the scorecard in terms of health improvement compared with a control group?

R: Thanks for the comment. We believe that a RCT would certainly be appropriate and is a topic of future investigation.

(c) Further to (2a), the value of the new composite measure derived in this study requires validation (i.e Methods - Line 288 - "Magenta scorecard is the summary of the score (mean score) of all the 6 domains". Are means scores the most appropriate measure of central

tenancy to use or are medians or modes more appropriate? In summarising scores from multiple measures across the health and wellbeing domains, of what value is a collective score number across domains? If the authors meaning is presentation of 6 summary numbers (as opposed to a single number), this should be clearly stated as such.

R: We agree that further validation is need. Presenting medians could be also of value, however, as we had a small amount of outliers, presenting means is easier to the journal´s audience to understant. In addition, as the scorecard is aimed to be a measure of long-term assessment of health and wellbeing, future research will consider mean differences along periods of time. Our aim (in the future) is to correlate mean differences scores, adherence to health/wellbeing programs and cost-effectiveness.

(d) In light of (2a) and (2b), could the authors consider describing this as a pilot study?

R: We agree. Now in the text:

"This is a single-center preliminary pilot study regarding the development of a health scorecard in our private healthcare system (Alice Healthcare, São Paulo, SP, Brazil)"

3. Methods - Line 26 - The authors report development of a mobile based e-Pro that measures patients health and wellbeing every 3-5 months and (Line 28) that "We PROSPECTIVELY collected data when patients onboard in our healthcare system'. The Ethics disclosure (Line 344) states "The study explored and exposed our Magenta Scorecard framework and used anonymous data COLLECTED PREVIOUSLY for operational/healthcare purposes. There was no analysis from health individual personal data. As anonymous preliminary data, it fits as waived from ethical approval (Resolution 674, Capítulo IX, Art. 26)".

R: Our data was collected previously for for operational/healthcare purposes.

Now in the text: "We collected data when patients onboard in our healthcare system'.

(a) If prospective data capture was performed by study participants using an app on their smartphones, this reviewer would expect that the protocol for, and conduct of this human research study would require approval from a university or hospital ethics committee, including informed consent obtained from all participants. If this was obtained, it should be stated in the manuscript.

R: Please see above.

(b) The description of the study cohort in the Methods section lacked detail regarding the sampling methodology (e.g. opportunistic, snowball etc). How were participants recruited? Were there any benefits or incentives offered to participants to participate? Further, it was unclear as to what inclusion and/or exclusion criteria were applied when screening participants for this study.

R: Now in the text: We included a convenience sample and included all the participants that responded the scorecard from december 2021 to november 2022. There was no incentives regarding to responding the scorecard.

(c) The protocol employed in this study is unclear. In Introduction - Line 90, the authors mention that this new Scorecard is "a method for patients to partake in their own selfcare." This reviewer expected a more detailed description as to how data was captured in this study in the Methods section. Was data captured prospectively into a mobile app by participants themselves, was it obtained by interviews with investigators or was data extracted

retrospectively from other systems to yield the analysis presented?

R: Data was collected retrospectively for our study purposes.

Now in the text:

We applied the score card every 3-5 months and patients questionnaire responded using a mobile-based app. As most of the information about the health plan (accredited network, available medical network, pending health actions) is within the mobile application, the scorecard was a pending action when joining the health plan and a new assessment was encouraged every 3 -5 months. We included a convenience sample, which is those that responded to the scorecard from december 2021 to november 2022. We included only patients that responded at the scorecard completely two times in a more than 3-month timespan. Incomplete scorecard responders were excluded. There was no extra incentives regarding to responding the scorecard besides measuring patient health and wellbeing status. Data was captured prospectively into a mobile app by participants themselves and data extracted retrospectively from our database for this manuscript needs.

(d) Readers would benefit from a brief narrative regarding the app itself and how participants interacted with it to capture data i.e sliding scale or dot point likert scales displayed on screen for each questionnaire used in this scorecard. Was a standard alone native IOS and/or Android app, a HTML web form or a mobile interface to a centralised medical records (or similar) database used to capture data. How were results entered by particpants communicated/sent to the investigators i.e. online update applied to central database, accrual on their own device or by email etc.

R: Now in the text:

Data was captured inside our healthcare standard alone native IOS and/or Android mobile-app. For data collected we used a sliding scale or descriptive alternatives according to scorecard phases. Data was sent to the investigator by online update applied to central database.

We also change Methods subtile #2 to: **Data capturing and guided approach**

(e) In Methods - Guided approach - Line 113, the authors state "We applied the score card every 3-5 months and patients questionnaire responded using the a mobile-based app" suggesting a prospective approach. How were participants prompted to provide responses (e.g prompted by the app or reminded by email etc).

R: Now in the text:

"As most of the information about the health plan (accredited network, available medical network, patient pending health actions) is within the mobile application, the scorecard was a pending action when joining the health plan and a new assessment was encouraged every 3 -5 months"

.

4. Re Results - Sample Characteristics - Line 251 - In considering participant attrition during this study, the authors report that this study included 5,757 particpants. On Line 258, 1,662

responded T0 and T1 …”
To assist readers, could the authors consider refining the description of the decomposition of participant numbers along the lines of standard CONSORT reporting and it’s constituent phases i.e., Enrollment, Allocation, Follow-up and Analysis.
e.g Of 5,757 potential participants considered for enrollment in this study, xxx participants were excluded after applying inclusion and exclusion criteria. 1,662 participants completed follow-up at both T0 and T1 milestones and were included in the final data analysis.

R: Now in the text:

Data was collected from December 2021 to November 2022 and included 5,757 participants. Of 5,757 potential participants considered for enrollment in this study, 4095 participants were excluded after applying inclusion and exclusion criteria. 1,662 participants completed follow-up at both T0 and T1 milestones and were included in the final data analysis.