

## Peer Review File

**Article information:** <http://dx.doi.org/10.21037/mhealth-20-143>

### Reviewer Comments

Comment 1: This manuscript reports on a series of three formative research studies to adapt and test the measures and instructions for an ecological momentary assessment (EMA) protocol. The assessments concern social comparisons (with other people) and positive/negative social interactions. The formative research is well-justified. The daily diary and EMA research literature currently resembles a "wild west" of measures and instructions. Although I doubt a gold standard will ever emerge that receives researcher consensus, these three studies offer other researchers a blueprint to help refine their measures, tailor them to the target population, and medical/social context. The three studies follow a logical progression- from initial pilot, to qualitative interviewing to measure refinement and re-testing with a new, but similar sample.

- Reply 1: We appreciate the reviewer's positive feedback on the organization and potential utility of this submission. We hope that it does provide a blueprint for future studies in this area.

Comment 2: I don't know the extent to which the sample of high risk CVD women offered qualitative responses that suggest their risk status figured into their (mis)interpretations of items/questions. I suspect others of the same age would probably report the same confusions and interpretations about responding. If the daily survey or EMA diary had items specific to their risk status, we would obtain a better idea of how confusion, ambiguity or misinterpretation obscures responses.

- Reply 2: We acknowledge that our highlighted examples from interviews do not specifically implicate CVD risk as a contributor to women's perceptions, and that it is not possible to know from these studies how different their feedback would be from those of women with lower risk, men, or other populations. We believe that this work is necessary and we now specify this as a future direction. In interviews, participants' risk status was relevant to presenting the broader context of the study and why we wanted their feedback about these and other items (including assessments of body satisfaction, physical pain, and physical activity motivation).

Comment 3: I suspect some methodologists would require a more systematic approach to the coding of the qualitative interviews (e.g., INVIVO). I am not a qualitative researcher so from my perspective what the present researchers did is adequate. Others more expert in qualitative methods might urge more systematic coding, coding agreement, etc. However, this paper does

not recommend against that option. The significant contribution of this research is it provides a useful set of steps to increase the reliability and validity of responding.

- Reply 3: We agree that a formal approach to coding qualitative feedback is a step beyond what most researchers would consider necessary or practical for preliminary steps toward intensive assessment studies. Our goal was not to identify underlying themes in questions or around social comparison more broadly but to explicitly solicit feedback on the items, thus the more direct approach to coding the qualitative data in the current study. We now acknowledge this as a potential limitation of our work in the General Discussion: “Further, as the focus of these studies was the identification of issues with items rather than identifying underlying themes in the qualitative interviews, we did not employ formal qualitative analysis methods. Future work may require adjustment to the specific formative research questions at hand (e.g., with whom are participants comparing or having social interactions) and may benefit from extended interviews that would facilitate the use of formal methods of analyzing qualitative feedback.”

Comment 4: Line 51: Found this to be confusing: item performance in what was compared to Study 1?

- Reply 4: We have amended this sentence to indicate that item performance in Study 3 was compared to that of Study 1.

Comment 5: Line 82: add intended to be sampled or sample targeted?

- Reply 5: We have changed this sentence to read: “Yet, designing these studies requires unique attention to item construction, response framing, instructions to participants, and the experiences of the target population (Conner and Lehman, 2012).”

Comment 6: Line 102: "self-evaluations relative to others" is an ambiguous. Less ambiguous to state: "self-evaluations compared to others"

- Reply 6: We have modified this and a previous instance of the same phrasing, as suggested.

Comment 7: Line 121: How about rephrase: "understand the kinds of experiences queried " ...

- We have edited this sentence as suggested.

Comment 8: line 110: did the authors mean to insert "after" following "even"?

- Reply 8: We have reworded this sentence to read “...which may present difficulties with recognizing that they have occurred and then reporting on them, even after short periods of time (e.g., hours).”

Comment 9: line 138-141: the "within" and "between" is perhaps not the best way to communicate these ideas, at least in this early part of the paper. What I think you mean is that a person's social perceptions and health behaviors may vary across the day, week or month and also differ from other people's social perceptions and health behaviors.

Comment 10: line 146: But you don't describe how within- versus between-subject information would be differentially informative and useful for descriptive inferences about daily life or for informing the choice of an intervention. I think a couple of examples should be presented to make this crystal clear.

- Replies 9 and 10: We agree that these points could have been clearer, and we have revised this section to read: “As both social perceptions and health behaviors vary within and across days, weeks, and months, as well as between people, (Elavsky et al., 2016; Reichenberger et al., 2018; Vella et al., 2008; Wheeler and Miyake, 1992), intensive assessments could be crucial for understanding at what level(s) these processes are associated among midlife women. To date, however, social perceptions and health behaviors have been assessed more often as stable individual differences than as experiences that vary over short and longer periods. Findings from studies that use intensive assessment methods could inform and optimize interventions at both levels (Curran et al., 2014; Dunton, 2017). For example, identifying women *for whom* certain social perceptions are associated with low engagement in physical activity could help to target appropriate interventions toward this subgroup of women. In contrast, identifying *when* or *in what contexts* certain social perceptions are associated with decreases in women’s physical activity could help to identify the appropriate timing or circumstances for exposure to intervention content or reminders about using specific behavioral skills (Arigo et al., 2020a).”

Comment 11: ln. 156: "introductions to social perceptions" What does that mean? (It is made clearer in subsequent sentences, but I suggest making it clear from the start.)

- Reply 11: We have reworded this sentence to read: “In protocols that employ intensive assessment designs such as EMA, definitions of and introductions to concepts such as social comparisons and interactions may affect reporting – and thus, conclusions about these experiences and their relations with health behaviors in at-risk groups – though few studies have systematically evaluated approaches to honing item wording and participant instructions.”

Comment 12: ln. 299: Wills (1981) observed that people often report that comparing with others is an unbecoming thing to do. So, they might do it less or are be less likely to report making comparisons.

- Reply 12: We appreciate the reviewer calling attention to this aspect of Wills’s work, and we have added reference to this specific finding in our Introduction (Line 81): “Some existing evidence also indicates that participants may be reluctant to report that they make social comparisons, due to seeing them as socially undesirable (Helgeson and Taylor, 1993; Wills, 1981).”

Comment 13: The comparison items, as they stand, do not provide questions or responses that are compatible with opinion comparisons. That is a not a serious omission; the focus here is on social interaction and comparisons. I would also suppose a full research protocol would also include behavioral reports of some kind (e.g., health behaviors; medical adherence). Attitudes and opinions probably would be less salient and important ----Unless researchers consider such constructs as illness risk. In any case, the manuscript should be clearer about the kind of comparison domains the formative studies address. If the authors think the same approach can be adapted to comparison of opinions, risk, etc. then they so state.

- Reply 13: We agree that our emphasis was on comparisons of the self (broadly) or of behavior, given the larger study’s focus on cardioprotective behaviors such as physical activity. Although we did not intend for our items to exclude comparisons of opinions, we acknowledge that this type of comparison was not emphasized in the feedback process and participants did not often offer these as examples. We have added this point to our General Discussion: “The methods of evaluation and interpretations of results thus focused on these aspects of intensive assessment data, and equally important aspects such as the dimension of comparison (e.g., appearance, wealth) and the source of social interactions (e.g., family, coworkers) are not described. With respect to comparison domain, the original theoretical model indicated that comparisons are made primarily on the bases of abilities and opinions (or attitudes; Festinger, 1954), though subsequent evidence has shown that the range of specific comparison dimensions is much wider (e.g., personality, appearance, wealth; Wheeler & Miyake, 1992). In the current work, the broader context of understanding relations between women’s perceptions and their cardioprotective behaviors led to an emphasis on comparisons of behavioral performance or global assessments (e.g., health status), which may be akin to the overarching domain of abilities, rather than those of attitudes or opinions. Although this series of studies was not designed to exclude comparisons of attitudes or opinions, they received less emphasis throughout the formative research process than comparisons of health or behavior.”