



A multi-study approach to refining ecological momentary assessment measures for use among midlife women with elevated risk for cardiovascular disease

Danielle Arigo^{1,2^}, Jacqueline A. Mogle^{3^}, Megan M. Brown^{1^}, Adarsh Gupta^{2^}

¹Department of Psychology, Rowan University, Glassboro, NJ, USA; ²Department of Family Medicine, Rowan School of Osteopathic Medicine, Stratford, NJ, USA; ³Prevention Research Center, Pennsylvania State University, State College, PA, USA

Contributions: (I) Conception and design: All authors; (II) Administrative support: MM Brown; (III) Provision of study materials or patients: D Arigo, A Gupta; (IV) Collection and assembly of data: D Arigo, MM Brown; (V) Data analysis and interpretation: D Arigo, JA Mogle; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Danielle Arigo, PhD. Assistant Professor of Psychology, Rowan University, 201 Mullica Hill Road, Robinson Hall 116G, Glassboro, NJ 08028, USA. Email: arigo@rowan.edu.

Background: Ample evidence indicates that everyday perceptions of the social environment can affect health behaviors; these include social comparisons (i.e., self-evaluations compared to others) and positive versus negative social interactions. However, relations between social perceptions and healthy behaviors have received little attention among specific medical populations for whom an improved understanding of behavioral determinants could inform updates to tailored interventions. Research methods that capture and differentiate between stable, person-level differences and dynamic, within-person variability in these relations would be particularly useful, both for identifying their nature in daily life and informing improvements to tailored interventions.

Methods: We conducted a series of three formative research studies to adapt and test the measures and instructions for an ecological momentary assessment (EMA) protocol with midlife women who had elevated risk for cardiovascular disease (CVD; e.g., current diagnosis of hypertension, type 2 diabetes). Study 1 involved a pilot test of initial EMA items, sent to participants' smartphones 5 times per day for 7 days ($N=13$; $M_{\text{Age}}=47$, $M_{\text{BMI}}=33.7 \text{ kg/m}^2$), as well as brief exit interviews to identify points of confusion and suggestions for improvement. Study 2 used 1-hour, individual qualitative interviews with a new sample to elicit women's perceptions of revised items and identify additional opportunities for refinement ($N=10$, $M_{\text{Age}}=52$, $M_{\text{BMI}}=29.8 \text{ kg/m}^2$). In Study 3, a new sample of participants completed 7 days of EMA with revised items and instructions (5 times per day; $N=13$, $M_{\text{Age}}=50$, $M_{\text{BMI}}=33.4 \text{ kg/m}^2$).

Results: Item performance in Study 3, including the frequencies of reporting social comparisons and interactions, was compared to that in Study 1 using multilevel modeling; these tests showed meaningful improvement in reporting patterns between Studies 1 and 3 (e.g., changes of $d=0.33-0.75$ where appropriate).

Conclusions: Together, findings from this series of studies demonstrate the utility of a multi-study approach to refining EMA methods for use with midlife women who have elevated CVD risk, which may generalize to other populations of interest.

Keywords: Intensive assessment; ecological momentary assessment (EMA); social comparison; social process; midlife women

[^] ORCID: Danielle Arigo, 0000-0002-7807-5913; Jacqueline A. Mogle, 0000-0002-6082-228X; Megan M. Brown, 0000-0001-6451-106X; Adarsh Gupta, 0000-0002-6131-7218.

Received: 13 September 2020; Accepted: 13 December 2020; Published: 20 October 2021.

doi: 10.21037/mhealth-20-143

View this article at: <http://dx.doi.org/10.21037/mhealth-20-143>

Introduction

The use of intensive assessment methods (i.e., repeated assessment within the same person over short periods) (1) to capture dynamic psychosocial experiences and their relations with health outcomes has risen exponentially over the past decade, in part due to increasingly sophisticated and powerful technologies that can be used to deploy and manage intensive assessment tools. For example, many studies now allow participants to complete assessments via their personal smartphones (*vs.* paper-and-pencil or study-owned device), and professional companies can program and oversee survey delivery for a fee (e.g., LifeData). Depending on the nature of a given research question, intensive assessment methods have unique advantages over approaches such as cross-sectional, longitudinal, and group-based experimental designs (2). For example, intensive assessment methods such as ecological momentary assessment (EMA) (3) allow for differentiation of variance in experiences and relations with outcomes, at the levels of the person (i.e., stable individual differences) and of the day, hour, or moment (i.e., fluctuation within the same person over short intervals) (4). As intensive assessments often take place in participants' daily lives, rather than in research centers or clinics, an additional advantage is greater ecological validity (5).

Yet, designing these studies requires unique attention to item construction, response framing, instructions to participants, and the experiences of the target population (6). Although recommendations exist for how to approach such decisions, there is little guidance for developing, adapting, or refining items for optimal use in intensive designs. Existing items have been developed to capture experiences over a wide range of time frames, from the current day to the current moment. For a given study, item stems may need revision to focus on the appropriate time frame (e.g., the last several hours) and to capture the real-world experience expected in that temporal window (7). Additionally, response options must be provided that closely match the operationalization of the construct as it occurs in this window. Also critical are that participants are able to understand the items and identify relevant experiences in order to report on them accurately, as researchers are not available to clarify questions or response options in the moment. Importantly,

these experiences and the language participants use to describe them likely will differ between populations sampled; this suggests that the language researchers use to provide instructions and the wording of survey items also should be tailored to the population of interest (8).

Despite such challenges of intensive assessment, this approach provides a window into the real-world environments that researchers believe are key to promoting health and healthy behaviors. For example, perceptions of the social environment often are of interest in health-related research, as these perceptions can influence key health outcomes (e.g., via stress responses or motivation for/engagement in healthy behavior). Such perceptions include social comparisons [i.e., self-evaluations relative to others (9)] and social interactions (positive *vs.* negative encounters with others). Existing intensive assessment work shows that these experiences vary based on individual differences and, critically, environmental and contextual shifts (10,11), suggesting there is a range of possibilities for defining these experiences and framing the associated survey items.

For example, social comparison can occur quickly and automatically, without much conscious processing (12), which may present difficulties with recognizing that they have occurred and then reporting on them, even after short periods of time (e.g., hours). Several studies also specify reporting on a limited number of comparison domains [e.g., appearance (13)] or indicate interest in only one direction of comparison, such as comparisons with others who are better off than the self [upward comparison (14)] or worse-off than the self [downward comparison (15)]. Some existing evidence also indicates that participants may be reluctant to report that they make social comparisons, due to seeing them as socially undesirable (16,17). With respect to social interactions, the exact definition of an interaction varies between studies (e.g., active face-to-face communication only *vs.* passive exposure online), and some protocols differentiate between the occurrence or number (quantity) of interactions and the perceived intensity of one or more interactions [quality (18)]. Accurately capturing these nuances and ensuring that participants understand the kinds of experiences queried in surveys may benefit from both advance consideration and explicit testing, with iterations to improve on an initial protocol.

Case example: intensive assessment among midlife women with elevated risk for cardiovascular disease (CVD)

Individuals with chronic medical conditions may be harder to reach, have lower socioeconomic status, and/or have lower tech literacy than healthy individuals (19-21). As such, decisions about intensive assessment methodology are particularly important for ensuring appropriate language and instructions that match patients' experiences. One such population is midlife women who have chronic conditions that increase their risk for CVD (e.g., hypertension, type 2 diabetes). This is a large and diverse group that has high healthcare utilization and costs (22), despite prevention and intervention efforts to lower their health risks. Existing evidence indicates that midlife women's cardioprotective behaviors, such as physical activity engagement and dietary choices, are influenced by their perceptions of their social environments [e.g., social support for healthy behavior, negative judgments from others (23-25)]. As both social perceptions and health behaviors vary within and across days, weeks, and months, as well as between people (26-29), 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 (2,30). 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 (31). Previous work has shown that intensive assessment (specifically, EMA) is feasible and acceptable among midlife women for capturing self-perceptions [e.g., of self-efficacy or physical activity engagement (32)]. To our knowledge, however, no intensive assessment study has focused on social comparisons or social interactions among midlife women with elevated CVD risk; consequently, there is little available information to guide decisions about such assessments with this population.

Aims of the present research

In protocols that employ intensive assessment designs such as EMA, definitions of and introductions to concepts such as social comparisons and social interactions may affect reporting, and thereby, affect conclusions about these experiences and their relations with health behaviors in at-risk groups. However, few studies have systematically evaluated approaches to honing item wording and participant instructions to capture these experiences. A better understanding of how to word and introduce EMA items among midlife women (or other populations of interest) is critical to ensuring the accuracy of intensive assessment reports, and to the validity of conclusions about relations between these reports and health behaviors in participants' daily lives. With these points in mind, the present series of studies was designed to inform refinements to intensive assessment items for use with midlife women experiencing elevated CVD risk. The ultimate goal of this work is to investigate between- and within-person relations between social perceptions and physical activity in this population, using an EMA design (31). In Study 1, we used EMA to pilot items with this population for seven days and collected feedback about their experiences with these items. For Study 2, we used a qualitative interview method to elicit additional feedback about the items and suggestions for improvements. Finally, for Study 3, we used a seven-day EMA protocol to evaluate the performance of modified items, relative to the original items. We present the following article in accordance with the STROBE reporting checklist for cohort studies (available at <http://dx.doi.org/10.21037/mhealth-20-143>).

Methods

These studies were conducted in accordance with the Declaration of Helsinki (as revised in 2013) and were approved by institutional ethics committees at The University of Scranton (Scranton, PA, Study 1; no number associated) and Rowan University (Glassboro, NJ, Studies 2 and 3; Pro2018002377). Written informed consent was taken from all individual participants.

Study 1: pilot EMA (Phase I) and exit interviews

Initial EMA items were generated based on literature review and the authors' experience with the constructs

Construct	Study 1	Study 3
Social comparisons		
Occurrence/non-occurrence	<i>Since you woke up/in the last 3 hours</i> , did you think about or evaluate yourself in comparison to someone else (or someone else in comparison to yourself)?	<i>Since you woke up/in the last 3 hours</i> , did you think about or evaluate yourself or your behavior in comparison to someone else (or someone else in comparison to yourself)?
Number of comparisons (overall)	How many times did you compare yourself to someone else <i>since you woke up/in the last 3 hours</i> ?	How many times did you compare yourself to someone else <i>since you woke up/in the last 3 hours</i> ?
Number of upward comparisons	[Number of] Comparisons to people who are superior, better, or more desirable than I am	[Number of] Comparisons to people who seem to be doing better than I am
Number of lateral comparisons	[Number of] Comparisons to people who are about the same as I am	[Number of] Comparisons to people who seem to be doing about the same as I am
Number of downward comparisons	[Number of] Comparisons to people who are inferior, worse, or less desirable than I am	[Number of] Comparisons to people who seem to be doing worse than I am
Social interactions		
Occurrence/non-occurrence of positive interactions	Did you have <u>positive</u> interactions with any of the following [people] <i>today/in the last 3 hours</i> ?	Did you have <u>positive</u> or pleasant social experiences with any of the following [people] today/in the last 3 hours?
Number of positive interactions	Total number of positive interactions (<i>since you woke up/in the last 3 hours</i>)	Total number of times you had positive/ pleasant social experiences since you woke up/in the last 3 hours
Occurrence/non-occurrence of negative interactions	Did you have <u>negative</u> interactions with any of the following [people] <i>today/in the last 3 hours</i> ?	Did you have <u>negative</u> or unpleasant social experiences with any of the following [people] today/in the last 3 hours?
Number of negative interactions	Total number of negative interactions (<i>since you woke up/in the last 3 hours</i>)	Total number of times you had negative/ unpleasant social experiences since you woke up/in the last 3 hours

Figure 1 Item wording for ecological momentary assessments of social comparisons and social interactions. Bolded text indicates item wording changes between Study 1 and Study 3. Italic text indicates the time frames of the corresponding reporting periods. Underlined text indicates emphasis added to differentiate items about similar experiences.

of interest (10,13,33-37). All items were indicated the reporting period as “since you woke up” (first survey of the day each day) or “in the last three hours” (all other surveys). Intentions for social comparison items were for participants to report on any instance of comparing an aspect of themselves or their behavior to that of others during the relevant reporting period. This included comparisons that were noteworthy for prompting an emotional reaction, but such a reaction was not a criterion for inclusion (38); comparisons might have remained salient by the time for the survey prompt for a range of other reasons, including providing useful information or confirming an existing opinion (39). Intentions for social interactions were for participants to report on any experience with another person or people that they perceived positively or negatively. This included both protracted interactions (such as meetings or conversations) and briefer encounters (such as someone holding a door or behaving rudely while driving). Both social comparisons and social interactions were intended to be unrestricted with respect to modality,

and were meant to include exposures that occurred face-to-face, via telephone, via traditional media such as television or magazines, and via online platforms such as Facebook. The initial items appear in *Figure 1*.

Data collection occurred from September 2017 to June 2018. The target sample size was 10–15 participants, selected to maximize the utility of both quantitative and qualitative data that could be collected on a pre-specified timeline. Recruitment used print and electronic advertisements, including materials placed in primary care offices in a small northeastern U.S. city and surrounding suburbs, to attract women between the ages of 40 and 60 who had one or more of the following health conditions: hypertension or prehypertension, type 2 diabetes or prediabetes, high cholesterol (hypercholesterolemia or hyperlipidemia), metabolic syndrome, or current smoker (or quit smoking in the past 3 months). A total of 17 individuals expressed interest in participating and were screened for eligibility; two declined to participate at the initial screening call and one was scheduled to begin but

did not attend her appointment. Participants who were confirmed as eligible and completed the study ($n=13$; $M_{\text{Age}}=47$, $M_{\text{BMI}}=33.7 \text{ kg/m}^2$) were predominantly Caucasian (77%) and had household incomes less than \$75,000 per year (62%). The largest subset was married (46%) and 39% had less than a bachelor's-level education. High cholesterol was the most frequent CVD risk condition (62%), followed by (pre)hypertension (54%; see *Table 1*) and the number of risk conditions per participant ranged from 1 to 3 ($M=1.77$).

After completing a telephone screening call to verify eligibility, participants attended an initial setup appointment to provide written informed consent, have their height and weight measured by research staff, and receive training in the EMA protocol. Training introduced the experiences to be assessed in momentary surveys (e.g., interactions with other people) but did not provide detailed definitions of these constructs or guidance for determining whether certain experiences “counted” toward estimated totals. Participants then engaged in signal-contingent EMA recording for the following seven days (i.e., completing surveys in response to prompts or signals), with signals to complete five surveys per day. Signals were sent to participants' personal smartphones as text messages with embedded links to electronic surveys, which appeared in browser windows when selected. Survey prompts were separated by at least 3.5 hours and survey schedules were adjusted to align with participants' typical sleep/wake times. Schedules did not differ between weekdays and weekends. Participants were asked to complete each survey within one hour of receiving it. At the end of seven days, participants returned for a face-to-face exit interview that focused on eliciting qualitative feedback about the EMA process and item wording. Participants received \$25 gift codes for Amazon.com for completing the study. Overall EMA survey completion was 84%.

Statistical analysis

Descriptive statistics, including moment- and day-level means for items of interest and variability at the between- and within-person levels [intraclass correlation coefficients (ICCs)], were used as initial indicators of item performance. Multilevel models in SAS 9.4 (PROC MIXED; SAS Institute, Cary, N.C.) with maximum likelihood estimation were employed to account for the nested structure of the data (i.e., moments within days within participants).

Study 2: qualitative interviews

Items of interest were revised based on findings from Study

1 and subjected to qualitative feedback from the target population. Data were collected from October to December 2018. The sample size was not pre-specified; recruitment stopped when feedback reached saturation. Women who met the same criteria for Study 1 were recruited via print and electronic advertisements. Active recruitment also took place in primary care clinics in a mix of urban and suburban locations in the northeastern U.S., where patients' medical charts were screened for eligibility and patients meeting the above criteria were offered the opportunity to learn about the study from research staff. A total of 19 individuals expressed interest in participating and four women did not return our calls for screening. Of the 15 participants screened for eligibility, one was scheduled but then cancelled due to surgery, and four were scheduled but did not attend their appointments. The final sample included 10 women ($M_{\text{Age}}=52$, $M_{\text{BMI}}=29.8 \text{ kg/m}^2$). As in Study 1, this new sample was predominantly Caucasian (80%) and married (60%); 40% had household incomes less than \$100,000 per year and received less than a bachelor's-level education. Type 2 diabetes/prediabetes was the most frequent CVD risk condition (60%), followed by high cholesterol and (pre)hypertension (50% each; see *Table 1*). The number of CVD risk conditions per participant ranged from 1 to 4 ($M=2.11$). After providing written informed consent, participants engaged in one-hour interviews with research staff, which took place either at the research center or in the primary care clinic. Height and weight were measured by research staff for those attending at the research center; these measurements were taken from medical charts for patients who participated in the clinic.

Individual interviews focused on eliciting feedback about EMA measures, including how items were interpreted, confusion about item wording, the appropriateness of response options, and any hesitation to provide accurate answer. Participants were asked not to provide their immediate answers to the items, but to verbalize their reactions to each item and their process for determining how they would answer each one. Participants received \$25 for completing the interview. Recruitment stopped at 10 participants due to saturation. Emerging themes from these interviews were identified and categorized, and items were modified based on feedback.

Study 3: pilot EMA (Phase II) and comparison with Phase I

Women who met the same criteria for Studies 1 and 2 were recruited via print and electronic advertisements as well as

Table 1 Demographic information for participants in Studies 1, 2, and 3

Demographic characteristics	Study 1 (n=13)	Study 2 (n=10)	Study 3 (n=13)
Age	47.31 (5.22)	52.20 (6.49)	50.23 (5.25)
BMI (kg/m ²)	33.70 (9.27)	29.80 (4.37)	33.43 (5.22)
Number of CVD risk conditions	1.77 (0.93)	2.11 (1.17)	1.77 (0.93)
Racial/ethnic identification			
Hispanic/Latina	2 (15%)	2 (20%)	1 (8%)
Asian/Pacific Islander	1 (8%)	0 (0%)	0 (0%)
White	10 (77%)	8 (80%)	12 (92%)
Marital status			
Never married	4 (31%)	0 (0%)	3 (23%)
Divorced/separated	3 (23%)	3 (30%)	0 (0%)
Married	6 (46%)	6 (60%)	8 (62%)
Widowed	0 (0%)	1 (10%)	2 (15%)
Household income			
Less than \$25,000	1 (8%)	1 (10%)	1 (8%)
\$25,000-\$50,000	2 (15%)	0 (0%)	3 (23%)
\$50,000-\$75,000	6 (46%)	1 (10%)	0 (0%)
\$75,000-\$100,000	1 (8%)	2 (20%)	7 (54%)
\$100,000 or more	3 (23%)	6 (60%)	2 (15%)
Highest level of education attained			
High school graduate/GED	2 (15%)	3 (30%)	3 (23%)
Associate's/technical degree or partial college	3 (23%)	1 (10%)	1 (8%)
Bachelor's degree	2 (15%)	2 (20%)	5 (38%)
Graduate/professional degree	6 (46%)	4 (40%)	4 (31%)
CVD risk condition			
High cholesterol	8 (62%)	5 (50%)	6 (46%)
(Pre)hypertension	7 (54%)	5 (50%)	7 (54%)
Prediabetes/type 2 diabetes	5 (38%)	6 (60%)	6 (46%)
Metabolic syndrome	2 (15%)	0 (0%)	1 (8%)
Current smoker or quit in last 3 months	1 (8%)	3 (30%)	3 (23%)

Participants may have had more than one CVD risk condition (total percent >100). Data are shown as mean (SD) or frequency (%).

on-site recruitment in primary care clinics (with the same procedures as Study 2). Data collection occurred between January and April 2019; the target sample size was selected to match that of Study 1 (n=13). A total of 19 women expressed interest in participating, and again, four women did not return our calls for screening. Of the 15 individuals

who were screened, one withdrew from the study due to a personal matter and one declined to enroll. The majority of participants ($M_{Age} = 50$, $M_{BMI} = 33.4$ kg/m²) again were Caucasian (92%) and married (62%). The largest subset had less than a bachelor's level education (69%) and 30% had household incomes less than \$50,000 per year. (Pre)

hypertension (54%), type 2 diabetes (46%), and high cholesterol (46%) were the most frequent CVD risk conditions (see *Table 1*), and the number of risk conditions per participant ranged from 1 to 4 ($M = 1.77$).

As in Study 1, potential participants completed a telephone screening call to verify eligibility; those who were eligible and interested attended an in-person setup appointment with research staff. All participants provided written informed consent before receiving instructions for completing EMA surveys. These instructions included detailed review of how to define social “experiences” (i.e., the number of interactions, not people interacted with) and social comparisons and additional discussion to ensure understanding. Participants received handouts reviewing these concepts to take home with them. Over the next seven days, EMA prompts were sent to participants’ personal smartphones 5 times per day, with at least 3 hours between prompts. Prompts came as text messages with embedded links to electronic surveys. Survey schedules were matched to align with participants’ typical sleep/wake times and schedules did not differ between weekdays and weekends. Participants were again asked to complete each survey within one hour of receiving it. At the end of seven days, participants returned for a face-to-face exit interview. Participants received \$15 for completing the initial session and \$30 for engaging in EMA, with a \$10 bonus for completing 80% or more of the momentary surveys. The overall survey completion rate was 80%.

Statistical analysis

The same descriptive statistics presented for Study 1 were used as initial indicators of item performance and inferential tests examined differences in item performance between samples (Study 1 *vs.* Study 3). Given the greater emphasis on normalizing and clearly defining the experiences of negative social interactions and social comparisons in Study 3, we expected to observe higher likelihoods of occurrence and higher numbers of both events reported in Study 3 versus Study 1. In addition, given that the number of social interactions was specified as the number of interaction instances, rather than the number of people interacted with, we expected to observe lower likelihoods of occurrence and lower numbers of positive interactions in Study 3 versus Study 1. These hypotheses were tested using multilevel models to account for the nested structure of the data, using SAS 9.4 PROC GLIMMIX for reported occurrence (yes/no) and PROC MIXED for numbers reported; study was treated as a dichotomous variable.

Results

Study 1

Descriptive statistics for variables of interest are displayed in *Table 2*. All ICCs indicated meaningful variance at the within-person (moment) level, with the number of positive interactions showing the largest proportion of variability between-person. Participants reported making social comparisons on 50% of days; positive interactions were reported on 99% of days, whereas negative interactions were reported on only 36% of days. Means for the numbers of negative social interactions and social comparisons reported (overall and each direction) were below 1.0 per survey. For the total number of social comparisons, for example, this corresponded to 1.65 comparisons per day ($SE = 0.50$). In contrast, the mean number of positive social interactions was above 4.12 per survey ($SE = 0.81$), or 18.04 per day ($SE = 3.68$).

Exit interviews with participants revealed response styles along two noteworthy dimensions. First, 5/13 women indicated some hesitancy to make or report social comparisons, summarized by one woman’s assessment that “*I don’t worry about other people, I just do me.*” Probing with the remaining eight women revealed that several instances of comparison were not recognized as belonging to this category and likely were not reported. For example, all 13 women indicated that they occasionally looked to coworkers or friends as behavioral models of how to save time with work or household tasks (i.e., they recognized seeing these people as upward targets, or “doing better” than they were on these dimensions), though 10 said that they likely hadn’t counted such instances as comparison.

Similarly, discussions about responses to social interaction items suggested that at least 6/13 participants based their totals on the number of people involved in a given interaction, rather than the number of interaction instances. For example, one woman explained:

“I was at my niece’s birthday party and it was a nice time. I haven’t seen my family recently and I expected there to be lots of comments about me avoiding them, but everyone seemed happy to see me and happy to be there. Probably 15 people showed up. Yeah, I counted everyone in the house that day.”

Interestingly, when asked for examples of negative interactions, participants seemed to focus on individual interactions, rather than those with a group, even if groups were involved. One woman provided the following example:

“My boss just got on my case. He was really angry about a small mistake, and I don’t think he knew people could hear him.”

Table 2 Descriptive statistics for social comparisons and social interactions (moment level) and comparisons between Study 1 and Study 3

Social experiences	Study 1		Study 3		Between-study difference: t[df], P, effect size (OR/d)
	B (SE)	ICC	B (SE)	ICC	
Social comparisons					
Occurrence/non-occurrence	1.30 (0.56)		1.79 (0.61)		t[116]=0.62, P=0.54, OR =1.63
Number of comparisons (overall)	0.39 (0.12)	0.22	0.42 (0.10)	0.16	t[150]=0.14, P=0.71, d=0.14
Number of upward comparisons	0.13 (0.05)	0.18	0.20 (0.07)	0.14	t[81]=1.51, P=0.14, d=0.49
Number of lateral comparisons	0.26 (0.10)	0.15	0.11 (0.04)	0.50	t[84]=-4.65, P<0.001, d=0.93
Number of downward comparisons	0.04 (0.03)	0.05	0.11 (0.04)	0.08	t[75]=1.59, P=0.12, d=0.52
Social interactions					
Occurrence/non-occurrence of positive interactions	3.05 (0.70)		2.71 (0.64)		t[127]=0.44, P=0.67, OR =0.72
Number of positive interactions	4.12 (0.81)	0.52	2.39 (0.27)	0.16	t[131]=-1.86, P=0.06, d=0.56
Occurrence/non-occurrence of negative interactions	0.33 (0.23)		2.26 (0.40)		t[127]=4.21, P<0.001, OR =6.67
Number of negative interactions	0.18 (0.16)	0.15	0.54 (0.07)	0.02	t[143]=3.18, P=0.002, d=0.75

ICC, intraclass correlation coefficient.

Not just other people who work there—customers, too, and a bunch of people made comments and gave me weird looks after. It was awful... I said that was one interaction in the survey that day, the one with my boss. But I guess I could have counted the ones with workers and customers, too.”

Thus, findings from Study 1 raised the possibility that reports for the frequency of negative interactions and social comparisons were too low, and offered insight into reasons for this pattern. Although the observation of more positive than negative interactions is consistent with some previous intensive assessment work (33,37), midlife women in this study reported a higher overall frequency of positive interactions and a larger discrepancy between the reported numbers of positive and negative interactions than participants in previous studies. It is possible that perceptions of social interactions differ between midlife women with CVD risk conditions from those of other populations (e.g., younger adults). However, qualitative feedback suggested that the definition of an interaction could be clearer—specifically, whether to count the number of people involved in a social interaction or the number of individual instances (regardless of the number of people), which seemed to differ between reports of positive and negative interactions.

With respect to social comparisons, reports of their overall frequency among midlife women with CVD risk conditions was similar to published reports with other populations. Yet, many of these used methods distinct

from those employed in the present study. For example, several studies have used *event-contingent recording* of social comparisons with college students (29,40), which requires participants to recognize when they make a comparison and complete a report in response to this awareness. Event-contingent recording may require more cognitive effort on the part of the participant than responding to signals, particularly when the experience of interest is expected to occur with some regularity (41), and signal-contingent methods may provide less biased aggregated estimates of frequency (1). Other EMA studies have focused on a restricted range of comparison targets or dimensions, such as comparisons with romantic partners (42) or comparisons of appearance (14,43). Given both the repeated prompting to complete surveys and the unrestricted range of comparison dimensions in the present study, it is reasonable to expect detection of higher numbers of comparisons than in previous work.

Again, it is possible that the population of interest for this study makes comparisons at a different rate than those assessed previously, and we are unaware of any previous intensive assessment studies of social comparison frequency with midlife women to use as a benchmark. However, qualitative feedback offered alternative explanations for our quantitative results, such as a perception that making social comparisons is undesirable and lack of awareness of one's own comparison activity. As noted, participants received little guidance to define social constructs, and it

is unclear whether more explicit instruction could affect recording (10). Based on the exit interviews, it was clear that additional revision of items could be useful. We conducted a qualitative follow-up study using cognitive interviewing to clarify perceptions of item construction, response framing, and instructions.

Study 2

Predominant themes emerging from qualitative interviews were (I) confusion about the constructs identified by each item, (II) perceptions that reports of specific experiences represented negative judgments about the self or others, and (III) the amount of effort or time required for deciding on each response. With respect to confusion, 6/10 participants expressed a desire to clarify the boundaries of “social interactions.” For example, one woman stated:

“I wouldn’t have included bumping into someone who said something rude, or people making comments on Facebook. I thought this meant longer conversations, like, being with the other person.”

Similarly, one participant said:

“Oh okay, ‘interaction’ made me think it had to be face to face, you know?”

Suggestions regarding this point (i.e., “interaction” as indicating in-person communication) focused on modifying the wording of the item to ask about “social experiences,” with additional explanation during setup appointments. Subsequent inquiries about such changes were met with positive feedback, and 7/10 participants endorsed changing the wording to assess “social experiences.” The other 3 participants expressed no preference but indicated that this change would not be problematic. Further feedback about items related to this construct indicated that the descriptors “positive” and “negative” also restricted responses. As one woman put it:

“I would have thought you meant something specific, like a birthday [positive] or a death in the family [negative]. Something major. But sounds like you mean anything I like or don’t like, right?”

When asked whether adding the descriptors “pleasant” and “unpleasant” to these items would help to clarify, 8/10 participants said that they would prefer these changes and that they would make the intent of the items clearer. For example:

“I’d say yes to having an unpleasant experience, but I don’t know about negative. What does that say about me? And pleasant just sounds better, too.”

The remaining 2 participants reported that these changes would be agreeable but not necessary.

Participants showed similar, and perhaps stronger, responses to social comparison items. Initially, 3/10 participants claimed that they never made comparisons, and 2/10 reported that they made comparisons but attempted not to focus on them. In response to research staff providing examples of comparisons (as in Study 1), however, all 10 women indicated that they had these experiences. For example, research staff offered the example of seeing a coworker as more efficient or effective than the self and learning helpful tips from that person. As one woman stated:

“Oh gosh, yeah that’s happened to me. And I can see it happening again. I didn’t think about it as comparing myself—that [comparing myself] sounds bad, but the work example isn’t. Maybe I just needed it explained.”

In fact, 8/10 women indicated that additional explanation to clarify “comparing myself” would be helpful, and with this explanation, they could imagine reporting the occurrence of comparison at least once per day. Regarding specific item wording, 6/10 women also expressed concern about the descriptors “superior” and “inferior” to describe comparison targets. One woman explained:

“That sounds so judge-y. I don’t know everything about their circumstances, so maybe I shouldn’t judge. How do I know that I’m actually superior?”

When the intent of this language was clarified by research staff—indicating that these are just descriptions of how we see the world, not judgments about others—4/10 women explicitly recommended changing the wording to clarify. All 10 women responded positively to the suggestion that the wording be changed to “comparisons to people who seem to be doing better than I am” (for superior/upward) and “comparisons to people who seem to be doing worse than I am” (for inferior/downward).

Finally, 5/10 women expressed concern that estimating the number of interactions and comparisons they experienced would take a great deal of time and effort. All 5 of these women, as well as the remaining 5, conveyed relief in response to the explanation that they should spend no more than a few seconds on these items, which were intended to capture their subjective assessments and recollections of interactions and comparisons that remained salient at the end of a 3-hour period (rather than a precise account of every experience).

Consequently, qualitative interviews provided specific avenues for improving both item wording and instructions.

A subset of participants exhibited the anticipated reluctance to report on (or lack of awareness of) their social comparisons, and some participants also expressed hesitation to report on social interactions. With additional discussion and guidance, however, most hesitation resolved fairly quickly, and awareness of comparisons increased. These findings suggested that an introduction to the concepts of social comparisons and interactions, with examples as described above, could be extremely useful for improving the accuracy of EMA reports. Findings also implied that the aforementioned wording changes, summarized in *Figure 1*, could further improve reporting patterns by increasing consistency between the intention of each item and the items as they appeared in each momentary survey. Consequently, specific discussion and instructions regarding social perceptions were added to the initial setup appointment (with checks to ensure correct understanding), and this appointment was supplemented with a handout that summarized each construct and instructions for completing surveys (e.g., “choose the number of social experiences/comparisons that seems right to you—don’t worry about the exact number”). Participants were encouraged to consult these handouts if they experienced confusion during days of EMA data collection in Study 3.

Study 3

Table 2 shows descriptive statistics for variables of interest and comparisons for these variables between studies. As in Study 1, ICCs in Study 3 indicated meaningful variance at the within-person (moment) level; the ICC for positive interactions dropped from 0.52 in Study 1 to 0.16 in Study 3, and in Study 3, the number of lateral comparisons (i.e., those perceived as doing “about the same” as the self) showed largest proportion of variability between-person. Participants reported making social comparisons on 58% of days (*vs.* 50% in Study 1; $\chi^2=1.12$, $P=0.29$). Positive interactions were again reported on 99% of days ($\chi^2=0.02$, $P=0.88$), and negative interactions were reported on 76% of days (*vs.* 36% in Study 1; $\chi^2=25.96$, $P < 0.001$).

Further, the odds of reporting the occurrence of a comparison at a given survey (*vs.* reporting that no comparisons occurred) were 1.63 times higher in Study 3 than in Study 1 (see *Table 2*). This (non-significant) difference in reporting produced an average of 1.77 surveys per participant with (*vs.* without) reports of comparisons occurring in Study 3, relative to 1.23 per participant in Study 1 ($d=0.33$). The average number of comparisons

reported per survey was 0.42 in Study 3, relative to 0.39 in Study 1, corresponding to a daily difference of 1.65 to 2.0 comparisons per day (Study 1 *vs.* Study 3, respectively). This pattern was similar for upward and downward comparisons, with higher average numbers of comparisons per survey and day in Study 3 versus Study 1 (see *Table 2*). Reporting an occurrence of upward comparison was 4.04 times higher in Study 3 than in Study 1, and reporting an occurrence of downward comparison was 2.34 times higher in Study 3 than Study 1. None of these between-study differences were statistically significant, however (all $P>0.30$). In contrast, the odds of reporting an occurrence of lateral comparisons was 0.11 times lower in Study 3 than Study 1, and the average number of lateral comparisons reported per survey was lower in Study 3 than in Study 1 ($d=0.93$; see *Table 2*).

With respect to social interactions, as expected, the odds of reporting the occurrence of a negative interaction (versus reporting that no negative interactions occurred) at a given survey were 6.67 times higher in Study 2 than in Study 1 ($P<0.001$). The average number of negative interactions per survey was significantly higher in Study 3 than in Study 1 (i.e., 0.18 *vs.* 0.54 interactions per survey; $d=0.75$; see *Table 2*). Conversely, the odds of reporting the occurrence (*vs.* non-occurrence) of a positive interaction was 0.28 times lower in Study 3 than in Study 1 ($P=0.67$), and the average number of positive interactions per survey was noticeably lower in Study 3 than in Study 1 (i.e., 4.12 *vs.* 2.39 per survey; $d=0.56$).

Although some of the observed effects were not statistically significant (possibly due to modest sample sizes for each study), direct comparisons between findings from Studies 1 and 3 showed differences in the expected directions. Specifically, participants reported making social comparisons and having negative social experiences more often in Study 3 than in Study 1. Of note, the frequency of reporting lateral comparisons decreased from Study 1 to Study 3, whereas the frequencies of reporting upward and downward comparisons increased. Given that participants in these studies were different groups of midlife women who were recruited from separate geographic areas, it is possible that the observed discrepancies between studies were due to actual differences between these women’s experiences. As the eligibility criteria and recruitment methods were similar across studies, and as the two geographic locations were similar with respect to demographics, however, it is more likely that the revised wording of items assessing upward comparisons, downward comparisons, and negative interactions reduced reluctance to report on these

experiences (see *Figure 1*). Conversely, reports of positive experiences were somewhat lower in Study 3 than Study 1. This may have been due to the modified instructions for reporting on social interactions (i.e., explicit guidance to report the number of events, rather than the number of others interacted with) and/or revised item wording (see *Figure 1*).

Discussion

Taken together, this series of studies presents a process for adapting and evaluating items for intensive assessments within a specific population of interest. As this type of formative research is recommended but not frequently published, our goal was to demonstrate one possible approach to iterative revisions of item instructions, item content, and response options (6). This approach employed a combination of quantitative and qualitative methods to understand the language participants used to describe their social perceptions and translate that into intensive assessments that best capture the experiences of interest. We now discuss the implications of this work for intensive assessments with midlife women with elevated CVD risk, as well as for intensive assessments more generally.

Previous pilot work with healthy midlife women showed that intensive assessments of experiences associated with physical activity (e.g., self-efficacy for activity) were feasible and acceptable (32). Although item performance was evaluated in this research, there was no description of modifications made to item wording or instructions or the process for evaluating the effects of these changes. As such, the present study extended this work in several ways. First, it focused on midlife women with elevated CVD risk (i.e., those with chronic health conditions), who may differ from healthy women in ways that are important to technology-based intensive assessment protocols [e.g., (44)]. Second, this series of studies showed the feasibility and acceptability of intensive assessments to this specific, at-risk population, and focused on gathering their feedback to inform adaptations to the method. Third, the influence of these adaptations was explicitly evaluated, and showed improved item performance that was associated with moderate-to-large effect sizes.

A critical implication of this work is that language mattered, in both the instructions to participants and item construction. With respect to instructions, it was important to use language that helped participants understand the experiences of interest. Although detailed training and

instructions often are provided to participants in intensive assessment studies (45,46), exit interviews that explored the types of experiences participants believed to be relevant were key in determining the limitations of our protocol. For example, it was not clear to participants that comparisons of workplace behavior, which may be beneficial for improving their work performance, should be counted as an instance of social comparison. Further discussions with participants assisted in refining instructions and improving the ability of our assessments to match more closely with their lived experiences. Direct comparisons of responses between Studies 1 and 3 supported the notion that changes resulted in better alignment with expected frequencies of social comparisons and interactions occurring in daily life.

With respect to item construction, participants' feedback indicated that the language used in our initial items could lead to social desirability bias in reporting, particularly for upward and downward social comparisons and negative social interactions. This type of bias might lead to an underreporting of the experiences of interest if participants perceive their reports as reflecting negatively on themselves, but could lead to overreporting of other types of behaviors (e.g., positive health behaviors). For example, participants tend to overreport engaging in healthy eating behaviors (i.e., eating more vegetables and less fat) relative to their actual food intake as assessed by biomarkers of nutrition (47). Although the researcher is not present when assessments are completed in daily life, as they may be in laboratory- or clinic-based research, similar presentation biases may remain active for participants and could reduce the ecological validity of assessments (7). Attention to this possibility in future work could improve the accuracy of participants' reports of social perceptions and related experiences.

In this series of studies, the authors made deliberate efforts to recruit diverse groups of midlife women, and were able to attract small subsets who identified as women of color and disadvantaged socioeconomic categories. As is common in intensive assessment research, however, participants were predominantly Caucasian and well-educated (10,48). Their experiences and interpretations of item wording and instructions may not represent those of the larger population, and it will be critical for future formative work with intensive assessment protocols to address more diverse perspectives.

In addition, the present approach employed both quantitative and qualitative methods to inform decisions about protocol adaptations, and the process reported here was restricted to understanding the frequencies of

and levels of variability in participants' reports of social perceptions. 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 dimension, the original theoretical model indicated that comparisons are made primarily on the bases of abilities and opinions [or attitudes; (9)], though subsequent evidence has shown that the range of specific comparison dimensions is much wider [e.g., personality, appearance, wealth; (29)]. 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.

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. These limitations notwithstanding, the present series of studies highlights the utility of multi-stage, multi-method formative mHealth research with a specific population of interest. Additional work is needed to further explore approaches to this preliminary stage of intensive assessment work, and to understand optimal item construction and participant instructions for distinct populations.

Acknowledgments

A subset of these findings were presented as a poster at the 2019 biannual meeting of the Society for Ambulatory Assessment (Syracuse, NY). The authors would like to thank Kristen Pasko, M.A. for her assistance with data collection and management.

Funding: This work was supported by the U.S. National Heart, Lung, and Blood Institute (grant number K23

HL136657) to Danielle Arigo and internal funding awarded to Danielle Arigo. The second author's time was supported by the U.S. National Institute on Aging (grant number R01 AG062605) to Jacqueline A. Mogle.

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist for cohort studies. Available at <http://dx.doi.org/10.21037/mhealth-20-143>

Data Sharing Statement: Available at <http://dx.doi.org/10.21037/mhealth-20-143>

Peer Review File: Available at <http://dx.doi.org/10.21037/mhealth-20-143>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/mhealth-20-143>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. These studies were conducted in accordance with the Declaration of Helsinki (as revised in 2013) and were approved by institutional ethics committees at The University of Scranton (Scranton, PA, Study 1; no number associated) and Rowan University (Glassboro, NJ, Studies 2 and 3; Pro2018002377). Written informed consent was taken from all individual participants.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Moskowitz DS, Russell JJ, Sadikaj G, et al. Measuring people intensively. *Can Psychol* 2009;50:131-40.
2. Dunton GF. Ecological momentary assessment in physical

- activity research. *Exerc Sport Sci Rev* 2017;45:48-54.
3. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. *J Happiness Stud* 2003;4:35-52.
 4. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu Rev Psychol* 2011;62:583-619.
 5. Smyth JM, Juth V, Ma J, et al. A slice of life: Ecologically valid methods for research on social relationships and health across the life span. *Soc Pers Psychol Compass* 2017;11:e12356.
 6. Conner TS, Lehman BJ. Handbook of research methods for studying daily life. *Getting started: Launching a study in daily life*; New York, NY: The Guilford Press; 2012:89-107.
 7. Schwarz N. The science of real-time data capture: Self-reports in health research. Chapter 2, Retrospective and concurrent self-reports: the rationale for real-time data capture. New York, NY: Oxford University Press, Inc.; 2007:11-26.
 8. Uher J. Quantitative data from rating scales: an epistemological and methodological enquiry. *Front Psychol* 2018;9:2599.
 9. Festinger L. A theory of social comparison processes. *Hum Relat* 1954;7:117-40.
 10. Arigo D, Mogle JA, Brown MM, et al. Methods to assess social comparison processes within persons in daily life: a scoping review. *Front Psychol* 2020;10:2909.
 11. Liu H, Xie QW, Lou VW. Everyday social interactions and intra-individual variability in affect: a systematic review and meta-analysis of ecological momentary assessment studies. *Motiv Emotion* 2019;43:339-53.
 12. Gilbert DT, Giesler RB, Morris KA. When comparisons arise. *J Pers Soc Psychol* 1995;69:227-36.
 13. Leahey TM, Crowther JH, Mickelson KD. The frequency, nature, and effects of naturally occurring appearance-focused social comparisons. *Behav Ther* 2007;38:132-43.
 14. Pila E, Barlow MA, Wrosch C, et al. Comparing the body to superior others: Associations with daily exercise and body evaluation in men and women. *Psychol Sport Exerc* 2016;27:120-7.
 15. Affleck G, Tennen H, Urrows S, et al. Downward comparisons in daily life with chronic pain: Dynamic relations with pain intensity and mood. *J Soc Clin Psychol* 2000;19:499-518.
 16. Helgeson VS, Taylor SE. Social comparisons and adjustment among cardiac patients 1. *J Appl Soc Psychol* 1993;23:1171-95.
 17. Wills TA. Downward comparison principles in social psychology. *Psychol Bull* 1981;90:245-71.
 18. Zhaoyang R, Sliwinski MJ, Martire LM, et al. Social interactions and physical symptoms in daily life: Quality matters for older adults, quantity matters for younger adults. *Psychol health* 2019;34:867-85.
 19. Berkman ND, Sheridan SL, Donahue KE, et al. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011;155:97-107.
 20. Fisher EB, Coufal MM, Parada H, et al. Peer support in health care and prevention: Cultural, organizational, and dissemination issues. *Annu Rev Public Health* 2014;35:363-83.
 21. Shaw KM, Theis KA, Self-Brown S, et al. Peer reviewed: Chronic disease disparities by county economic status and metropolitan classification, behavioral risk factor surveillance system, 2013. *Prev Chronic Dis* 2016;13:E119.
 22. Centers for Medicare & Medicaid Services [Internet]. HealthData.gov 2018 [cited 8 Sep 2020]. Available online: <https://healthdata.gov/agencies/centers-medicare-medicaid-services>
 23. Im EO, Ko Y, Chee E, et al. Clusters of midlife women by physical activity and their racial/ethnic differences. *Menopause* 2017;24:417-25.
 24. Janssen I, Dugan SA, Karavolos K, et al. Correlates of 15-year maintenance of physical activity in middle-aged women. *Int J Behav Med* 2014;21:511-8.
 25. Walsh A, Simpson EEA. Health cognitions mediate physical (in) activity and walking in midlife women. *Maturitas* 2020;131:14-20.
 26. Elavsky S, Kishida M, Mogle JA. Concurrent and lagged relations between momentary affect and sedentary behavior in middle-aged women. *Menopause* 2016;23:919-23.
 27. Reichenberger J, Richard A, Smyth JM, et al. It's craving time: Time of day effects on momentary hunger and food craving in daily life. *Nutrition* 2018;55-56:15-20.
 28. Vella EJ, Kamarck TW, Shiffman S. Hostility moderates the effects of social support and intimacy on blood pressure in daily social interactions. *Health Psychol* 2008;27:S155-62.
 29. Wheeler L, Miyake K. Social comparison in everyday life. *J Pers Soc Psychol* 1992;62:760-73.
 30. Curran PJ, Howard AL, Bainter SA, et al. The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. *J Consult Clin Psychol* 2014;82:879-94.
 31. Arigo D, Brown MM, Pasko K, et al. Rationale and design of the Women's Health And Daily Experiences Project:

- Protocol for an ecological momentary assessment study to identify real-time predictors of midlife women's physical activity. *JMIR Res Protoc* 2020;9:e19044.
32. Ehlers DK, Huberty J, Buman M, et al. A novel inexpensive use of smartphone technology for ecological momentary assessment in middle-aged women. *J Phys Act Health* 2016;13:262-8.
 33. Arigo D, Pasko K, Mogle JA. Daily relations between social perceptions and physical activity among college women. *Psychol Sport Exerc* 2020;47:101528.
 34. Arigo D, Cavanaugh J. Social perceptions predict change in women's mental health and health-related quality of life during the first semester of college. *J Soc Clin Psychol* 2016;35:643-63.
 35. Dunton GF, Atienza AA, Castro CM, et al. Using ecological momentary assessment to examine antecedents and correlates of physical activity bouts in adults age 50+ years: a pilot study. *Ann Behav Med* 2009;38:249-55.
 36. Schumacher LM, Thomas C, Ainsworth MC, et al. Social predictors of daily relations between college women's physical activity intentions and behavior. *J Behav Med* 2021;44:270-6.
 37. Zhaoyang R, Sliwinski MJ, Martire LM, et al. Age differences in adults' daily social interactions: an ecological momentary assessment study. *Psychol Aging* 2018;33:607-18.
 38. Patrick H, Neighbors C, Knee CR. Appearance-related social comparisons: the role of contingent self-esteem and self-perceptions of attractiveness. *Pers Soc Psychol Bull* 2004;30:501-14.
 39. Suls J, Martin R, Wheeler L. Social comparison: Why, with whom, and with what effect? *Curr Dir Psychol Sci* 2002;11:159-63.
 40. Locke KD. Connecting the horizontal dimension of social comparison with self-worth and self-confidence. *Pers Soc Psychol Bull* 2005;31:795-803.
 41. Wheeler L, Reis HT. Self-recording of everyday life events: Origins, types, and uses. *J Pers* 1991;59:339-54.
 42. Pinkus RT, Lockwood P, Schimmack U, et al. For better and for worse: Everyday social comparisons between romantic partners. *J Pers Soc Psychol* 2008;95:1180-201.
 43. MacIntyre RI, Heron KE, Braitman AL, et al. An ecological momentary assessment of self-improvement and self-evaluation body comparisons: Associations with college women's body dissatisfaction and exercise. *Body Image* 2020;33:264-77.
 44. Zhang Y, Lauche R, Sibbritt D, et al. Comparison of health information technology use between American adults with and without chronic health conditions: Findings from the National Health Interview Survey 2012. *J Med Internet Res* 2017;19:e335.
 45. Hufford MR. The science of real-time data capture: Self-reports in health research. Chapter 4, Special methodological challenges and opportunities in ecological momentary assessment. New York, NY: Oxford University Press, Inc.; 2007:54-75.
 46. Scott SB, Graham-Engeland JE, Engeland CG, et al. The Effects of Stress on Cognitive Aging, Physiology and Emotion (ESCAPE) Project. *BMC Psychiatry* 2015;15:146.
 47. Brunner E, Juneja M, Marmot M. Dietary assessment in Whitehall II: Comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. *Brit J Nutr* 2001;86:405-14.
 48. Zapata-Lamana R, Lalanza JF, Losilla JM, et al. mHealth technology for ecological momentary assessment in physical activity research: a systematic review. *PeerJ* 2020;8:e8848.

doi: 10.21037/mhealth-20-143

Cite this article as: Arigo D, Mogle JA, Brown MM, Gupta A. A multi-study approach to refining ecological momentary assessment measures for use among midlife women with elevated risk for cardiovascular disease. *mHealth* 2021;7:53.