MEASUREMENT COMPARABILITY OF ELECTRONIC AND PAPER ADMINISTERED VISUAL ANALOGUE SCALES: A REVIEW OF PUBLISHED STUDIES

Byrom B[1], Eremenco S[2], Muehlhausen W[3], Howry C[4], Watson C[5], Bodart S[6], Platko JV[7], Elash CA[8]

1: CRF Bracket, London, United Kingdom, 2: Critical Path Institute, Tucson, AZ, USA, 3: Muehlhausen Ltd, Cloughjordan, Ireland, 4: assisTek, Scottsdale, AZ, USA, 5: ERT, Nottingham, United Kingdom, 6: CRF, Bracket, Beaconsfield, QC, Canada, 7: CRF Bracket, Plymouth Meeting, PA, USA, 8: YPrime Inc, Malvern, PA, USA.

INTRODUCTION

Visual analogue scales (VASs) are commonly used in the assessment of a variety of health-related constructs including pain [1,2,3], mood [4], quality of life [5,6] and patient satisfaction [7], and have been found to provide reliable and valid data [8]. The VAS is defined in the FDA's PRO Guidance [9] as "*a line of fixed length (usually 100 mm) with words that anchor the scale at the extreme ends and no words describing intermediate positions.*" Traditionally, the VAS has been implemented on paper (pVAS) with a 100 mm (10 cm) line to facilitate measurement of 101 discrete points (0 to 100) with a metric ruler [10]. More recently, with the advent of electronic modes of clinical outcome assessment (COA) data collection such as electronic diaries, an eVAS can be implemented on a line of any length, provided the response fields allow for selection of precisely 101 discrete pixels or points on the line. VASs are brief and simple to administer and are particularly useful when assessing a single construct with many perceptible gradations due to the continuous nature of the scale and the 101 possible response options. These characteristics also contribute to the scale's high sensitivity to change [11]. Because the VAS uses few words, the vocabulary level of the respondent is generally not a concern, provided the anchor descriptors are straightforward, common terms. People with visual impairments can typically see a VAS easily, and most patients have sufficient dexterity to use the VAS [12], although patients with dexterity issues, for example due to arthritis or orthopaedic trauma to their hands or arms, may have difficulty targeting a very specific point on the scale to indicate their desired response [11]. While the VAS is commonly used, some research indicates that children younger than 6 or 7 may have difficulty using the VAS [13]. In addition, older adults may have more difficulty using a VAS than younger respondents [11], with some research showing increased age associated with incorrect response on a paper VAS [14].

The VAS has additional limitations related to the visual nature of the scale. For example, a VAS is not appropriate for telephone interview-based or interactive voice response (IVR) system- or voice assistant-based data collection because the respondent must be able to see the scale in order to select a response. Measurement of the score on paper with a metric ruler may introduce human error [12] and additional scoring error may be introduced if the response is ambiguous, i.e., not a clear mark within the confines of the scale anchors [8]. Additionally, photocopying the pVAS can change the length of the line, making the comparison between distances measured on the original and the photocopied version difficult [11,14]. These limitations of the pVAS can be mitigated by implementation on a screen-based electronic platform (see Figure 1). With an eVAS, the pixels along the VAS line are invisibly categorized into 101 equal regions so that position of the response can be converted into a numeric score from 0 to 100. This approach prevents ambiguous or invalid responses and eliminates the need for manual measurement with a ruler, hence removing that source of human error. Although eVAS is now commonplace, and there is a plethora of data in the published literature to demonstrate the psychometric equivalence of pVAS and eVAS scores [15,16,17], researchers still regularly question whether eVASs with various line lengths are psychometrically comparable to the 10 cm pVAS.

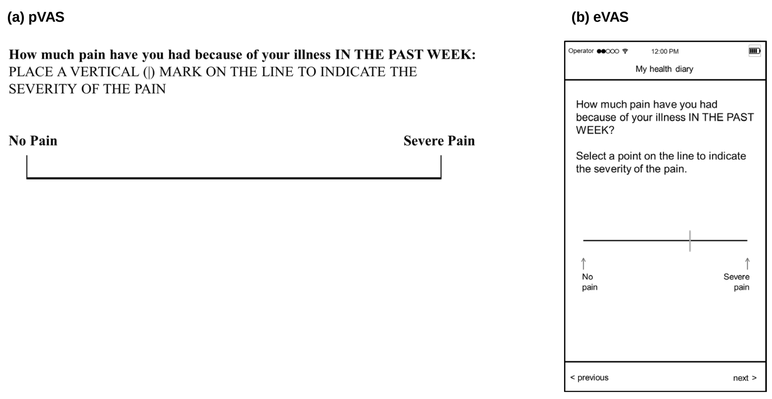

Figure 1. Typical representations of pVAS and eVAS based on The Stanford Health Assessment Questionnaire Disability Index HAQ-DI) [18]

OBJECTIVE

The objective of the current work was to review publications comparing data collected using 100 mm pVAS and eVASs of various lengths on different electronic platforms to answer the question whether line length and mode of data collection (pVAS or eVAS) affect scale measurement properties.

METHODS

We performed an informal literature search for studies evaluating the eVAS compared to pVAS. The search was conducted using the OneSearch library database provided by Nottingham Trent University (Nottingham, UK) [19], and used search terms: "visual analogue scale" and "equivalence" or "validation." Articles providing quantitative comparison between eVAS and pVAS were added to our list of previously-identified publications.

The following information (when available) was catalogued from each article: participant characteristics - disease indication, age range, sample size; eVAS and pVAS length; electronic device used; patient-reported outcome measures (PROMs) studied; and findings and conclusions of the evaluation.

RESULTS

Nineteen articles (Table 1) were included in the evaluation of measurement comparability [1-4, 20-34]. Eight studies (42%) were conducted in general population volunteers; the remainder included patients with pain (21%), rheumatoid arthritis (16%), panic disorder (5%), multiple sclerosis (5%), and non-small cell lung cancer (5%). One study was conducted in 8 to 10 year olds [30], and the remaining studies included adults up to 86 years old. All studies were crossover comparisons of paper to at least one form of electronic data collection. Studies varied in size from 12 to 355 participants (median: 65 participants). VAS items included measures of pain, fatigue, global health, appetite, anxiety, and alcohol effects.

Electronic modes included: personal digital assistant (PDA) (n=8), personal computer (PC) (n=5), smartphone (n=3), tablet (n=3), feature-phone (n=2), and smartwatch (n=1) (Figure 2). eVAS length was not reported for 5 studies, and ranged from 2 to 4 cm (n=3, min: 2.1 cm), 4.1 to 6 cm (n=5), 6.1 to 8 cm (n=3), 8.1 to 10 cm (n=2), and > 10 cm (n=2, max: 28.9 cm) in the remainder (Figure 3).

Authors of all studies (19/19) concluded pVAS and eVAS were comparable. However, 2 studies (11%) reported trends towards higher scores on eVAS vs. pVAS [1,2], 3 (16%) towards lower scores on eVAS [4,22,30], and 2 studies (11%, Apple Newton and Palm device) indicated eVAS scores may be lower than pVAS at the scale ends [31,34]. A later study reported no scale-end effects [21]. Differences between eVAS and pVAS were considered not clinically relevant although 3 publications (16%) recommended paper and electronic versions should not be used interchangeably [30,31,34] because of these trends indicating systematically higher or lower scores by format.

Ref.	n	Population	Age	pVAS	eVAS	Electronic mode *	PROM **
[1]	30	Rheumatoid arthritis	Range: 49-70	100 mm	-	PDA	Pain, fatigue, global health
[2]	38	Rheumatoid arthritis	Mean: 58 [SD = 13]	100 mm	-	PDA	Pain, fatigue, global health
[3]	155	Chronic pain	Range: 19-69	100 mm	40-80 mm	SP, Tablet	Pain
[4]	71	Panic disorder and healthy	Range: 17-72	100 mm	200 mm	PC	Anxiety
[20]	35	General population	Range: 22-62	100 mm	40 mm	FP	-
[21]	30	General population	Medians: 34 (M), 31 (F)	100 mm	70 mm	PDA	Appetite
[22]	22	General population	Range: 56-86	100 mm	100 mm	Tablet	Pain
[23]	24	General population	Range 19-57	100 mm	50 mm	PDA	-
[24]	88	Rheumatoid arthritis and Psoriatic arthritis	Mean: 54 [SD = 11]	-	-	PC	Pain
[25]	189	Chronic pain	Range: 18-82	-	-	PC	SF-MPQ (pain intensity)
[26]	86	Non-small cell lung cancer	-	100 mm	53 mm	PDA	LCSS
[27]	355	-	-	-	50-800 pixels	PC	-
[28]	200	Chronic pain	Mean: 56.5 [SD = 14]	100 mm	50 mm (est)	PDA	Pain
[29]	104	Multiple sclerosis and General population	Mean: 49 [SD = 9]	100 mm	55, 100 mm (est)	SP, Tablet	-
[30]	12	General population	Range: 8-10	100 mm	24 mm	SW	Appetite
[31]	12	General population	Mean: 30 [SD = 12]	100 mm	66 mm	PDA	Appetite
[32]	65	General population	Range: 19-54	100 mm	21 mm	FP	Alcohol effects
[33]	98	Chronic pain	Mean: 44 [SD = 16]	100 mm	289, 96 mm	PC, SP	-
[34]	20	General population	Mean: M: 37 [SD = 13] F: 32 [SD = 9]	100 mm	52 mm	PDA	Appetite

* FP = feature-phone , PDA = personal digital assistant, SP = smartphone, PC = personal computer, SW = smartwatch. ** SF-MPQ = Short form McGill Pain Questionnaire, LCSS = Lung Cancer Symptom Scale.

Table 1. Summary of articles Included

RESULTS (cont.)

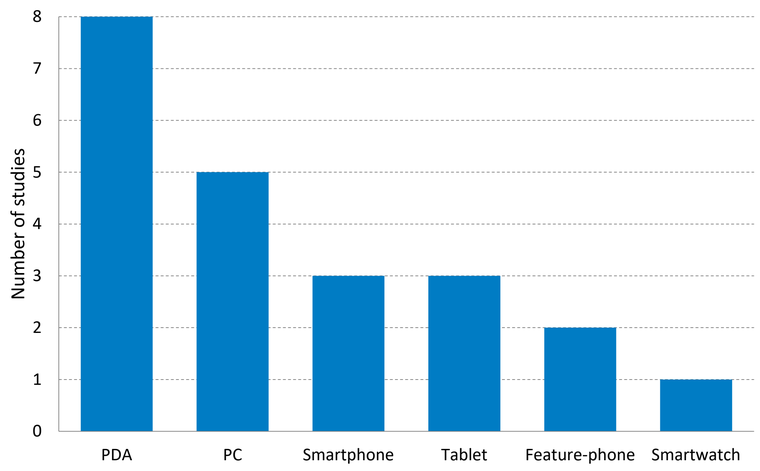

Figure 2. Modes of electronic VAS included in the review of comparability with paper

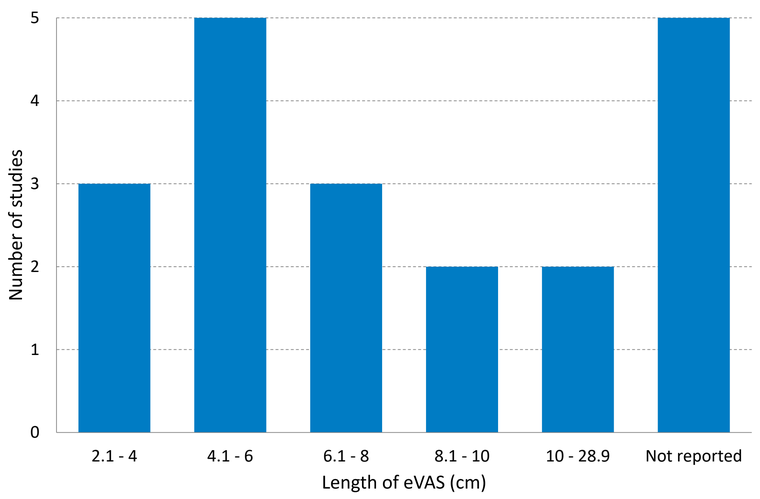

Figure 3. Lengths of electronic VAS included in the review of comparability with paper

DISCUSSION

Limitations of this research

- This literature review was not intended to be a comprehensive review of evidence regarding the comparison of eVAS with pVAS. Rather, it was designed to address whether there is sufficient evidence available to claim that pVAS are comparable to eVAS and to assess whether or not the length of the eVAS influences its measurement properties.
- Twelve articles (63%) were at least 10 years old and only 5 articles (26%) were published in the last 5 years. Similarly, Table 1 shows only 5 studies for which either a smartphone, a tablet, or a smartwatch was used. Considering the recent improvements in electronic data collection methods and technologies for clinical trials, it would be helpful to review more recent literature to corroborate these findings with current technology.
- This research is general and does not focus on specific therapeutic areas or populations. While one study presented findings in children aged 8 to 10 years old, subsequent reviews could focus on the use of VAS by older adults, children, or patients with visual impairment or other physical restrictions.

eVAS best practices implementation

- When migrated to an electronic format, a VAS needs to be implemented according to design best practices (e.g., ePRO Consortium [35]).
- Some articles [19,31] presented an implementation that did not strictly follow these best practices. However, it was not possible to verify best practice eVAS implementation for most of the studies reviewed. Departures from implementation best practice may potentially result in differences between eVAS and pVAS.

DISCUSSION (cont.)

Acceptability of the VAS

- No clinically relevant difference between pVAS and eVAS was found in this literature review.
- Some scale-end effects were detected in the studies using an Apple Newton and a Palm Pilot system for data collection. These are relatively old technology, and the design of the device may have led to this effect.
- Based on the publications reviewed, the length of the eVAS does not affect its measurement properties.
- Recent review articles [8,10] suggest superiority of the numeric rating scale (NRS) to the VAS in adult pain measures because it was found easier to administer and score, and had both higher patient acceptability and better psychometric properties. These review articles also suggest the VAS to be more difficult for certain patient populations to understand. While we believe this is likely, more evidence is needed to substantiate this assertion.

Recommendations for future VAS use

- The use of a VAS should be carefully evaluated for implementation in a clinical trial as some patients may have difficulty understanding the scale or may be unable to use it because of visual or physical impairment.
- The ePRO Consortium does not recommend the use of VAS for newly developed PROMs because of the aforementioned difficulties for some patients and the inability to implement a VAS scale on an IVRS system. While IVRS is less used, in future we may see greater adoption of voice assistants such as Amazon Alexa or Google Assistant for data collection in clinical studies.

CONCLUSIONS

There is evidence in the literature supporting the comparability of eVAS and pVAS regardless of the VAS length. When implementing a VAS on a screen-based electronic mode, it is critical to follow industry best practices for faithful migration to diminish clinically meaningful statistical differences with pVAS.

ACKNOWLEDGMENTS

Critical Path Institute is supported, in part, by Critical Path Public-Private Partnerships Grant Number U18 FD005320 (effective 2015-2020) from the U.S. Food and Drug Administration. Support for the ePRO Consortium comes from membership fees paid by members of the ePRO Consortium (https://c-path.org/programs/epro/).

REFERENCES

[1] Kvien TK, Mowinckel P, Heiberg T et al. Ann Rheum Dis 2005; 64: 1480-1484.
[2] Heiberg T, Kvien TK, Dale O et al. Arthritis Rheum 2007; 57: 454-460.
[3] Byrom B, Doll H, Muehlhausen W et al. Value Health 2018; 21: 581-589.
[4] van Duinen M, Rickelt J and Griez E. Prog Neuropsychopharmacol Biol Psychiatry 2008; 32: 1045-1047.
[5] deBoer AGEM, van Lanschot JJB, Stalmeier PFM et al. Qual Life Res 2004, 13:311-320.
[6] Petersohn S, Ramaekers BLT, Olie RH et al. Qual Life Res. 2019 Mar 30. doi: 10.1007/s11136-019-02166-0. [Epub ahead of print]
[7] Voutilainen A, Pitkäaho T, Kvist T et al. J Adv Nurs. 2016 Apr;72(4):946-57.
[8] Safikhani S, Gries KS, Trudeau JJ et al. J Patient Rep Outcomes. 2018 Sep 6;2:40.
[9] Food and Drug Administration 2009.
[10] Gries K, Berry P, Harrington M et al. J Patient Rep Outcomes. 2018 Sep 6;2:41.
[11] Briggs M and Closs JS. J Pain Symptom Manage. 1999 Dec;18(6):438-46.
[12] Flaherty SA. AANA J. 1996 Apr;64(2):133-40.
[13] Shields BJ, Palermo TM, Powers JD et al. Child Care Health Dev. 2003 Jul;29(4):281-90.
[14] Jensen MP, Karoly P and Braver S. Pain. 1986 Oct;27(1):117-26.
[15] Campbell N, Ali F, Finlay AY et al. Qual Life Res 2015; 24: 1949-1961.
[16] Muehlhausen W, Doll H, Quadri N, et al. Health Qual Life Outcomes 2015;13:167-187.
[17] Gwaltney CJ, Shields AL, Shiffman S. Value Health 2008;11: 322-333.
[18] Bruce B and Fries JF. Health Qual Life Outcomes 2003; 1: 20- 25
[19] https://ntu-primo.hosted.exlibrisgroup.com/primo-explore/search?vid=NTU_VU10 (Accessed 27MAR19)
[20] Kreindler D, Levitt A, Woolridge N et al. Psychiatry Res 2003; 120: 165-177.
[21] Almiron-Roig E, Green H, Virgili R et al. Appetite 2009; 53: 465-468.
[22] Bird M-L, Callisaya ML, Cannell J et al. Interact J Med Res 2016;5(1):e3.
[23] Jamison RN, Gracely RH, Raymond SA et al. Pain 2002; 99: 341-347.
[24] Cunha-Miranda L, Barcelos F, Miguel C et al. Ann Rheum Dis 2014; 73: 1155.
[25] Cook AJ, Roberts DA, Henderson MD et al. Pain 2004; 110: 310-317.
[26] Hollen PJ, Gralla RJ, Stewart JA et al. Support Care Cancer 2013; 21: 165-172.
[27] Reips U-D and Funke F. Behav Res Methods 2008; 40: 699-704.
[28] Junker U, Freynhagen R, Längler K et al. Curr Med Res Opin 2008; 24: 1797-1806.
[29] Kos D, Raeymaekers J, Van Remoortel A et al. Clin Rehabil 2017; 31: 1215-1225.
[30] Rumbold PLS, Dodd-Reynolds CJ and Stevenson E. Appetite 2013; 69: 180-185.
[31] Stratton RJ, Stubbs RJ, Hughes D et al. Eur J Clin Nutr 1998; 52: 737-741.
[32] Tiplady B, Cairns W, Sturdee M et al. Conference presentation: Psychobiology Section of the British Psychological Society, 2005, Low Wood, Windermere, UK.
[33] Delgado DA, Lambert BS, Boutris N et al. JAAOS 2018; 2: e088.
[34] Whybrow S, Stephen JR and Stubbs JR. Eur J Clin Nutr 2006; 60: 558-560.
[35] https://c-path.org/wp-content/uploads/2018/09/BestPractices2_Response_Scales.pdf (Accessed 27MAR19)

For further information, please contact:
Bill Byrom, bill.byrom@crfhealth.com

ISPOR 2019 Annual Congress
New Orleans, LA, USA
18-22 May 2019