American Journal of Epidemiology
Vol. 150, No. 10
Printed in U.S.A.

Bias and Precision in Visual Analogue Scales: A Randomized Controlled Trial

Agnès Paul-Dauphin, Francis Guillemin, Jean-Marc Virion, and Serge Briançon

Various types of visual analogue scales (VAS) are used in epidemiologic and clinical research. This paper reports on a randomized controlled trial to investigate the effects of variations in the orientation and type of scale on bias and precision in cross-sectional and longitudinal analyses. This trial was included in the pilot study of the SU.VI.MAX (supplementation by antioxidant vitamins and minerals) prevention trial in France in 1994. Six types of VAS (simple, middle-marked, graphic rating, graduated, graduated-numbered, and numerical rating) and two orientations (horizontal and vertical) were used to measure three symptoms of ear, nose, and throat infection at 2-month intervals in 870 subjects. Differences between scales were analyzed by comparing variances (Levene's test) and means (variance-covariance analysis for repeated measures). Scale characteristics were shown to influence the proportion of zero and low values (i.e., there was a floor effect), but not mean scores. The precision of measurements varied cross-sectionally according to the type of scale, but no differences were observed in the precision of measurement of change over time. In conclusion, the characteristics of VAS seem to be important in cross-sectional studies, particularly when symptoms of low or high intensity are being measured. Researchers should try to reach a consensus on what type of VAS to use if studies are to be compared. *Am J Epidemiol* 1999;150:1117–27.

bias, epidemiology; measurement error; pain measurement; visualization

Visual analogue scales (VAS) are often used in epidemiologic and clinical research to measure the intensity or frequency of various symptoms, particularly pain. They are generally completed by patients themselves but are sometimes used to elicit opinions from health professionals (1, 2). VAS are more sensitive to small changes than are simple descriptive ordinal scales (3) in which symptoms are rated, for example, as mild or slight, moderate, or severe to agonizing (4, 5).

The most simple VAS is a straight horizontal line of fixed length, usually 100 mm. The ends are defined as the extreme limits of the parameter to be measured (symptom, pain, health) (6) orientated from the left (worst) to the right (best). In some studies, horizontal scales are orientated from right to left, and many investigators use vertical VAS (4). Scott and Huskisson (7) reported no difference between horizontal and vertical VAS in a survey involving 100 subjects, but other authors have suggested that the two orientations differ with regard to the number of possible angles of view (8). Reproducibility has been shown to vary along a vertical 100-mm VAS (9) and along a horizontal VAS (10). The choice of terms to define the anchors of a scale has also been described as important (4, 5).

VAS can be presented in a number of ways, including the following: scales with a middle point, graduations or numbers (numerical rating scales), meter-shaped scales (curvilinear analogue scales), "box-scales," scales consisting of circles equidistant from each other (one of which the subject has to mark), and scales with descriptive terms at intervals along a line (graphic rating scales or Likert scales) (4). Numerical rating scales or number scales (11) consist of numbers without a line, although the term is also sometimes used to refer to graphic rating scales.

Comparisons of measurements usually show good correlations between types of scale (7, 10–14). One study (15) analyzing the preferences of subjects for 12 different scales failed to find any universal favorite but observed that numerical rating scales with descriptive terms were preferred with regard to ease of use and accuracy of representation.

Other than in a paper by Cline et al. (16), no recommendations have been drawn in the literature about the effect of the presentation of a VAS on its metric prop-

Received for publication March 4, 1998, and accepted for publication March 15, 1999.

Abbreviation: VAS, visual analogue scale(s).

From the Ecole de Santé Publique, Faculté de Médecine, Vandoeuvre-les-Nancy, France.

Reprint requests to Dr. S. Briançon, Ecole de Santé Publique, Faculté de Médecine, 9 avenue de la Forêt de Haye, B.P. 184, 54505 Vandoeuvre-les-Nancy Cedex, France.

erties or about the consequences of choosing a particular type.

The aim of the present study was to investigate whether variations in orientation and type of VAS influence bias or precision of measurement in cross-sectional and longitudinal analyses of data from a randomized controlled trial. Six different VAS and two orientations were used to measure three symptoms of ear, nose, and throat infection at 2-month intervals.

MATERIALS AND METHODS

Population

The study sample consisted of 1,128 males and females involved in a pilot study of a prevention trial (SU.VI.MAX) (17). They were aged 35–61 years and drawn from the general population living in France.

Measures

All subjects were mailed questionnaires relating to age, sex, and assessment of health status. Direct questions were asked about the presence of any chronic ear, nose, and throat disease and the occurrence of ear, nose, and throat infection within the previous 2 months. In addition, a VAS was used to explore three ear, nose, and throat symptoms, specifically nasal obstruction (blocked nose), runny nose, and sore throat.

Six types of scale were used (figures 1 and 2):

- simple VAS
- middle-marked VAS
- graphic rating scale (a graduated verbal descriptive scale, with 10-mm graduations and the following descriptive terms along the line: not at all, a little, moderately, a lot, enormously or hugely)
- VAS graduated every 10 mm but without numbers
- VAS graduated and numbered every 10 mm
- numerical rating scale with numbers every 10 mm but no line

Scales could be orientated either horizontally or vertically. They were all 100 mm long and accompanied by text describing the extreme limits from "not at all" (no symptom) to "enormously" (for blocked nose) or "hugely" (for runny nose and sore throat) (the worst possible level of the symptom). Subjects were instructed to put a mark on the scale according to the intensity of their symptoms. Results were expressed in millimeters from zero (no symptom) to 100 (worst possible level of the symptom).

The same questionnaire, including the same questions and the same VAS, was sent at two time points (T1 and T2) 2 months apart.

Randomization

Each subject was assigned a unique type and orientation of VAS by two-level randomization, in a six by two factorial design stratified by age and sex. The first level of randomization assigned the orientation (horizontal or vertical), and the second level assigned the type of scale and was balanced every six subjects.

Statistical analysis

Five variables were derived according to the presence or absence of the following characteristics of the scales:

- middle mark: presence (middle-marked VAS, graphic rating scale, graduated VAS, graduated-numbered VAS, numerical rating scale) versus absence (simple VAS)
- numbers: presence (graduated-numbered VAS, numerical rating scale) versus absence (simple VAS, middle-marked VAS, graphic rating scale, graduated VAS)
- graduations: presence (graphic rating scale, graduated VAS, graduated-numbered VAS, numerical rating scale) versus absence (simple VAS, middle-marked VAS)
- text along the line: presence (graphic rating scale) versus absence (simple VAS, middle-marked VAS, graduated VAS, graduated-numbered VAS, numerical rating scale)
- line: presence (simple VAS, middle-marked VAS, graphic rating scale, graduated VAS, graduated-numbered VAS) versus absence (numerical rating scale)

Groups were compared using the chi-square test for qualitative variables and analysis of variance for quantitative variables.

Because of the presence of a high proportion of zero ratings for all symptoms, and because it was assumed that there is error on neither true zero nor true 100 scores (12, 18), but that some zero ratings might be the result of an error on the low scores, separate analyses of the proportion of zero ratings and of low ratings (from one to nine) were conducted.

Proportions of zero values and of values from one to nine, according to orientation and type and characteristics of the scales, were compared by analysis of covariance with repeated measures. Adjustment was made for chronic disease and recent ear, nose, and throat infection.

Measurements with a rating of nine or less at time 1, time 2, or both were excluded from further analyses. At the other extremity of the scales, the small proportion of values higher than 90 did not permit the same

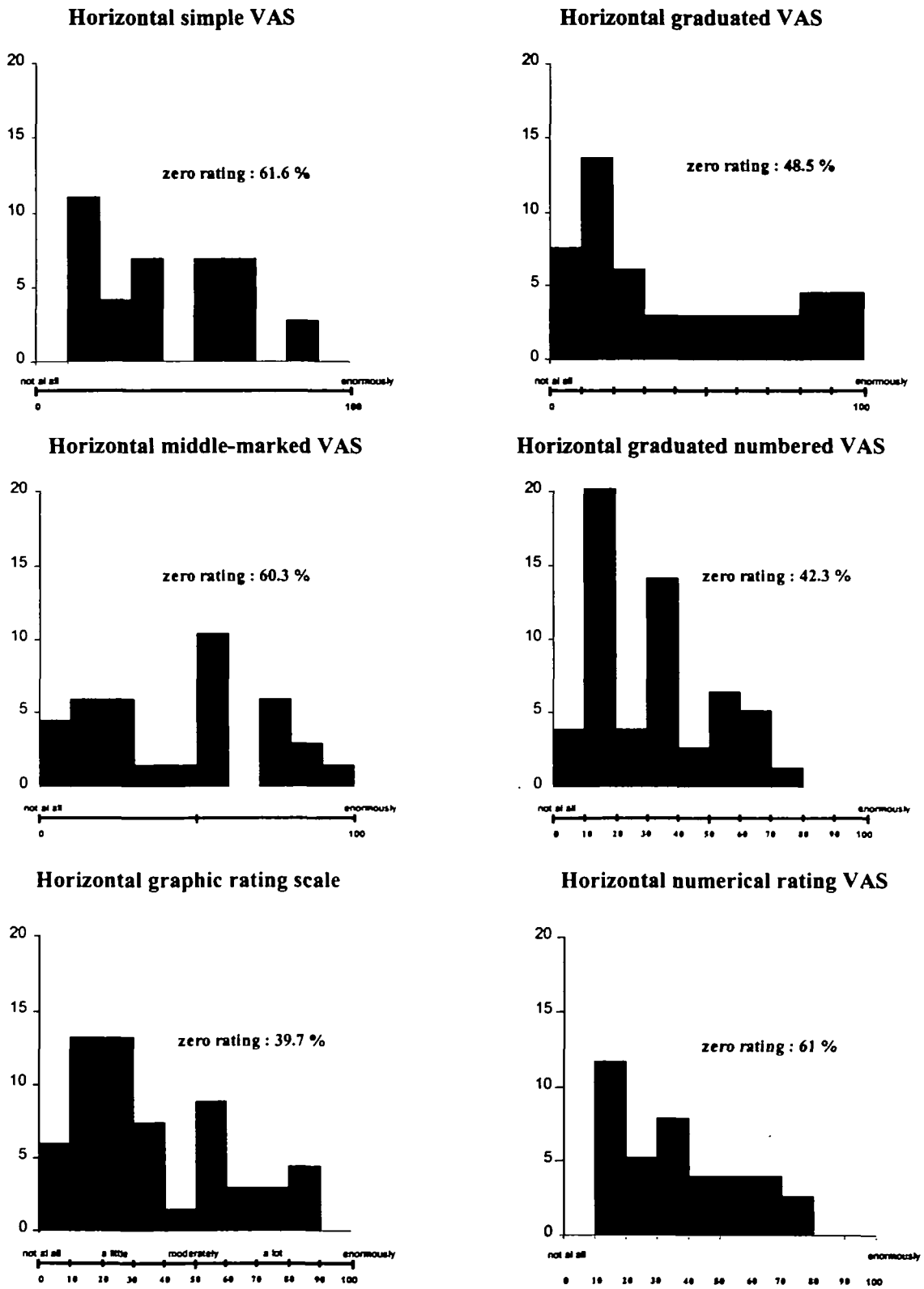

FIGURE 1. Distribution of non-zero ratings on horizontal scales for the symptom "blocked nose" at time 1, SU.VI.MAX prevention trial pilot study, France, 1994. SU.VI.MAX, supplementation by antioxidant vitamins and minerals; VAS, visual analogue scales. Percentages of zero rating appear in the text on each graph.

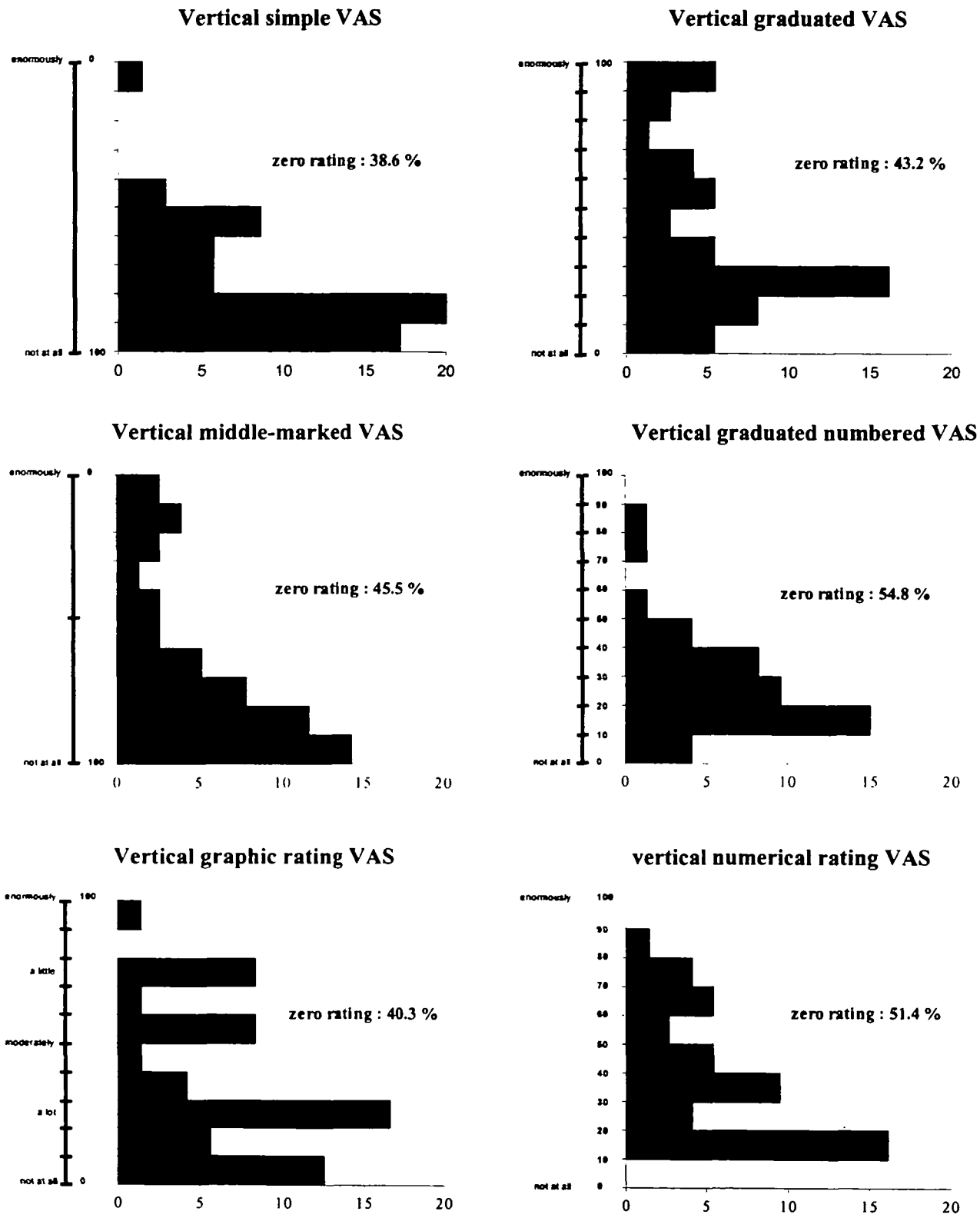

FIGURE 2. Distribution of non-zero ratings on vertical scales for the symptom "blocked nose" at time 1, SU.VI.MAX prevention trial pilot study, France, 1994. SU.VI.MAX, supplementation by antioxidant vitamins and minerals; VAS, visual analogue scales. Percentages of zero rating appear in the text on each graph.

analysis to be conducted as for small values; however, subjects with values higher than 90 were excluded from further analysis for similar reasons.

Using only subjects with ratings between 10 and 90 at the two measurement times, mean score levels were compared by analysis of variance-covariance with repeated measures. Adjustment was made for chronic disease and recent ear, nose, and throat infection, allowing for a cross-sectional comparison (between subjects) of means of the sum at time 1 and time 2 and a longitudinal comparison (within subjects) of means of the difference between times 1 and 2.

The precision of the cross-sectional measurement and of the measurement of change between time 1 and time 2 was analyzed according to the different types of scale for each symptom by comparing variances using Levene's test of equality of variances.

Because of the large number of statistical tests, $p < 0.01$ was considered significant, and analyses and hypotheses corresponding to these p values are to be interpreted strictly on an exploratory basis.

All statistical analyses were performed using BMDP statistical software (19).

RESULTS

Of the 1,128 subjects included, 963 returned the first questionnaire and 903 (80 percent) returned both. A total of 870 respondents (77 percent) completed all three ear, nose, and throat symptom scales at both time 1 and time 2 and were included in the analysis.

Characteristics of these subjects according to the randomization groups are presented in table 1. No statistically significant differences were observed between groups with regard to age, sex, or ear, nose, and throat infection/chronic disease.

Regardless of the group and the symptom, between time 1 and time 2 the percentage of zero ratings decreased, the percentage of ratings between one and nine increased, and the mean values for ratings between 10 and 90 decreased. The distributions of the non-zero ratings for the symptom "blocked nose" on each scale at time 1 are presented in figures 1 and 2.

The percentage of zero ratings varied according to symptoms from 49.0 percent to 60.1 percent at time 1 and from 45.4 percent to 61.4 percent at time 2 (table 2). Orientation had no apparent influence on the mean percentage of zero ratings at the two measurements, but it had a significant effect in the longitudinal analysis: the percentage of zero ratings decreased between time 1 and time 2 on the horizontal scales but increased on the vertical scales (table 3).

The type of scale had a significant effect on the cross-sectional measurement of symptoms, with higher percentages of zero ratings being observed on

TABLE 1. Characteristics of the subjects according to groups of randomization as defined by orientation and type of visual analogue scale (VAS), SU.VI.MAX* prevention trial pilot study, France, 1994

Orientation of scales	Type of scales	No.	Mean age (years)	Sex (% male)	ENT* infection 2 months before first questionnaire (%)	ENT infection 2 months before second questionnaire (%)	Presence of an ENT chronic disease (%)
All		870	48.6 (6.3)†	40.8	51.0	49.2	14.3
Horizontal	Simple VAS	73	49.0 (6.5)	37.0	49.3	53.4	21.9
	Middle-marked VAS	68	48.1 (6.4)	45.6	50.0	41.2	13.2
	Graphic rating scale	68	49.1 (6.9)	41.2	50.0	47.1	10.3
	Graduated VAS	66	48.6 (6.5)	33.3	48.5	50.0	13.6
	Graduated-numbered VAS	78	49.4 (6.0)	44.9	44.9	56.6	15.4
	Numerical rating scale	77	49.2 (6.2)	42.9	46.8	53.2	13.0
	All	430	48.9 (6.4)	40.9	48.1	49.8	14.7
Vertical	Simple VAS	70	47.8 (6.6)	42.9	61.4	52.9	11.4
	Middle-marked VAS	77	48.0 (6.1)	41.6	59.7	49.4	14.3
	Graphic rating scale	72	47.3 (6.1)	36.1	55.6	43.1	9.7
	Graduated VAS	74	49.2 (6.7)	41.9	45.9	47.3	12.2
	Graduated-numbered VAS	73	48.9 (6.0)	41.1	39.7	47.9	12.3
	Numerical rating scale	74	49.0 (6.1)	40.5	60.8	51.6	23.0
	All	440	48.6 (6.3)	40.7	53.9	48.6	13.9

* SU.VI.MAX, supplementation by antioxidant vitamins and minerals; ENT, ear, nose, and throat.
† Numbers in parentheses, standard deviation.

TABLE 2. Percentage of zero ratings according to orientation, type, and characteristic of the visual analogue scale (VAS), SU.VI.MAX* prevention trial pilot study, France, 1994

	No.	Time 1			Time 2		
		Blocked nose	Runny nose	Sore throat	Blocked nose	Runny nose	Sore throat
All	870	49.0	50.7	60.1	45.4	46.6	61.4
Orientation							
Horizontal	430	52.2	53.5	63.0	43.0	42.8	58.8
Vertical	440	45.7	48.0	57.3	47.7	50.2	63.9
Type							
Simple VAS	143	50.3	46.9	58.0	43.4	43.4	59.4
Middle-marked VAS	145	52.4	55.2	60.0	49.0	51.0	66.2
Graphic rating scale	140	40.0	45.7	55.0	37.9	40.0	49.3
Graduated VAS	140	45.7	49.3	58.6	44.3	42.1	58.6
Graduated-numbered VAS	151	48.3	53.0	64.9	45.0	47.0	65.6
Numerical rating scale	151	56.3	53.6	63.6	52.3	55.0	68.2
Characteristic							
Middle mark							
Yes	727	48.7	51.4	60.5	45.8	47.2	61.8
No	143	50.3	46.9	58.0	43.4	43.4	59.4
Number							
Yes	302	52.3	53.5	64.2	48.7	51.0	66.9
No	568	47.2	49.3	57.9	43.7	44.2	58.5
Graduations							
Yes	582	47.8	50.5	60.7	45.0	45.0	60.7
No	288	51.4	51.0	59.0	46.2	46.2	62.8
Text							
Yes	140	40.0	45.7	55.0	37.9	37.9	49.3
No	730	50.7	51.6	61.0	46.8	46.8	63.7
Line							
Yes	719	47.4	50.1	59.4	43.9	43.9	59.9
No	151	56.3	53.6	63.6	52.3	52.3	68.2

* SU.VI.MAX, supplementation by antioxidant vitamins and minerals.

the graduated-numbered VAS and numerical rating scales and lower percentages on the graphic rating scale and graduated VAS. The differences were mainly related to the presence of numbers, text, or lines: percentages of zero ratings were higher for scales with numbers and lower for scales with text or lines. A significant interaction was seen between orientation and the presence of graduations for runny nose in cross-sectional analyses and for sore throat in longitudinal analyses ($p < 0.01$). In scales with graduations, the percentage of zero ratings was lower when the orientation was horizontal rather than vertical.

The percentage of ratings between one and nine varied from 6.2 percent to 7 percent at time 1 and from 12.8 percent to 13.6 percent at time 2 (table 4). In cross-sectional analysis, orientation had no effect on the average percentage of ratings between one and nine at the two measurements. However, a significant effect was seen in longitudinal analysis: the percentage of ratings between one and nine increased between time 1 and time 2 on horizontal scales but remained unchanged on vertical scales (table 5).

The type and characteristics of scales had a significant effect on cross-sectional measurement of symptoms. Proportions of values between one and nine were lower when scales had a middle, numbers, or graduations, but they were higher when the scales presented a text. No ratings between one and nine were observed on scales with no line. A significant interaction was observed between orientation and the presence of a middle mark ($p < 0.01$ for all three symptoms): on vertical scales, percentages of values between one and nine were lower when there was a middle mark, whereas there was only a weak difference for horizontal scales. An interaction was also observed between orientation and the presence of graduations for the three symptoms ($p < 0.01$): on vertical scales only, the proportion of small values was lower in the presence of graduations.

The analysis performed on zero ratings and ratings between one and nine combined showed neither an

TABLE 3. Effect of orientation, type, and characteristics of scales on the percentage of zero ratings, SU.VI.MAX* prevention trial pilot study, France, 1994

	p value for statistical significance of the tests† (*n* = 870)					
	Cross-sectional analysis			Longitudinal analysis		
	Blocked nose	Runny nose	Sore throat	Blocked nose	Runny nose	Sore throat
Orientation	‡			0.0056	0.0026	0.01
Type	0.0028	0.017	0.01			
Interaction orientation x type	0.09	0.0087	0.09			0.0036
Characteristics§						
Middle mark						
Numbers	0.05	0.04	0.007			
Graduations						
Text	0.0015	0.02	0.002			0.03
Line	0.003	0.01	0.03			
Interaction of each characteristic with orientation						
Middle mark			0.07	0.06		
Numbers		0.08		0.006		
Graduations	0.03	0.0016	0.02			0.005
Text	0.05	0.08				0.02
Line				0.09		

* SU.VI.MAX, supplementation by antioxidant vitamins and minerals.
† Analysis of variance-covariance for repeated measures, with adjustment for chronic disease and current infection.
‡ Blanks indicate a *p* value > 0.1.
§ Adjustment for orientation.

effect of orientation in cross-sectional analysis nor an interaction between orientation and time in longitudinal analysis. It did, however, reveal an effect of the type of scale for one symptom ($p < 0.01$) and an effect of the presence of graduations in all symptoms in cross-sectional analysis.

Table 6 presents the means and standard deviations of ratings between 10 and 90 for each of the three symptoms. The analysis showed no effect of orientation and type of scale, either cross-sectionally or longitudinally. In cross-sectional analysis, the different characteristics had no significant effect on mean ratings. In longitudinal analysis, there was a small, nonsignificant effect of orientation ($p < 0.05$) and type of scale ($p < 0.05$).

Comparison of the variance of ratings between 10 and 90 at the time 1 measurement according to the characteristics of scales showed an effect of the type of scale ($p < 0.002$ for the symptom "runny nose" with a higher variance on the graduated VAS) and an effect of the presence of numbers, which was associated with lower variances for two symptoms at time 1 ($p < 0.007$ and $p < 0.04$ for the symptoms "blocked nose" and "runny nose," respectively). Different characteristics had no apparent effect on the variance of the measurement of change between time 1 and time 2 (tables 6 and 7).

DISCUSSION

This is the first randomized trial to compare different types of scale both cross-sectionally and longitudinally. In previously published studies (1, 4, 5, 7, 10, 12–15, 20), different types of VAS were given to the same subjects, assuming that they were in a steady state and that the first answers would not be recalled when the second scale was being completed. The order of presentation was sometimes randomized (7, 8, 12, 14). However, this design did not allow investigators to clearly identify differences in level or precision either cross-sectionally or longitudinally. The design implemented here, despite lower levels of control for interindividual variability, permitted the results of each type of scale (category) to be differentiated, thereby elucidating the influence of the choice of a type of VAS on the level and precision of ratings.

Assume that the observed rating for a subject i on a VAS j is given as follows:

$$x_{ij} = \mu + s_i + e_{ij} \quad (1)$$

TABLE 4. Percentage of low ratings (between 1 and 9) according to orientation, type, and characteristics of scales, SU.VI.MAX* prevention trial pilot study, France, 1994

	No.	Time 1			Time 2		
		Blocked nose	Runny nose	Sore throat	Blocked nose	Runny nose	Sore throat
All	870	6.2	8.4	7.0	13.6	13.3	12.8
Orientation							
Horizontal	430	3.5	5.8	4.7	16.5	14.4	15.6
Vertical	440	8.9	10.9	9.3	10.7	12.3	10.0
Type							
Simple VAS*	143	8.4	15.4	13.3	17.5	16.8	15.4
Middle-marked VAS	145	9.7	12.4	7.6	20.0	20.0	16.6
Graphic rating scale	140	9.3	10.0	9.3	22.9	20.7	22.9
Graduated VAS	140	6.4	8.6	7.9	12.1	12.9	12.9
Graduated-numbered VAS	151	4.0	4.0	4.6	9.9	10.6	9.9
Numerical rating scale	151	0.0	0.7	0.0	0.0	0.0	0.0
Characteristics							
Middle mark							
Yes	727	5.8	7.0	5.8	12.8	12.7	12.2
No	143	8.4	15.4	13.3	17.5	16.8	15.4
Number							
Yes	302	2.0	2.3	2.3	5.0	5.3	5.0
No	568	8.5	11.6	9.5	18.1	17.6	16.9
Graduations							
Yes	582	4.8	5.7	5.3	11.0	10.8	11.2
No	288	9.0	13.9	10.4	18.8	18.4	16.0
Text							
Yes	140	9.3	10.0	9.3	22.9	20.7	22.9
No	730	5.6	8.1	6.6	11.8	11.9	10.8
Line							
Yes	719	7.5	10.0	8.5	16.4	16.1	15.4
No	151	0.0	0.7	0.0	0.0	0.0	0.0

* SU.VI.MAX, supplementation by antioxidant vitamins and minerals; VAS, visual analogue scale.

where μ is the actual latent average value of the group (μ unknown), s_i is the departure from the group mean of the subject i, and e_{ij} is the error due to VAS j for the subject i.

If a VAS measures the latent value without bias, the mean of the e_{ij} is 0, and the mean of the x_{ij} ($\bar{x}$) is μ. Bias is the systematic difference between the measured mean value ($\bar{x}$) and the actual latent average value (μ). As no VAS type can be considered a gold standard to measure subjective symptoms, this bias cannot be estimated in itself. The similar mean levels between scales suggest at the very least a bias of similar magnitude that might be null.

This study reveals an effect of orientation and of the types and the characteristics of VAS on the percentages of zero and small ratings. Graphic rating scales with text, scales without lines, and scales with graduations differ from other scales in that respect. The effect is particularly important to consider in light of the fact that studied symptoms are relatively rare in a presumably healthy adult population. The results might have been different had other subjective states been investigated, such as perceived health or pain among sick subjects. The most significant effect with regard to the measurement of change over time was related to the orientation of the scales. The relation between different characteristics and percentages of zero ratings was inverse to that between characteristics and ratings of one to nine; and then, the change in the percentage of ratings from zero to nine combined did not relate to orientation. Differences in the number of zero ratings have not been analyzed in previous studies, and means according to different types of scales have usually been compared without taking into account possible variations in floor effect.

Although subjects with no symptoms are assumed to rate with no error, subjects with weak symptoms are likely to wrongly rate zero, depending on the type and orientation of the scale. Thus, care is necessary when measuring symptoms of low or high intensity, that is, those close to either anchor term of the VAS. Consequently, the use of scales anchored with terms

TABLE 5. Effect of orientation, type, and characteristics of scales on distribution of low ratings (between 1 and 9), SU.VI.MAX* prevention trial pilot study, France, 1994

	p value for statistical significance of the tests† (*n* = 870)					
	Cross-sectional analysis			Longitudinal analysis		
	Blocked nose	Runny nose	Sore throat	Blocked nose	Runny nose	Sore throat
Orientation	‡			$<10^{-5}$	0.0061	$<10^{-5}$
Type	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	0.04		0.02
Interaction orientation x type	0.0015	$<10^{-5}$	0.0001			0.03
Characteristics§						
Middle mark	0.05	0.0016	0.007			
Numbers	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	0.01		0.06
Graduations	0.0001	$<10^{-5}$	0.0025			
Text	0.0004	0.02	0.0007	0.03	0.06	0.005
Line	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	0.008	0.06	0.03
Interaction of each characteristic with orientation						
Middle mark	0.004	0.0016	0.0014			0.05
Numbers		0.07		0.05		0.02
Graduations	0.0001	$<10^{-5}$	$<10^{-5}$			
Text	0.1	0.1	0.02	0.02		0.1
Line				0.044		0.05

* SU.VI.MAX, supplementation by antioxidant vitamins and minerals.
† Analysis of variance-covariance for repeated measures, with adjustment for chronic disease and current infection.
‡ Blanks indicate a *p* value > 0.1.
§ Adjustment for orientation.

TABLE 6. Mean intensity (mm) of symptoms according to orientation and type of scales (restricted to ratings between 10 and 90), SU.VI.MAX* prevention trial pilot study, France, 1994

	No.	Time 1			Time 2		
		Blocked nose	Runny nose	Sore throat	Blocked nose	Runny nose	Sore throat
All	870	35.9 (21.7)†	32.0 (22.7)	32.8 (20.8)	29.9 (19.2)	27.7 (18.7)	27.4 (19.5)
Orientation							
Horizontal	430	37.0 (21.1)	31.4 (21.8)	33.4 (21.2)	32.8 (20.1)	27.0 (17.8)	27.5 (20.3)
Vertical	440	34.9 (22.2)	32.7 (23.7)	32.3 (20.6)	27.3 (18.1)	28.4 (20.5)	27.3 (18.8)
Type							
Simple VAS*	143	36.8 (21.2)	36.4 (25.2)	32.7 (14.5)	34.6 (22.1)	27.7 (19.6)	29.1 (27.7)
Middle-marked VAS	145	38.8 (20.7)	30.1 (15.7)	31.8 (22.4)	36.8 (24.2)	32.5 (18.3)	25.0 (14.2)
Graphic rating scale	140	37.6 (23.5)	32.1 (22.5)	35.9 (21.5)	30.8 (18.9)	26.8 (18.3)	28.6 (18.0)
Graduated VAS	140	40.1 (25.1)	37.2 (28.7)	34.8 (22.9)	26.0 (14.6)	29.3 (18.9)	21.4 (17.4)
Graduated-numbered VAS	151	29.1 (19.1)	25.2 (19.5)	26.0 (16.5)	26.9 (16.6)	23.9 (18.3)	30.9 (18.5)
Numerical rating scale	151	35.0 (19.4)	32.3 (20.8)	34.2 (24.3)	28.1 (19.2)	28.6 (19.0)	28.4 (23.4)

* SU.VI.MAX, supplementation by antioxidant vitamins and minerals; VAS, visual analogue scale.
† Numbers in parentheses, standard deviation.

or labels likely to induce floor or ceiling effects, or the investigation of populations with a high proportion of the highest or lowest level of a phenomenon, would be expected to result in increased measurement error.

Scale characteristics do not affect mean levels of measurements overall and therefore do not seem to introduce any differential bias or bias of unequal magnitude. However, higher scores on vertical rather than horizontal scales cannot, as previously reported in the literature

TABLE 7. Precision of time 1 (T1) rating and of rating change (T1 – T2) according to orientation, type, and characteristics of scales, SU.VI.MAX* prevention trial pilot study, France, 1994

	p value by Levene's test for comparison of variances†					
	Time 1			Time 1 – Time 2		
	Blocked nose	Runny nose	Sore throat	Blocked nose	Runny nose	Sore throat
Orientation	‡					
Type	0.03	0.002	0.07		0.04	
Characteristics						
Middle mark			0.02			
Numbers	0.007	0.04		0.09	0.10	
Graduations			0.08			
Text				0.05		
Line			0.04			0.10

* SU.VI.MAX, supplementation by antioxidant vitamins and minerals.
† *p* values over 0.10 not shown.
‡ Blanks indicate a *p* value > 0.1.

(2, 5, 7), be ignored. It might have been detected using a symptom rating with less skewed distribution.

The precision of a measure is estimated by the variance of x_{ij}, which is equal to the sum of the variances of the s_i and the e_{ij} (1). As the different types of scales were randomly assigned, it is assumed that the variances of the s_i are equal in each group and that differences in precision between scales can be estimated by comparing variances of x_{ij}. Some differences in precision were observed on cross-sectional measurements, whereas in longitudinal analysis precision appeared to be consistent, whatever the characteristics of the scales. This conclusion is of importance for clinical trials, where the primary interest is in the detection of change over time. The power of a trial should not be influenced by the type of scale used.

In conclusion, the characteristics of VAS seem to be important in cross-sectional studies, particularly when measuring symptoms of low intensity (floor effect) or high intensity (ceiling effect). Orientation appears to have a critical influence on measurement and requires further investigation. If results of different studies are to be comparable, researchers should try to reach a consensus on the types of scale to be used. Differences between mean scale levels illustrate differences in measurement bias and are a strong argument for choosing a unique type of tool as a standard for use in all studies.

ACKNOWLEDGMENTS

This work was supported by a scholarship grant (no. 94-5-2497) from Ministère de l'Enseignement Supérieur et de la Recherche.

REFERENCES

1. Huskisson EC. Measurement of pain. Lancet 1974;2:1127–31.
2. Miller MD, Ferris DG. Measurement of subjective phenomena in primary care research: the visual analogue scale. Fam Pract Res J 1993;13:15–24.
3. Keele KD. Pain chart. Lancet 1948;2:6–8.
4. Scott J, Huskisson EC. Graphic representation of pain. Pain 1976;2:175–84.
5. Sriwatanakul K, Kelvie W, Lasagna L, et al. Studies with different types of visual analog scales for measurement of pain. Clin Pharmacol Ther 1983;34:234–9.
6. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. New York: Oxford University Press, 1989.
7. Scott J, Huskisson EC. Vertical or horizontal visual analogue scales. Ann Rheum Dis 1979;38:560.
8. Aun C, Lam YM, Collett B. Evaluation of the use of visual analogue scale in Chinese patients. Pain 1986;25:215–21.
9. Dixon JS, Bird HA. Reproducibility along a 10 cm vertical visual analogue scale. Ann Rheum Dis 1981;40:87–9.
10. Joos E, Peretz A, Beguin S, et al. Reliability and reproducibility of visual analogue scale and numeric rating scale for therapeutic evaluation of pain in rheumatic patients. J Rheumatol 1991;18:1269–70.
11. Wilkie D, Lovejoy N, Dodd M, et al. Cancer pain intensity measurement: concurrent validity of three tools—finger dynamometer, pain intensity number scale, visual analogue scale. Hosp J 1990;6:1–13.
12. Downie WW, Leatham PA, Rhind VM, et al. Studies with pain rating scale. Ann Rheum Dis 1978;37:378–81.
13. Littman GS, Walker BR, Schneider BE. Reassessment of verbal and visual analog ratings in analgesic studies. Clin Pharmacol Ther 1985;38:16–23.
14. Ferraz MB, Quaresma MR, Aquino LRL, et al. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022–4.
15. Hyland ME, Sodergren SC. Development of a new type of global quality of life scale, and comparison of performance and preference for 12 global scales. Qual Life Res 1996;5:469–80.
16. Cline ME, Herman J, Shaw ER, et al. Standardization of the visual analogue scale. Nursing Res 1992;41:378–80.
17. Hercberg S, Preziosi P, Briançon S, et al. A primary prevention trial using nutritional doses of antioxidant vitamins and

minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study—design, methods, and participant characteristics. SUpplementation en VItamines et Mineraux AntioXydants. Control Clin Trials 1998;19: 336–51.
18. Maxwell C. Sensitivity and accuracy of the visual analogue scale: a psycho-physical classroom experiment. Br J Clin Pharmacol 1978;6:15–24.
19. Dixon WJ. BMDP statistical software manual. Berkeley, CA: University of California Press and John Wiley & Sons, Inc, 1992.
20. Grunberg SM, Groshen S, Steingass S, et al. Comparison of conditional quality of life terminology and visual analogue scale measurements. Qual Life Res 1996;5:65–72.