LETTER TO THE EDITOR

COMMENTS ON "MEASURES OF INDIVIDUAL AND GROUP CHANGES IN ORDERED CATEGORICAL DATA"

While Sonn & Svensson (1) are to be congratulated on employing an interesting statistical technique to aid in the analysis of ordinal data, questions arise from their paper. First, we do not know why ADL was measured with a scale that resulted in a 3.82% loss of subjects from the analysis because of misfit, and we wonder if use of such a scale in epidemiological studies is justified when other scales without this problem exist. Other work by Sonn & Hulter-Åsberg (2) refers to this issue of misfit with reference to Katz et al. (3) as a form of justification. However, we could find no rationale in this reference that justifies the use of a scale with misfit properties.

Secondly, we consider the transformation to a dichotomous measure of the three-category classification of independent, partly dependent and dependent to be confusing, as it results in independence being defined as "capable of performing a task either independently or with assistance", while dependence is defined as "not capable of performing a task" or "capable of performing a task with assistance". In other words, the dichotomous classification introduces ambiguity by failing to form mutually exclusive categories.

Thirdly, we do not know from the paper how to interpret the Relative Position (RP) statistic. Its value will be dependent on differences between the two samples, as well as on the number of categories and the sample size. At what point does one say that the distributions are significantly different from each other, and can one calculate this level of significance?

REFERENCES

1. Sonn, U. & Svensson, E.: Measures of individual and group changes in ordered categorical data: application to the ADL staircase. Scand J Rehab Med *29:* 233–242, 1997.
2. Sonn, U. & Hulter Åsberg, K.: Assessment of activities of daily living in the elderly. A study of a population of 76-year-olds in Gothenburg, Sweden. Scand J Rehab Med *23:* 193–202, 1991.
3. Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A. & Jaffe, M. W.: Studies of illness in the aged. The Index of ADL: a standardised measure of biological and psychosocial function. J Am Med Assoc *185:* 914–919, 1963.

Accepted April 15, 1998

H. G. Dickson and F. Köhler
Liverpool Health Service, P.O. Box 103, Liverpool NSW 2170, Australia

Response to the Letter by Dickson & Köhler

Sir,

We appreciate the great interest in our work, as there are many issues that can be discussed regarding measurement scales. Recently, Hand (5) and Cox (2) also pointed to the fact that little attention is paid to the relationship between measurement theory and assessments, although this relationship is fundamental to the choice of statistical approach to data. Your questions offer an opportunity to discuss these issues further.

It is true that there is a risk of acquiring misfits when using hierarchical conditional scales, that is, assessment not following the conditional structure defined according to a manual. This particular problem does not exist in summation scales, but these pose other problems, as discussed in our paper (3, 8, ref. 1 above). Concerning the ADL Staircase, it is important to conclude that assessments categorized as "Others" (misfit) are not missing values, as each person can be described in each item of the scale (ref. 2 above). In applying the scale to comparative evaluations, the category "Other" does not have to be excluded. By definition, a person classified as "Other" is more dependent than one classified as ADL-step 0 or ADL-step 1, and more independent than one classified as ADL-step 9/10. Furthermore, in each assessment one can always determine improvement or deterioration when changes occur in that person (7, ref. 2 above). Of course, the proportion "Others" should be as small as possible, otherwise it might indicate that the wrong measurements have been chosen for the study group. As the aim of the present study was to demonstrate a new statistical approach for analysis of change in ordered categorical scales, detailed analyses of

"Others" were not included. However, those 11 persons out of 382 (3%) who were classified as "Others" are presented in Table I. The subjects could also be further described in detail. The misfit of classifications occurs only on a single occasion.

According to the ADL Staircase, individual performance can be described not only in ADL steps 0–9/10, but the recording can also be used in other ways, e.g.:

- overall independence or dependence (ADL step 0/ADL steps 1–9/10),
- or in three areas of independence, dependence only in I-ADL, dependence in I-ADL as well as in P-ADL,
- or as a separate description of each activity (ref. 2 above, 9).

In these examples, the need for the category "Other" is eliminated.

The measurement process in instrument development includes both theoretical definitions of the concept to be measured and operational definitions of the attributes to be assessed. How items are defined varies greatly, as does how they are operationalized in different scales/instruments. Certainly this will affect both the reliability and validity of a scale/instrument. This is a general problem with all scales, hierarchical or not.

By definition, an activity consists of a group or cluster of tasks or actions which are parts of or components of the activity. The activity assessed may be defined in one word, such as dressing or cleaning, or, as in this case, with a more detailed description of what is meant by e.g. dressing (getting all needed clothing from closets and drawers and getting dressed, including fasteners, and putting on a brace, if worn) or cleaning (housecleaning, vacuuming, washing floors, etc.). In the ADL Staircase, each activity is defined and operationalized. The level "Partly dependent" is used to facilitate the assessment, and should be looked upon as a guideline to make it easier to determine whether a person can be classified as dependent or independent. It is a way of sharpening the definition so as to avoid ambiguity in the categories "dependent" or "independent". For instance, in dressing and cleaning, "Partly dependent" means needing very little help (tying shoes or getting assistance very seldom, e.g. to take carpets outdoors). Accordingly, these persons are assessed as independent, as they get very little help. In other activities such as shopping, persons who perform the activity only together with another person are assessed as dependent.

Our paper (ref. 1 above) demonstrates a statistical approach to a comprehensive evaluation of change ordered categorical data. The method is developed rating scales, and it takes account of the non-met properties of the data. The method makes it possible identify and measure the level of change in AD attributed to the group separately from the level individual variability within the group. The level ordered-preserved change in common for the group measured by the Relative Position (RP), and the level individual change is measured by the Relative Ra Variance (RV).

A systematic group change towards higher levels ADL dependence on the second occasion, for examp implies a non-zero, positive RP value. Individu changes apart from the group change imply a positi RV value (see ref. 1 above p. 235). Both RP and RV a rank-invariant measures, which means that they rema unaffected by re-labelling of categories and also, to so extent, by the number of categories. Scales with a sm number of categories are less sensitive to changes, whi implies loss of measurable information (1, 4). Thus, common with other statistical measures, RP and R depend on the responsiveness of scales.

Both measures are independent of sample size. Th standard error (SE) of a measure is, by definitio dependent on sample size. The more numerous th observations, the higher the confidence in the conclu sions drawn from the measures. Recent simulatio studies show that the measures are approximatel normally distributed. Hence, approximate confidenc intervals (CI) and hypothesis tests can be performed.

A 95% CI for the Relative Position of the populatio can be estimated by RP 1.96 SE(RP), and, for th individual level of change, by RV 1.96 SE(RV). A te statistic of the null hypothesis of unchanged ADL leve for the group over time is $z = RP/SE(RP)$, where z i normally distributed. Corresponding expressions hol for RV.

In our paper (page 237), the level of systematic grou change in ADL dependence for subjects aged 70–7

Table I. *Subjects classified as "Other" (x) according t the ADL Staircase at ages 70, 73 and 76 years*

	70	73	76	f
ADL step	0	x	5	7
	1	x	5	1
	3	6	x	1
	3	x	5	1

Table II. *The measures of change in ADL level and the differences in the measures between the groups men (M) (n = 166) and women (W) (n = 205); the jackknife standard errors for the measures are given in parentheses*

	Men	Women	Differences between men and women	95% CI of differences
Systematic change in position for group RP (SE)			RP_M–RP_W	RP_M–RP_W
70–73 years of age	0.092 (0.025)	0.035 (0.018)	0.057 (0.015)	0.027 to 0.087
73–76 years of age	0.221 (0.036)	0.178 (0.028)	0.043 (0.022)	0 to 0.086
Individual changes RV (SE)			RV_M–RV_W	RV_M–RV_W
70–73 years of age	0.0018 (0.0012)	0.0017 (0.0009)	0.0001 (0.0007)	−0.0013 to 0.0015
73–76 years of age	0.020 (0.008)	0.012 (0.005)	0.008 (0.0045)	−0.001 to 0.017

years was RP = 0.257 (SE = 0.024). The 95% CI of the estimated true level of group change for the elderly population is 0.21 to 0.30 (0.257 1.96 SE(RP)), and this interval differs significantly from zero, i.e. from unchanged ADL status (p 0.0001 as z = 10.7). As shown in our paper (page 238), the level of individual dispersion from the common pattern of group change in ADL dependence during the same time period was small: RV = 0.0112 (SE(RV) = 0.0036). However, the 95% CI from 0.004 to 0.018 confirms the slight departure from zero ($p = 0.002$, z = 3.11), which indicates the presence of some individual changes apart from the common ordered preserved group change.

The present study is a one-sample step-wise evaluation of changes in ADL dependence among elderly persons between 70, 73 and 76 years of age. By means of the statistical approach, it was possible to calculate the contribution to the change from the subsamples of men and women. The value of RP of the total group is the weighted mean of the RP values of the subgroups, denoted A and B. Thus, $RP = (n_A RP_A + n_B RP_B)/(n_A + n_B)$, where n_A and n_B denote the size of the subgroups A and B (see notations in our paper). The systematic change in ADL dependence for the whole group between 70–76 years of age is RP = 0.26, and the subgroup of men contributed more than the subgroup of women to this change, as RP was 0.30 in men and 0.22 in women (see Table III, page 240). By this one-sample design with gender as an explanatory variable, it is possible to test the hypothesis of no group and individual changes over time for the whole group. The 95% CIs for the level of group changes, measured by RP, for subjects 70–73 and 73–76 years of age were 0.03 to 0.09 (p 0.0001) and 0.15 to 0.24 (p 0.0001), respectively. The 95% CIs for the level of individual changes, measured by RV, for subjects 70–73 and 73–76 years of age were 0.0004 to 0.004 ($p = 0.01$) and 0.007 to 0.02 ($p = 0.0001$), respectively. The p-values are adjusted for multiple tests by means of the Bonferroni-Holm sequential multiple test procedure (6).

The observed gender difference between the patterns of change with increasing age, and the contributions to the measures of RP and RV might result in the formulation of a new hypothesis of differences in the levels of individual and group changes between the two groups of men and women. The measures of RP and RV are separately calculated for each group. Table II shows the measures of RP and RV and the 95% CI for the difference in RP between the groups of men (M) and women (W). The 95% CI for the difference between the groups is calculated by $(RP_M - RP_W)$ 1.96 $SE(RP_M - RP_W)$, where SE is based on the weighted standard deviation of the groups.

According to the two-sample comparison, there is a significant difference ($p = 0.0001$) in the group change in ADL dependence between men and women aged 70–73 years, as the 95% CI of the difference in RP ranges from 0.027 to 0.087.

Furthermore, note the printing errors on page 238. The figures should be $3 \times 95^2 + 1 \times 91^2$.

REFERENCES

1. Cox, D. R., Fizpatric, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J. & Jones, D. R.: Quality-of-life assessment: can we keep it simple? J R Stat Soc A *155:* 353–393, 1992.
2. Cox, D. R.: The current position of statistics: a personal view. Inter Stat Rev *65(3):* 261–290, 1997.
3. Feinstein, A. R., Joseph, B. R. & Wells, C. K.: Scientific and clinical problems in indexes of functional disability. Ann Intern Med *105:* 413–420, 1986.
4. Guyatt, G. H., Deyo, R. A., Charlson, M., Levine, M. N. & Mitchell, A.: Responsiveness and validity in health status

measurement: a clarification. J Clin Epidemiol *42:* 403–408, 1989.
5. Hand, D. J.: Statistics and the theory of measurement. J R Stat Soc A *159:* 445–492, 1996.
6. Holm, S.: A simple sequentially rejective multiple test procedure. Scand J Stat *6:* 65–70, 1979.
7. Katz, S. & Akpom, A.: A measure of primary sociobiological functions. Int J Health Serv *6:* 493–507, 1976.
8. Merbitz, C., Morris, J. & Grip, J. C.: Ordinal scales and foundation of misinference. Arch Phys Med Rehab *70:* 321, 1989.
9. Sonn, U.: Longitudinal studies of dependence in daily life activities among elderly persons. Scand J Rehab Med *Suppl. 34:* 1996.

Ulla Sonn, OT, PhD and Elisabeth Svensson, PhD, Biostatistician