Scales in the ALSWH

ALSWH surveys contain numerous health measurement scales, most of which have been validated in published work. Prior to using these scales in analyses it has been the practice in the study to evaluate the statistical properties of these scales in the survey in which they were first included for each age cohort. The evaluation includes implementation of the scoring recommended in the scale's validation and an assessment of its utility in the ALSWH data. The main objective is to endorse the use of the scale as recommended. Where this is not supported by the analysis, an alternative is proposed. Users of the ALSWH datasets are strongly advised to read relevant sections before using scales in statistical analysis. In particular, the use of a single item from a scale rather than the scale score is rarely appropriate.

The following procedure has been developed by the Data Management Committee as a guide to scale evaluation.

Scale Evaluation Procedure

Background
Summarise the concepts involved in the measure/scale; summarise the validation study for the measure/scale including study sample, text of items, response format and scoring, results of factor analyses, scoring of derived scales.

Item Responses
Describe the distribution of item responses (percent for each response, mean and standard deviation, number and percent missing) in ALSWH data. Describe the percent of missing items across complete scale and/or sub-scales.

Scale Reliability
Estimate scale reliability with item-to-total correlations (correlation of an individual item with the item total, excluding that item) and Cronbach's alpha coefficient (a measure of internal consistency). Streiner and Norman[1] suggest that alpha should exceed 0.7, but be no higher than 0.9 as very high values of alpha may indicate redundancy of some items. Others[2] recommend a minimum alpha coefficient of 0.5 for scales to be used in comparing groups of consumers. The ALSWH has set 0.6 as a minimum measure of internal consistency and requires item-to-total correlations to exceed 0.5. Examination of scale stability is beyond the scope of the ALWSH which administers surveys to each cohort once every 3 years.

Factor Analysis
Determine the number of factors supported by the data based on the consensus from the eigenvalues-greater-than-one rule, parallel analysis[3] and Velicer's MAP test.[3] Parallel analysis is based on simulations and recommends retaining the number of factors with observed eigenvalues greater than the corresponding 95th percentile of simulated values. The MAP test recommends the number of factors associated with the minimum average squared correlation.

Although parallel analyses and the MAP test usually suggest the same number of factors, this will not always be the case. When there are differences, O'Connor[3] recommends increasing the number of simulations used for the parallel analysis and scrutinising the average squared correlations for close calls. These two procedures tend to err (when they do) in opposite directions: parallel analysis tends to over-extract and the MAP test tends to under-extract factors.

Use factor analysis to determine whether the ALSWH data support the hypothesized groupings (if any) of the items. Where there is more than one factor, use the rotated factor pattern from principal components analysis with varimax rotation to check the groupings of the items and the contribution of each item to the factors. Other rotation methods should be considered when the factor pattern remains unclear. Items can be considered to load onto a particular factor if loadings exceed 0.5 and there is little cross-loading onto other factors.

Variables with communality (the amount of variance each item shares with all other items) of less than 0.5 should be considered for deletion.[4a] However, any decision to delete variables with lower communalities should take into account the variable's overall contribution to the research[4b] and any history of the inclusion of the variable in other research using the scale in conjunction with the other factor analysis results.

In the absence of a clear conclusion about the coherence of factors, it may be necessary to assess criterion validity (also known as convergent or concurrent validity). This involves developing clear hypotheses stating which measures will be associated with the factor(s), and in which direction. Measures expected not to be associated with the factor should also be proposed. Associations with these factors are evaluated with Pearson correlation coefficients and with differences in unadjusted means. These other measures may be concurrent (measured at the same time) or predictive (measured at a time after the proposed factor).

Latent trait analysis
Common factor analysis is not applicable to dichotomous data.[5] The magnitude of correlation between items is affected by the endorsement proportions (percentage of sample who responded "yes") of dichotomous items[4] and complex factorial structures can result from the factor analysis of dichotomous data with differing endorsement proportions, even when items are from a one-dimensional scale.[6] Latent trait analysis is the method ALSWH has used in these circumstances. A full description is provided in Section 2 – Goldberg Anxiety and Depression Scale.

Scores
Evaluate scoring options for complete cases. Calculate factor scores as the sum of items in each scale/sub-scale weighted by the appropriate standardised factor scoring coefficient from rotated factor analyses; scoring coefficient from rotated analysis will not be available for a one-factor solution. Calculate summed scores as the unweighed sum of item scores (or the unweighted mean of item scores). Calculate any other composite score detailed in the validation paper for a scale.

Determine the preferred score. Compare the distributional properties (mean, standard deviation, median, skewness, coefficient of variation, range) of all scores. Calculate the correlation of scores. Hair suggests that where the summated scale is well-constructed, valid and reliable it is preferable to factor scores.[4c] Factor scores are not replicable across studies, as scoring coefficients must be estimated independently in each study population.[4d] In contrast un-weighted, sum-based scores are more readily compared with other populations[4c] and tend to overcome problems of measurement error.[4b] So summed scores will be preferred to the factor score unless particular defects are found.

Visually evaluate the distribution of the preferred score for normality.

Missing items
To overcome the problem of missing data in the derived variable, some missing items may be imputed. Generally the imputed value will be the mean of the non-missing items. The number of items for which ALSWH allow imputation is usually 1 in each 6 scale items or as recommend by the developer of the scale.

Categorisation
Evaluate categorisation of scores. Develop appropriate categories for non-normal scores and consider the utility of categories for normally distributed scores. The most meaningful categorisation may include recognised/recommended cut-points, quantiles or values that have a specific interpretation in terms of the scale items.

Recommendation for usage
Recommend usage as a continuous or categorical measure.

SAS code
Include the SAS code used to derive any scores or categories recommended.

References

1. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use,* 2nd Ed. Oxford (UK): Oxford University Press; 1995. Page 65
2. Davies AR, Ware JEJ. *GHAA's consumer satisfaction survey and user's manual*, 2nd Ed. Washington, DC: The Group Health Association of America; 1991. Page 39
3. O'Connor.BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behaviour Research Methods, Instruments & Computers* 2000;32(3):396-402
4. Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate Data Analysis*. 5th Ed. Prentice Hall; 1998. a Page 115; b Page 116; c Page 120; d Page 119
5. Clarke D, Mackinnon A, Smith G, MacKenzie D, Herrman H. Dimensions of psychopathology in the medically ill: A latent trait analysis. *Psychosomatics* 2000;41(5):418-425
6. Mackinnon A, Christensen H, Form A, Henderson A, Scott R, Korten A. A latent trait analysis of an inventory designed to detect symptoms of anxiety and depression using an elderly community sample. *Psychological Medicine* 1994;24:977-986