NIH Public Access
Author Manuscript
Stat Methods Med Res. Author manuscript; available in PMC 2013 December 01.

Published in final edited form as:
Stat Methods Med Res. 2012 December ; 21(6): 599–619. doi:10.1177/0962280210391444.

Scale Development Based on Likelihood Cross-Validation

George J. Knafl[1], **Jane K. Dixon**[2], **Jean P. O'Malley**[3], **Margaret Grey**[2], **Janet A. Deatrick**[4], **Agatha Gallo**[5], and **Kathleen A. Knafl**[1]

[1] School of Nursing, University of North Carolina at Chapel Hill

[2] School of Nursing, Yale University

[3] Clinical and Translational Research Institute, Oregon Health & Science University

[4] School of Nursing, University of Pennsylvania

[5] School of Nursing, University of Illinois at Chicago

Abstract

The use of likelihood cross-validation for guiding the scale development process is formulated and demonstrated, including choosing the number of factors, assessing item-factor allocations suggested by rotations, adjusting those allocations, reducing the number of factors, removing items, and assessing the applicability of scales to subjects other than those for whom it was originally developed. An example analysis is presented on the development of scales to measure how parents caring for a child with a chronic condition view their family's management of that condition.

Keywords

childhood chronic conditions; cross-validation; factor analysis; model selection; scale development; survey instruments

1 Introduction

Factor analysis is used in scale development to model the covariance for survey instrument item responses. A set of preliminary items are developed that address a certain theoretical domain. A cross-sectional study is conducted surveying a relatively large number of subjects who respond to the newly developed preliminary instrument as well as to related established survey instruments for construct validity purposes along with relevant subject characteristics. When the items are developed to address specific dimensions within the theoretical domain of the instrument, confirmatory factor analysis (CFA) can be used to assess how well those dimensions are addressed by such theory-based scales. However, there is no guarantee that this will result in an effective item-scale allocation so that an exploratory factor analysis (EFA) may still be required. EFA is also required when there is no a priori information on issues like the appropriate number of scales, the allocation of items to those scales, and which items to discard. Factors underlie EFA and CFA models, representing unmeasured, latent constructs and so EFA and CFA are special cases of classical latent variable models.[1]

Request for reprints and software should be sent to George J. Knafl, School of Nursing, Campus Box 7460, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7460. gknafl@unc.edu..

There is a need to evaluate alternative EFA/CFA models, and Knafl and Grey[2] developed a likelihood cross-validation (LCV) approach for these purposes. They demonstrated the usefulness of LCV in evaluating EFA/CFA models, but only for item responses to established survey instruments. LCV can also be used to assist in scale development, providing objective support for decision-making at various steps in the analysis, including choosing the number of factors, assessing item-factor allocations suggested by rotations, adjusting those allocations, reducing the number of factors, removing items, and assessing the applicability of developed scales to other types of subjects. While there are similarities between developing a new scale and assessing an established scale, scale development involves more complex issues. The items of established scales have already been screened for relevance and allocated to associated scales. Those scales also have meaningful interpretations as well as established reliability based on Cronbach's alpha and construct validity. On the other hand, items considered in scale development have to be allocated to new scales and substantial numbers are likely to be expendable, thereby complicating the development process. Moreover, those scales need to be shown to have meaningful interpretations and acceptable reliability and construct validity. Consequently, a novel approach for scale development guided by LCV is formulated and then demonstrated through example analyses addressing development of scales measuring family management of childhood chronic conditions. LCV provides objective criteria for judging different choices made as part of scale development, but is not sufficient by itself. Other issues, including conceptual meaning, reliability, and construct validity, need to be considered as well by all members of the research team, not just the statistician.

2 Model evaluation using likelihood cross-validation

2.1 Likelihood cross-validation (LCV)

Cross-validation is used sometimes in factor analysis, and more generally in structural equation modeling,[3,4] but usually the type known as learning-testing (or splitting or holdout). Data for subjects are randomly split into two disjoint subsets, the learning (or training or calibration) set for fitting the model and the test (or holdout or validation) set for evaluating that model's performance. Knafl and Grey[2] used a more extensive k-fold likelihood-based cross-validation scheme, with data for subjects randomly partitioned into $k>1$ disjoint subsets called folds. Using k-fold cross-validation[5,6] instead of learning-testing allows model estimation and evaluation to be based on all the data rather than on separate subsets of the data.

Let $\mathbf{y}_s$ denote the vector of I item responses for the sth subject out of n total subjects. Assume that these are realizations of a random vector $\mathbf{y}$ with mean vector μ and covariance matrix Σ sampled independently from the same population. Let $S=\{1,\cdots,n\}$ be the set of subject indexes and denote by μ(S) and Σ(S;UN) the maximum likelihood estimates of μ and Σ under a multivariate normal distribution with unstructured covariance matrix. In other words, μ(S) is the average over the indexes in S of the n observed response vectors $\mathbf{y}_s$ and Σ(S;UN) is the average of the n cross-product matrices $(\mathbf{y}_s-\mu(S))\cdot(\mathbf{y}_s-\mu(S))^T$. For simplicity of notation, parameter estimates are represented as functions of the data S used in their computation without hat symbols. For any other covariance structure model M, denote the corresponding maximum likelihood estimate of Σ by Σ(S;M). Maximum likelihood is used to estimate parameters in this formulation, even parameters like the item variances typically estimated instead with unbiased estimates based on restricted maximum likelihood.

Denote by $L(\mathbf{y};\mu,\Sigma)$ the likelihood for $\mathbf{y}$ when it is treated as multivariate normally distributed with mean vector μ and covariance matrix Σ. For any nonempty set S′ of indexes in S and any covariance structure model M, denote by μ(S′) and Σ(S′;M) the maximum likelihood estimates of μ and Σ computed with the response vectors $\mathbf{y}_s$ for indexes s in S′.

Let S_h for $1 \le h \le k$ denote k folds that partition S into nonempty subsets with nonempty complements $S_h^c = S \backslash S_h$. In the example analyses, subjects are randomly assigned to folds. Denote by $S_{h(s)}$ the unique fold containing the index s. The contribution to the LCV score for the sth subject is given by

$$\mathrm{LCV_s} = \mathrm{L}\left(\mathbf{y}_s; \mu\left(\mathrm{S_{h(s)}}^c\right), \Sigma\left(\mathrm{S_{h(s)}}^c; \mathrm{M}\right)\right)$$

or equivalently

$$\log\left(\mathrm{LCV_s}\right) = -\left(\mathbf{y}_s - \mu\left(\mathrm{S_{h(s)}}^c\right)\right)^{\mathrm{T}} \cdot \Sigma^{-1}\left(\mathrm{S_{h(s)}}^c; \mathrm{M}\right) \cdot \left(\mathbf{y}_s - \mu\left(\mathrm{S_{h(s)}}^c\right)\right)/2 - \log\left((2 \cdot \pi)^{\mathrm{I}} \cdot \left|\Sigma\left(\mathrm{S_{h(s)}}^c; \mathrm{M}\right)\right|\right)/2.$$

The LCV score is then given by

$$\mathrm{LCV} = \Pi_{\{1 \le s \le n\}} \mathrm{LCV_s}^{1/m}$$

where m=I·n is the total number of item responses for all subjects. Models with larger LCV scores are better models with better predictive capability and more compatible with the available data as measured using deleted predictions determined by the fold assignment.

LCV appears to have been first formulated by Stone,[7] but just for the leave-one-out (LOO) case with each observation in its own fold. LOO LCV has been used to choose the bandwidth parameter for kernel estimation,[8,9] for growth curve modeling,[10] and for variable selection.[11] A repeated form of learning-testing LCV has been used for cluster analysis[12] while k-fold LCV has been used for adaptive regression modeling[13,14] and for factor analysis.[2]

2.2 LCV ratio tests

Since LCV scores are based on likelihoods, they can be compared using LCV ratio tests analogous to likelihood ratio tests. Stone[7] originally proposed and justified these tests for the LOO case. Knafl, Fennie, & O'Malley[14] applied them with k-fold LCV in the context of adaptive repeated measures modeling. The unnormalized LCV score, that is, LCV^m, equals the standard likelihood if parameter estimates are based on data for all the subjects combined rather than on data for subjects within complements of folds. As long as the fold sizes n_h are small in comparison to the number n of subjects and n is large, deleted estimates $\mu(S_h^c)$ and $\Sigma(S_h^c;M)$ are based on data for large numbers $n-n_h$ of subjects in the complements S_h^c of the folds S_h. Consequently, standard asymptotic results are reasonably applied to deleted estimates $\mu(S_h^c)$ and $\Sigma(S_h^c;M)$ of μ and Σ so that −2 times the log of ratios of unnormalized LCV scores are asymptotically χ^2 distributed,[7] as for ratios of standard likelihoods. When applied to nested EFA models with different numbers of factors, the degrees of freedom parameter equals the change in the number of parameters, i.e., the number I of items times the decrease in the number of factors, because the removal of a factor reduces the number of loading parameters by 1 for each item.

2.3 Factor analysis models

Likelihoods for factor analysis models are based on the multivariate normal distribution with factor analytic covariance structure. Items are often standardized which will be assumed in this formulation, but factor analysis based on raw items can be addressed similarly. Specifically, let $\mathbf{z}_s = \mathbf{V}^{-½} \cdot (\mathbf{y}_s - \mu)$ be the I×1 vector of standardized items where $\mathbf{V}$ is the I×I diagonal matrix of variances for the I items. These are modeled in EFA as $\mathbf{z}_s = \Lambda \cdot \eta + \mathbf{u}_s$ where

η is the F×1 vector of F latent common factors (or just factors for short) with zero mean vector and independent with unit variances, Λ is the I×F matrix of loadings $\lambda_{i,f}$ of the I items on the F common factors (and so all items are assumed to load on all factors), and $\mathbf{u}_s$ is the I×1 vector of unique factors with zero mean vector, I×I diagonal covariance matrix Ψ of variances for the I unique factors (and so they are independent of each other), and independent of η. Consequently, the covariance matrix for $\mathbf{y}_s$ is $\Sigma=\mathbf{V}^{½}\cdot(\Lambda\cdot\Lambda^T+\Psi)\cdot\mathbf{V}^{½}$ where the correlation matrix $\Lambda\cdot\Lambda^T+\Psi$ for $\mathbf{y}_s$ or the covariance matrix for $\mathbf{z}_s$ is constrained to have unit diagonal entries.

CFA models also are considered in example analyses with each item loading on at most one factor thereby constraining the loading matrices Λ. Loadings for these CFA models are sometimes pre-specified and sometimes estimated. While estimated loadings generate better LCV scores, pre-specified loadings are used when that better reflects the issue under investigation. These models also allow for the more general case of correlated common factors rather than assuming them to be indepedent as in EFA models.

The EFA formulation is commonly just used for factor extraction, but LCV is used to support a variety of other aspects of the analysis as well. LCV scores can be computed for alternative EFA models to assess the impact of different numbers of factors and different factor extraction procedures and for CFA models corresponding to scales suggested by rotated EFA loadings to assess the impact of different rotations and of different specifications for inter-factor dependence.

Models for the example analyses of Section 4 were computed in SAS Version 9.2 (SAS Institute, Inc., Cary, NC) using the EFA procedure PROC FACTOR (with the HEYWOOD option to set communalities greater than 1 to 1), the CFA procedure PROC CALIS, and the matrix language procedure PROC IML. SAS macros for conducting the computations are available at http://www.unc.edu/~gknafl/software.html (accessed 10/30/10) or by request from the first author. PROC FACTOR and PROC CALIS discard data for subjects with missing item values, and so these missing values need to be imputed to avoid loss of data. Reliability and construct validity use person mean substitution since they are based on specific item-scale allocations while EFA and CFA computations use item mean substitution since item-scale allocations have not yet been determined.[15] More sophisticated imputation alternatives could have been used instead, e.g., corrected mean substitution[15] or the NORM approach of Schaefer[16] for handling multivariate normal data. Alternately, full information maximum likelihood[17] can be used instead to compute parameter estimates without the need for imputation, but the scale development analysis issues addressed here apply in any case.

2.4 Accounting for removed items

In order to assess whether or not items are reasonably removed or not, LCV scores need to be based on all I available items so that they are comparable. If models are based on different numbers of items, LCV scores could change because the number of items has changed, not necessarily because of adjustments to the model. All I items are considered by setting all the factor loadings of removed items to zero, that is, $\lambda_{i,f}(\mathrm{CFA})=0$, $1\le f\le F$, for each removed item i. Removed items also are treated as having unstructured covariance. Treating them as independent generates worse LCV scores and seems unrealistic. To accomplish this, the covariance matrix Ψ for the unique factors is partitioned into two submatrices (and is left zero elsewhere): a diagonal submatrix Ψ_1 of unique factor variances for the non-removed items so that these remain independent of each other and a submatrix Ψ_2 of unstructured correlations (not more general covariances due to the constraints on $\Lambda\cdot\Lambda^T+\Psi$) for the removed items.

2.5 Choosing numbers of folds and factors

The number F of factors can be chosen by maximizing the LCV score in F. The results of Knafl and Grey[2] suggest that this selection will be robust to the number k of folds as long as k is not too small. Consequently, they recommend setting the number of folds to the first local maximum in k for $\max_F\{LCV(k,F)\}$ over scores LCV(k,F) for EFA models with F factors using k folds. This choice addresses the need for k to be not too small while limiting the computations. Subsequent analyses would use this value of k as well. Their results also suggest that smaller numbers of factors than the one with maximal LCV score may be parsimonious alternatives in the sense that they generate competitive LCV scores using fewer factors.

2.6 Rotating factors

Rotations R of the loadings of factors for an EFA model can be assessed using the approach of Knafl and Grey[2] based on CFA models for the item-factor allocations they suggest, with items allocated to only one factor, the one on which they load most highly in the rotated EFA solution. In other words, item i is allocated to factor f(i) satisfying

$$|\lambda_{i,f(i)}(\mathrm{EFA},\mathrm{R})| = \max_f \{|\lambda_{i,f}(\mathrm{EFA},\mathrm{R})| : 1 \le f \le F\}$$

where $\lambda_{i,f}$ (EFA,R) denote the rotated loadings generated by R applied to the EFA loadings. The CFA models considered for assessing rotations use unit loadings, that is, ±1 depending on signs for their strongest loadings in the EFA model, and 0 loadings on all other factors. In other words,

$$\lambda_{i,f(i)}(\mathrm{CFA},\mathrm{R}) = \mathrm{sign}(\lambda_{i,f(i)}(\mathrm{EFA},\mathrm{R})) \quad \text{and} \quad \lambda_{i,f}(\mathrm{CFA},\mathrm{R}) = 0 \quad \text{for} \quad f \ne f(i).$$

Items are allocated to single factors so that the CFA models correspond to the common practice of computing scales by summing disjoint sets of items. Unit loadings are used since then an observed item value equals the sum of a latent true score plus error as in classical test theory. More general models can be considered with estimated loadings and/or items loading on more than one factor. All common factors in these and other CFA models are treated as correlated.

2.7 Adjusting item-factor allocations

The use of unit loadings requires specification of codings for items, that is, which items are negatively oriented and so need reverse coding and which are positively oriented and so do not. Codings can be determined for scales suggested by rotations using signs of rotated loadings for EFA models. However, if an item is reallocated to another factor for some item-factor allocation, its coding may need to be changed. Having to make this decision can be avoided by using CFA models with estimated rather than unit loadings. Appropriate codings for associated unweighted scales (i.e., scales computed as sums of items with unit weights) are then determined by signs of the estimated loadings. Estimated loadings also provide measures of the strength of relationships between items and the factors to which they are allocated. For these reasons, CFA models with estimated loadings are used in identifying an effective item-factor allocation on which to base unweighted scales. As before, each item i is allocated to only one factor f(i) and so has only one nonzero loading $\lambda_{i,f(i)}$(CFA).

An improved item-factor allocation can be identified using the following systematic search. At each step of the allocation adjustment process, reallocate items one at a time to each of the other factors as well as to no factors (i.e., by setting all the item's loadings to zero and so removing the item; see Section 2.3). This requires computation of LCV scores for F different models per item (corresponding to reallocation of the item to the F–1 factors other than its currently allocated factor as well as to none of the factors). The computational effort can be limited by considering for reallocation only a specific number J of the I items, the ones with the weakest standardized loadings (i.e., smallest in absolute value) in the CFA model for the current allocation. Each adjustment step still requires substantial amounts of computation, involving comparison of J·F different models to select the next reallocation, but this can be far less than the I·F models needed for an exhaustive search (e.g., J=10 is used in example analyses out of I=65 items with F=10 factors). While there is no guarantee that this produces the same results as an exhaustive search, the results should not be too different. Items with weak loadings are the ones whose reallocation would likely improve the model. The adjustment process continues as long as the LCV score improves.

2.8 Combining factors

An item-factor allocation may be based on too many factors. Evidence of this includes reliability for associated scales below the conventional acceptable cutoff of 0.70[18,19] or factors with small numbers of associated items (e.g., three or less). One way to reduce the number of factors is to combine pairs of factors, allocating items for the two factors to a single combination factor. However, only problematic factors need be considered for combination. LCV can be used to support the factor combination process. At each step, the two factors to be combined next are the ones whose combination generates the best LCV score of all pairs of factors under consideration at that step. Since LCV scores typically decrease with factor combination, the process is continued as long as the next combination generates a scale with improved reliability. Other criteria for combining factors could be considered like combining factors on the basis of correlations between associated scales, but the use of LCV seems more objective.

2.9 Removing factors and reassigning their items

Factor combination may not resolve all problems, e.g., combined factors may still have unacceptable reliability. This can be resolved by removing problematic factors, but the reassignment of some of their items to remaining factors might improve those factors. Item reassignment can be based on LCV, computing scores for CFA models with each removed item allocated to each remaining factor. The best of these LCV scores then determines the next item-factor reassignment. Continue as long as this generates an improved LCV score.

2.10 Removing items from scales

So far, items might be removed due to item-factor adjustments or removal of factors to which they are allocated. Other items may be expendable as well. This is addressed by systematically removing currently allocated items one at a time. The next item removed is the one whose removal generates the CFA model with the best LCV score as long as that score improves. CFA models with unit rather than estimated loadings are used to address the impact of item removal on unweighted scales. As before, removed items have all zero loadings and unstructured covariance.

Item elimination is commonly based on strength of loadings. Items are assigned in disjoint sets to the factors for which they have strong loadings (i.e., loadings large in absolute value) in rotated EFA solutions. Conventional cutoffs for a strong loading typically range from 0.3 to 0.5. Items with strong loadings on more than one factor are usually eliminated because they do not reflect a unique latent construct.[20] Items that do not load strongly on any of the factors provide little information about those factors and so would be eliminated as well.[21] These conventional item elimination approaches can be assessed through LCV scores for associated CFA models with unit loadings, estimated inter-factor correlations, and eliminated items removed as before by setting their loadings on all factors to zero and treating them as having unstructured covariance.

2.11 Consultation with the research team

Preliminary scales are generated at each of the steps of the scale development process that can be reviewed by research team members to assess those scales and to decide whether further scale development is needed. The current item-scale allocation with items ordered on absolute values of their standardized loadings (e.g., Table 7) can be used to assess the conceptual meaning and interpretability of associated scales. Counts of the number of items allocated to factors as well as reliability and construct validity coefficients for the scales should also be reported in order for the research team to decide if these are acceptable or not. Furthermore, specific adjustments within each of the scale development steps (e.g., which items are reallocated to which factors, which factors are combined, which factors are removed, which items are reassigned to which factors, and which items are removed) can also be reviewed for appropriateness by the research team.

2.12 Preliminary item assessment

Items are often eliminated prior to factor analysis on the basis of their summary statistics. Items with means close to the center of the range of item values and with relatively high standard deviations are desirable.[18] Inspection of item correlations is also recommended to assess subjectively whether relationships between items are sufficiently strong for identified factors to correspond to actual latent constructs.[19] However, all available items are used in example analyses. Decisions about eliminating items are based as described above on LCV scores.

2.13 Applicability of developed scales to other groups of subjects

Applicability of scales developed using responses for one group of subjects to responses for another group of subjects can be addressed through LCV. Using responses for the second group of subjects, apply the item-factor reallocation process of Section 2.7 starting from the allocation based on item responses for the first group of subjects. The scales are reasonably applicable to the second group of subjects if the final LCV score for the reallocation process is not too different from the initial LCV score. One important example is the assessment of the impact of missing item value imputation, which can be accomplished by applying the reallocation process to item responses for subjects with no missing item values starting from the item-factor allocation determined from data with imputed item responses. Another important example involves family measures developed using responses from one type of family member (e.g., the mother) but to be completed by other types of family members (e.g., fathers and/or children).

3 The example survey data

The Family Management Measure (FaMM)[22] (see http://nursing.unc.edu/research/famm/, accessed 10/30/10) was developed using the Family Management Style Framework[23,24] to assess parents' perceptions of family management of a child's chronic condition. Items were developed to address eight theoretical dimensions: child identity, family focus, future expectations, illness view, management approach, management mindset, parental mutuality, and parenting philosophy.[24] A telephone survey was conducted of parents of a child with a chronic condition utilizing I=65 preliminary items. Example scale development analyses use responses from the n=349 partnered mothers from this survey. Partners and single mothers also participated, but their responses were not used in developing the scales. Single mothers did not respond to parental mutuality items and the imputation of these items would be inappropriate. The inclusion of partner responses would violate the assumption of independence for responses from different subjects. However, the applicability of the scales developed from partnered mothers to responses from single mothers and from partners can be addressed as described in Section 2.13.

3.1 Responses to items

FaMM items have response values 1-5 with 1 meaning strongly disagree, 5 strongly agree, and intermediate values unlabeled. To obtain extra subjective information about items, surveyed parents were allowed to classify non-response (i.e., not 1-5) as “don't know,” “not applicable,” or “refused.” Additionally, item values could be missing as well. Of the m=65·349=22 685 possible responses to the I=65 FaMM items by the n=349 mothers, only 222 (1.0%) of these are non-response values of any kind. Non-response is thus quite limited in this data set. Consequently, all non-response values are imputed for example analyses as described in Section 2.3.

3.2 Number of subjects

Standard conventions for numbers of subjects needed for developing scales with factor analysis typically range from 5 to 10 subjects per item with 300 subjects considered large enough to relax these restrictions.[18] The example scale development analyses are based on responses for 5.4 (349/65) subjects per item, at the low end of the conventionally acceptable range, but the number of subjects (349) is well above the cutoff of 300 for relaxing these restrictions. The results of MacCallum et al.[25] indicate that there is an interplay between sample size, the level of communalities, and the ratio of number of items to number of factors (the overdetermination). They found that, for 6.7 items per factor, factor analysis can be effective, even for small sample sizes and low communalities, suggesting that results for example analyses with at least 6.5 items per factor (65 items with 10 factors, the maximum number considered in the analyses) are dependable.

3.3 Item summary statistics

FaMM item means range from 1.36-4.76 (without reverse coding). Seven items have extreme means either less than 1.5 (item 42) or greater than 4.5 (items 23, 30, 39, 40, 52, and 65). Item standard deviations range from 0.63-1.65. The 5 items (23, 30, 39, 42, and 52) with the least variability have standard deviations ranging from 0.63-0.78. All of these items also have extreme mean values. The strongest relationship, as measured by the maximum percent explained variability (i.e., squared bivariate correlations or R^2 expressed as a percentage), for a given FaMM item with each of the other items ranges from 2%-44%. There are 17 items (3, 5-7, 15, 20, 30, 36, 37, 39, 40, 42, 45, 46, 50, 56, and 64) at most weakly related to the other items in the sense that their strongest relationship with each of the other items is less than 10%. Four of these items also have extreme means so that, if items with extreme means, small variability, or weak relationships were all eliminated prior to factor analysis, a total of 20 (30.8%) would be eliminated. However, all I=65 available items are used in example analyses.

3.4 Assessment of the theory-based scales

Table 1 contains reliability values for the 8 possible theory-based scales computed for the n=349 mothers using all I=65 items with reverse codings based on item wordings as determined prior to data collection. Four of them, child identity, family focus, illness view, and parental mutuality, have acceptable reliability values of at least 0.70, and can be improved further through item deletion. Two other scales, future expectations and management mindset, have marginal reliability values above 0.60 but below 0.70. These can be improved by item deletion to 0.67 and 0.69, respectively, but still remain less than 0.70. The other two scales, management approach and parenting philosophy, have very poor reliability values of 0.33 and 0.46, respectively. They can be improved by item deletion, but only to poor levels of 0.50 and 0.58, respectively. Consequently, 4 of 8 of the theory-based scales have unacceptable reliability, even after item deletion, and so scales as specified solely through theory require further adjustments. Such adjustments of theory-based scales can be based on LCV, but it was decided instead to use EFA to develop completely data-driven scales. In any case, similar adjustments would be required for unacceptable theory-based scales as for EFA-generated scales.

4 Example scale development based on LCV

Example analyses are presented in this section to demonstrate the use of LCV in guiding scale development using responses for the n=349 partnered mothers of the FaMM survey.

4.1 Overview of the analysis

The analyses proceed through several steps, starting with an assessment of numbers of folds and of factors. Maximum likelihood estimation was used to extract factors and always converged. Which rotation to use is addressed next followed by a demonstration of how to adjust item-factor allocations using LCV. The adjusted factors then are combined into a reduced set of factors with removal of ineffective factors and reassignment of their items. After that, the resulting scales are improved through item removal, generating the allocation for the final FaMM scales. While this approach is effective in this case, alternative approaches can also be considered (see Section 4.11). Whatever process is used, LCV provides objective support for decisions made throughout that process. However, it is not the only issue that needs to be considered. Reliability of generated scales is also an important issue, as is their construct validity. Most importantly, generated scales should be interpretable and conceptually meaningful, and so the FaMM research team assessed results at the various stages on this basis.

4.2 Choosing numbers of folds and factors

4.2.1 Impact of the number of folds—Table 2 presents results of maximizing LCV scores over 0 to 15 factors using a selection of numbers k of folds. The maximum LCV score is attained at 10 factors for all k except k=2, indicating that choosing the number of factors through LCV can be robust to the number of folds as long as it is not too small. For all but k=2, there is at most a small difference in scores (and so they are reported to five digits to distinguish between values) of less than 1% from the best score (occurring at k=35), suggesting that the choice for the number of folds, if large enough, is not likely to have much impact on the scores for other analyses of these data. Consequently, the first local maximum in k, in this case k=15, is used in all subsequent analyses of the FaMM items. These results along with those reported by Knafl and Grey[2] indicate that small numbers k of folds can produce different results, but that otherwise the selected number F of factors is robust to the setting for k.

4.2.2 Alternative numbers of factors—Table 3 contains 15-fold LCV scores for EFA models with 0 to 20 factors. While 10 factors produce the best score, choices between 5-16 factors generate scores within 1% of that best score. Consequently, there can be a wide range of alternatives to the optimal number of factors generating LCV scores close to the best score, suggesting an explanation for the difficulty in general of choosing the number of factors: there are likely to be many competitive alternatives.

Table 3 also reports the LCV score for the model based on a general, unstructured covariance matrix for the items, which is 5.45% less than the best score for 10 factors. In comparison, the independent-items or 0-factors model generates a score 10.57% lower than the best score. Thus, the unstructured model is distinctly preferable to the independent-items model, indicating that there is a distinct amount of dependence between responses to the FaMM items. On the other hand, factor analysis models with positive numbers of factors improve on the unstructured model, and substantially so for many choices. Factor analysis models with moderate numbers of factors provide effective descriptions of the dependence in responses to the FaMM items, and they have far fewer parameters than the unstructured model (650 loadings for the 10-factors solution compared to 2 080 correlations between pairs of different items). See Knafl and Grey[2] for a comparison of LCV to other approaches for choosing the number of factors.

4.2.3 Parsimonious choices for the number of factors—Table 3 results also suggest that fewer factors than 10 might be parsimonious alternatives. This is addressed in Table 3 using LCV ratio tests. The maximum fold size is 34 (or 9.7%) of the 349 mothers, and so deleted parameter estimates are based on data for sufficiently large numbers of mothers (at least 315) for asymptotics to apply. LCV ratio tests reported in Table 3 are significant ($p<0.01$) for 0-4 factors and distinctly nonsignificant ($p\geq0.18$) for 5-9 factors, suggesting that 5-9 factors are parsimonious alternatives to the optimal choice of 10 factors. These alternatives have LCV scores distinctly less than 1% of the best score, suggesting that percent decreases of 1% or more may be considered noncompetitive alternatives to any factor analysis model based on the FaMM item responses, not just the 10-factors EFA model. A percent decrease of 1% may not seem very large, but it is substantive for these data with the very large number of m=22 685 measurements. Although smaller numbers of factors than 10 are competitive choices, the 10-factors solution is investigated in subsequent analyses to assess whether it provides an acceptable set of unweighted scales, or whether fewer factors are needed instead.

4.3 Rotating the factors

Table 4 contains LCV scores for CFA models corresponding to 22 different rotations of the 10-factors EFA loadings, including 6 orthogonal rotations, oblique versions and oblique promax extensions of these 6 orthogonal rotations, and 4 other oblique rotations, as well as not rotating.[26] Varimax rotation generates the best item-factor allocation, with associated CFA model having the best LCV score for all of the Table 4 rotations. The Harris-Kaiser oblique rotation generates the worst score for rotations 1.22% less than the best score. All other rotations generate competitive allocations with scores well within 1% (at most 0.54%) of the best score. There is not much impact to which rotation is used as long as it is not Harris-Kaiser. On the other hand, unrotated loadings generate a distinctly poor allocation with the worst overall score, 1.64% lower than the best score.

4.4 Adjusting item-factor allocations

Varimax rotation of the 10-factors EFA model generates the best LCV score for all the Table 4 rotations, and so provides a natural place to start to search for an improved allocation. Table 5 describes the results of these adjustments. Eight items are reallocated with one of these a removal. The LCV score for the final allocation is 0.23371. Even though the adjustment process reallocates eight (12.3%) of the 65 FaMM items, the LCV score only improves by a very small amount to 0.09% larger than the score of 0.23350 for the varimax-based allocation (with estimated loadings rather than unit loadings as in Table 4). Moreover, reliability scores are not distinctly better (reliabilities for scales based on factors 1-10 change from 0.89 to 0.90, 0.78 to 0.78, 0.58 to 0.62, 0.64 to 0.65, 0.65 to 0.61, 0.59 to 0.55, 0.75 to 0.75, 0.60 to 0.61, 0.76 to 0.65, and 0.68 to 0.68, respectively). Four scales have improved reliability, but three others have worse reliability. In any case, since the adjusted allocation is available for the FaMM items and generates the better score, it is used as the basis for subsequent analyses. However, both the varimax-based allocation and the adjusted allocation produce some problematic scales with either small numbers of associated items (i.e., factor 10 has only two items in both cases) and/or reliability scores below 0.70, suggesting that there should be less than 10 factors.

4.5 Combining factors

Table 6 contains factor combination results starting from the adjusted item-factor allocation. Three pairs of factors are combined, first factors 4 and 8, then factors 9 and 10, and finally factors 3 and 5. After each combination step, the reduced item-factor allocation is adjusted using the search process of Section 2.7, but this results in only one adjustment with item 3 reallocated from factor 5 to factor 6 at step 2. The first two combinations generate scales with reliability scores of at least 0.70. The scale generated in the third step has reliability 0.69 reasonably close to 0.70. Furthermore, combining this factor with factor 6, the one remaining factor with associated reliability less than 0.70, generates a scale with reliability 0.69 even though based on 20 items. Consequently, the combination process is terminated after three steps producing a 7-factor allocation.

4.6 Removing factors and reassigning their items

Six of the seven scales generated by factor combination have reliability at least 0.70 or close to 0.70. The exception is the scale corresponding to the original factor 6 (with item 3 added). It has a poor reliability of 0.57, even though based on 10 items. For this reason, this factor is dropped, leaving a 6-factor allocation. It is possible that some of the 10 items allocated to this factor might provide valuable contributions to the other factors. This is investigated by starting from the reduced 6-factors allocation with all 10 items (3, 7, 8, 18, 30, 37, 39, 45, 46, 53) of the original factor 6 removed, and then systematically reassigning items from among these 10 to one of the six reduced factors. Five items (8, 18, 30, 45, and 53) are reassigned by this process. Altogether six items remain removed, the other five items from the original factor 6 together with item 36 removed as part of the adjustments described in Table 5. Finally, the reassigned allocation is adjusted using the search process of Section 2.7. Only one item is reallocated, item 15 from the reduced factor 3 to the reduced factor 6.

The resulting reduced 6-factors allocation generates five of six scales with reasonable reliability of at least 0.70 and a sixth scale with reliability of 0.69. Removal of the 10 items of the original factor 6 causes the LCV score to decrease from 0.23266 (last entry of Table 6) to 0.23150. Item reassignment increases it to 0.23217. The whole process decreases the LCV score from 0.23371 (last entry of Table 5) by 0.66% to 0.23217. A decrease distinctly less than 1% seems an acceptable trade-off given that this solution produces distinct reliability benefits for resulting scales. Moreover, 6 factors is a parsimonious choice compared to the optimal choice of 10 factors (Table 3). The scales for the reduced allocation have items whose deletion would improve the reliability of those scales, but items are not deleted here on this basis. An item whose deletion improves its scale's reliability can still have sufficient value to retain it in the scale in the sense that its deletion results in a reduced LCV score.

4.7 Removing items from scales

4.7.1 The LCV-based approach—So far, six items have been removed, one (36) as part of the adjustment process of Section 2.7 applied to 10 factors and five others (3, 7, 37, 39, and 46) from the original factor 6. Other items might possibly be removed as well. The model can be improved by removing six more items (5, 6, 45, 64, 49, and 50, in that order) generating the final allocation. Of the five items previously reassigned from among items allocated to the original factor 6, only one (item 49) is removed again in this step, indicating that the reassignment step is needed in general. The LCV score for the initial unweighted scale model is 0.22736, which improves to 0.22878 with the removal of the extra 6 items.

When loadings are estimated for this final allocation, they are all highly significant with absolute standardized values ranging from 4.52 to 17.67 indicating that all remaining items provide substantive contributions to the final FaMM scales. The LCV score for the model with estimated loadings is 0.23172, an increase of 1.29%. Hence, weighted scales can generate distinct improvements in LCV scores compared to unweighted scales, but the research team decided to use unweighted FaMM scales to simplify their computation in practice.

4.7.2 Comparison to preliminary item elimination approaches—Twelve items are removed on the basis of LCV, including only one item (39) of the seven items with extreme mean values. On the other hand, most (11/12; all but item 49) of these LCV-removed items are among the 17 items with the weakest relationships (in terms of maximum percent explained variability; see Section 3.3), but six of the 17 weakest items are not removed by LCV with only three of these having extreme mean values. Removal of items on the basis of summary statistics would eliminate 20 items, substantially more (i.e., eight or 12.3% more) of the 65 available items. These extra items have positive impacts on the final FaMM scales, generating larger LCV scores when included than when removed as well as highly significant standardized loadings.

4.7.3 Comparison to item elimination based on the strength of loadings—Using loadings of FaMM items for the 10-factors EFA model rotated by varimax with the best LCV score in Table 4, a cutoff of 0.5 would eliminate 37 (56.9%) of the 65 FaMM items, all due to not loading strongly on any factors. Lowering the cutoff to 0.4 would eliminate 22 (33.8%) items, 18 due to not loading strongly on any factors and 4 others for loading strongly on multiple factors. Lowering it further to 0.3 would eliminate 21 (32.3%) items, 7 due to not loading strongly on any factors and 14 others for loading strongly on multiple factors. Conventional approaches based on the strength of loadings eliminate 21-37 items, substantially more (9-25 or 13.8%-38.5% more) than the 12 LCV-eliminated items, suggesting that these conventional approaches are likely in general to result in loss of substantial numbers of useful items. LCV scores for models corresponding to cutoffs of 0.3, 0.4, and 0.5 are respectively 0.22018, 0.22483, and 0.22348 or 3.89%, 1.86%, and 2.45% lower than the score of 0.22909 for the model based on varimax-rotated loadings without item elimination. The intermediate cutoff 0.4 generates the best score of these three, but all are distinctly inferior to not eliminating items at all. Thus, conventional item elimination criteria can result in substantive loss of information, i.e., eliminating sizeable numbers of effective items with highly significant standardized loadings and whose removal produces substantial reductions in LCV scores.

4.8 Assessment of the final FaMM scales

Even if scales are going to be unweighted, standardized loadings on associated factors can be used to order items on their importance to those scales, facilitating the interpretation of underlying constructs for the scales. Table 7 lists the four most important items for each of the final FaMM scales. Based on item wordings for their most important items, the FaMM scales corresponding to the final factors 1-6 have been named family life difficulty, parental mutuality, condition management ability, view of condition impact, condition management effort, and child's daily life, respectively[20] (or difficulty, mutuality, ability, impact, effort, and child for short). Thus, the final FaMM scales have meaningful interpretations for family management of childhood chronic conditions. Moreover, they are distinct from the original theoretical dimensions used to generate the items. Only mutuality (with all but one of its items from the parental mutuality dimension) and child (with all of its items from the child identity dimension) overlap with the theoretical dimensions. The other four scales are based on mixtures of items from several theoretical dimensions and so represent new latent constructs.

These scales also all have acceptable reliability at least 0.70 (ranging from 0.71-0.90). Several measures were collected for assessing the construct validity of the FaMM scales. One of these measures was the Functional Status II, a measure of child functioning.[27] Larger scores indicate better child functioning and so should correspond to lower difficulty, effort, and impact scores and higher ability, child, and mutuality scores. Correlations with child functioning for responses from partnered mothers (348 due to a missing child functioning value) are in expected directions (0.33, 0.43, −0.50, −0.37, −0.31, and 0.20 for ability, child, difficulty, effort, impact, and parental mutuality, respectively) with all significantly nonzero (all with $p<0.01$). Similar results held for the other available measures,[22] thereby providing evidence supporting the construct validity of the scales.

4.9 Impact of missing item value imputation

The final six FaMM scales were developed using imputed missing item values which might have affected the results. This issue is addressed here by considering how much the final item-factor allocation changes when it is adjusted using the item-factor adjustment process of Section 2.7 applied to the FaMM item responses from the 248 partnered mothers who responded to all 65 FaMM items. LCV scores are computed for these reduced mother data using 15 folds because analyses of the full partnered mother data suggest that the number of folds if large enough is not likely to have much of an impact. The final item-factor allocation generated using the full partnered mother data when applied to the reduced mother data generates an LCV score of 0.22944. The adjustment process started at this allocation reallocates 3 items with an improved LCV score of 0.22948, an increase of only 0.02%. While an improved allocation can be obtained by reallocating 3 (4.6%) of the 65 items, this improved allocation has little effect compared to using the allocation based on responses with missing item values imputed. Thus, the FaMM scales developed using item imputation are reasonably compatible with non-imputed responses. Imputation has not affected the results much, but allows the analyses to be based on more extensive data.

4.10 Applicability of FaMM scales to other sets of subjects

The final six FaMM scales will be used in practice with other types of family members including partners and single mothers, and so it is important to assess the applicability of the FaMM scales to these other family members. This issue can be addressed similarly to the assessment of the impact of imputation of Section 4.9 by considering how much the final item-factor allocation changes when adjusted using the item-factor adjustment process of Section 2.7 applied to FaMM item responses from other sets of family members. These analyses are not reported for brevity, but they lead to the conclusions that the FaMM scales developed using responses for only the partnered mothers are reasonably compatible with responses from partners and from single mothers.

4.11 Alternative scale development approaches

The example analyses followed a specific process that progressively led to a final set of scales for the example data. The steps in this analysis process can be conducted in alternate ways and orders, and may not be all necessary for some data. For example, when the number of factors generating the best EFA model produces some scales with unacceptable reliability, the number of factors can be reduced one at a time, using associated EFA models to generate new set of scales with one less scale, and continue until that produces a set of scales with acceptable reliability. This is not guaranteed to hold, and so it seems best to limit the search to numbers of factors generating competitive LCV scores (e.g., 5-10 factors for the example data). As another example, the item-factor allocation of Section 2.7 had little effect on LCV scores. This also held for the two cases reported by Knafl and Grey,[2] suggesting that this adjustment may not be necessary in scale development, which would save substantial amounts of computation. Similarly, it did not have much impact on factor combination and item reassignment after factor removal in the example analyses, and so may not be necessary for those steps as well. However, item-factor allocation adjustment is useful for assessing the applicability of scales to responses from different types of subjects than those used in their development. In any case, whatever scale development process is used, as long as it is sufficiently complete, resulting scales are likely to be competitive, if not the same as those generated by the process used in the example analyses.

5 Discussion

The example analyses demonstrate the usefulness of LCV for guiding scale development, including choosing how many factors to extract, which rotation to use, how to allocate items to factors, which factors to combine, which items to remove, and assessing the applicability of developed scales to other groups of subjects. A wide variety of alternative EFA and CFA models were considered to assess the use of LCV for supporting factor analysis, but, in practice, a different set of models would be used, adapted to the problem and data under analysis, possibly following a somewhat different process. However, the example analyses demonstrate that LCV-based scale development can be effective in producing scales with meaningful interpretations as well as acceptable reliability and construct validity.

5.1 Comparison to the classical approach

A classical EFA involves a series of subjective decisions. Items may be discarded prior to EFA on the basis of their summary statistics. The item correlation matrix is inspected to verify that sufficiently strong correlation exists to support conducting an EFA. A number of factors has to be chosen, then a rotation scheme to apply to those factors. A cutoff for a strong loading has to be chosen to be used to discard items that do not load strongly on any factors or that load strongly on multiple factors. LCV can be used to make the EFA scale development more objective and, as demonstrated in the example analyses, can produce much different results.

LCV provides an objective approach for choosing an initial number of factors and refining that choice until scales are generated with acceptable reliability, sufficient numbers of associated items, and interpretability for the underlying theoretical domain. LCV also provides an objective way to select from among the many alternative rotation schemes available for EFA. While the commonly used varimax procedure was chosen in the example analyses and in the two cases reported by Knafl and Grey,[2] there is no guarantee that it will always be an effective choice.

Item removal prior to EFA on the basis of summary statistics can result in elimination of sizeable numbers of effective items. In the example analyses, this eliminated 8 more (20 compared to 12) or 12.3% of the 65 available items than with LCV. Item removal on the strength of rotated EFA loadings eliminated from 9-25 more, or 13.8%-38.5%, of the 65 available items than with LCV. In both cases, the extra eliminated items were effective since LCV scores were larger with items included than with them removed. Furthermore, all the remaining items provided substantial contributions to their associated scales since they had very strong estimated loadings (with absolute standardized values of 4.52 and larger).

Inspection of the unstructured item correlation matrix is recommended to verify subjectively that item correlations are sufficiently strong overall to justify the existence of distinct factors.[19] The factor structure may be so weak that a 0-factors or independent-items model may better reflect the correlation structure. LCV can be used to assess this issue by comparing the LCV score for the 0-factors or independent-items model to the maximum LCV score in the number of factors. If the score for the 0-factors model is close to the best score, for example, within 1%, conducting a factor analysis is likely to generate deceptive results. On the other hand, in the scale development context, items are chosen to address related theoretical constructs, and so are likely to be distinctly correlated, so that the 0-factors model is unlikely to have a competitive LCV score. Factor analysis still might produce deceptive results though. If the unstructured model for the item covariances has a distinctly better LCV score than the best score for factor analysis models, those covariances have more complex structure than can be modeled effectively with factor analysis. However, the analyses reported here together with those reported by Knafl and Grey[2] suggest that factor analysis with a moderate number of factors is likely to provide an effective depiction of the item covariance structure, intermediate between the simplicity of the independent-items model and the complexity of the unstructured model.

When items have been developed to reflect theoretical dimensions as for the example data, a classical CFA can be conducted using reliabilities and fit statistics to confirm the pre-determined factor structure. However, it is quite likely for theory-based scales to require substantial adjustments as was the case for the example data. These can be based on classical modification indexes but can be conducted more objectively with LCV. Even if the theory-based scales have acceptable reliabilities and fit statistics, it can still be beneficial to conduct a completely data-driven EFA. The LCV score for the theory-based scales can be compared to the score for the data-driven scales to determine whether the theory-based scales provide a competitive alternative or that the data-driven scales provide a substantive improvement.

5.2 Issues for future research

There are still a variety of important issues for future research. The example analyses address the development of unweighted scales because the research team chose that alternative so that scale values can be more simply computed in practice. Further work is needed to investigate development of weighted scales. The example analyses demonstrate LCV-based scale development in the standard context with responses independent for different subjects. The data for this study, however, come from a survey involving one to two participating parents per family. Partners are excluded from those scale development analyses so that the independence assumption is not violated. Some of the mothers are single parents who do not respond to items addressing parental mutuality, and so they are excluded as well since it is inappropriate to impute item values on parental mutuality for single mothers without partners. It is possible to justify that the scales developed from partnered mother responses are applicable to responses from these other types of parents, but it would be preferable to base the scale development on the combined responses for all participating parents, while accounting for inter-parental correlation as well as with item means changing with parent type.

EFA and CFA models are based on a normality assumption that may be inappropriate for items with limited numbers of values (1-5 for the example data). It is possible to conduct residual analyses to assess this assumption,[2] but methods are needed to handle distinctly non-normal data. Extensions to the multivariate normal distribution accounting for skew[28] or for kurtosis[29] can be used to adjust for non-normality. LCV scores are then based on the adjusted likelihood functions. Also, more general item factor analysis (IFA) models treating items as categorical rather than as continuous variables would be more natural choices.[30] These models are not currently supported in SAS but are in other statistical software tools like JMP (SAS Institute, Inc, Cary, NC), LISREL (Scientific Software International, Lincolnwood, IL), and Mplus (Muthén & Muthén, Los Angelos, CA). LCV can be extended to IFA as well. While these alternative approaches are computationally more intensive, they may support development of better scales, and so there is a need to extend LCV to these contexts.

The example analyses address the case with each item allocated to only one factor, as is typical for scale development. However, items sometimes are allocated to multiple factors, and the use of LCV in that case still needs investigation. LCV also applies to structural equation modeling as well as to all types of latent variable modeling[1], not just to factor analysis. The LCV formulation applies without change to this more general context, but its usefulness for more general latent variable modeling still needs to be investigated. Furthermore, the use of LCV still needs to be investigated through simulation studies.

Acknowledgments

This work has been supported in part by NIH Grants R01 NR08048 and R01 AI057043 and by Oregon Health & Science University's Oregon Clinical and Translational Research Institute (OCTRI) through NIH Grant UL1 RR024140.

References

1. Rabe-Hesketh S, Skrondal A. Classical latent variable models for medical research. Statistical Methods in Medical Research. 2008; 17:5–32. [PubMed: 17855748]
2. Knafl GJ, Grey M. Factor analysis model evaluation through likelihood cross-validation. Statistical Methods in Medical Research. 2007; 16:77–102. [PubMed: 17484294]
3. Browne MW. Cross-validation methods. Journal of Mathematical Psychology. 2000; 44:108–32. [PubMed: 10733860]
4. Browne, MW.; Cudeck, R. Alternative ways of assessing model fit.. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Sage; 1993. p. 136-62.
5. Burman P. A comparative study of ordinary cross-validation, ν-fold cross-validation and the repeated learning-testing methods. Biometrika. 1989; 76:503–14.
6. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection.. In: Mellish, CS., editor. Proceedings of the 14th international joint conference on artificial intelligence.; Morgan Kaufman. 1995; p. 1137-43.
7. Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. Journal of the Royal Statistical Society, Series B. 1977; 39:44–7.
8. Azzalini A, Bowman AW, Härdle W. On the use of nonparametric regression for model checking. Biometrika. 1989; 76:1–11.
9. Staniswalis JG. The kernel estimate of a regression function in likelihood-based models. Journal of the American Statistical Association. 1989; 84:276–83.
10. Lee JC. Tests and model selection for the general growth curve model. Biometrics. 1991; 47:147–59. [PubMed: 2049496]
11. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. Applied Statistics. 1999; 48:313–29.
12. Smyth P. Model selection for probabilistic clustering using cross-validated likelihood. Statistics and Computing. 2000; 10:63–72.
13. Knafl GJ, Fennie KP, Bova C, Dieckhaus K, Williams AB. Electronic monitoring device event modelling on an individual-subject basis using adaptive Poisson regression. Statistics in Medicine. 2004; 23:783–801. [PubMed: 14981675]
14. Knafl, GJ.; Fennie, KP.; O'Malley, JP. Adaptive repeated measures modeling using likelihood cross-validation.. In: Bovaruchuk, B., editor. Proceedings second IASTED international conference on computational intelligence. ACTA Press; 2006. p. 422-27.
15. Huisman M. Imputation of missing item responses: Some simple techniques. Quality and Quantity. 2000; 34:331–351.
16. Schafer, JL. Analysis of incomplete multivariate data. Chapman & Hall; 1997.

17. Arminger G, Sobel ME. Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. Journal of the American Statistical Association. 1990; 85:195–203.
18. DeVellis, RF. Scale development: Theory and applications. 2nd. ed.. Sage; 2003.
19. Nunnally, JC.; Bernstein, IH. Psychometric theory. 3rd ed.. McGraw-Hill; 1994.
20. Hatcher, L. A step-by-step approach to using SAS for factor analysis and structural equation modeling. SAS Institute; 1994.
21. Floyd FJ, Widamen KF. Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment. 1995; 7:286–99.
22. Knafl K, Deatrick JA, Gallo A, Dixon JK, Grey M, Knafl GJ, O'Malley JP. Assessment of the psychometric properties of the Family Management Measure. Journal of Pediatric Psychology. in press. doi: 10.1093/jpepsy/jsp034.
23. Knafl K, Breitmayer B, Gallo A, Zoeller L. Family response to childhood chronic illness: Description of management styles. Journal of Pediatric Nursing. 1996; 11:315–26. [PubMed: 8908900]
24. Knafl K, Deatrick J. Further refinement of the Family Management Style Framework. Journal of Family Nursing. 2003; 9:232–56.
25. MacCallum RC, Widamen KF, Zhang S, Hong S. Sample size in factor analysis. Psychological Methods 1999. 1999; 4:84–99.
26. SAS Institute. SAS/STAT 9.1 user's guide. SAS Institute; 2004.
27. Stein R, Jessop D. Functional Status II: A measure of child health status. Medical Care. 1990; 28:431–38.
28. Azzalini A, Capitanio A. Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society, Series B. 1999; 61:589–602.
29. Shapiro A, Browne MW. Analysis of covariance structures under elliptical distributions. Journal of the American Statistical Association. 1987; 82:1093–7.
30. Wirth RJ, Edwards MC. Item factor analysis: Current approaches and future directions. Psychological Methods. 2007; 12:58–79. [PubMed: 17402812]

Table 1

Internal Consistency Reliability for Theory-Based Scales[a]

Scale	No. Items	Alpha[b]	Item Deletions[c]	Adjusted No. Items[d]	Adjusted Alpha[b]
child identity	9	0.73	46 39 49 15	5	0.78
future expectations	8	0.67	47	7	0.67
family focus	8	0.73	8	7	0.76
illness view	8	0.75	40 31	6	0.77
management approach	7	0.33	36 37 45 7	3	0.50
management mindset	9	0.60	5	8	0.69
parental mutuality	8	0.78	64 58 60 61 62	3	0.83
parenting philosophy	8	0.46	6 3	6	0.58

[a]Using the item-scale allocation for the theory-based dimensions with reverse codings determined from item wordings.

[b]Cronbach's alpha for the scale obtained by summing up associated items after reverse coding negatively coded items, first using all the items allocated to the scale and then after deleting all appropriate items.

[c]Items whose deletion improves the associated scale's reliability in the order corresponding to the largest possible improvement in the reliability continuing until no further improvement is possible. Items 5, 6, and 46 have small negative item-total correlations with their associated scales, and the reliabilities of those scales improve when codings for these items are switched, but the reliability for these scales improves even more when those items are deleted.

[d]A total of 20 of the 65 items are deleted including 11 of the 12 items removed in the final allocation used in Table 7, all but item 50.

Table 2

Choosing the Number of Factors Using LCV for a Selection of Numbers of Folds

No. of Folds	No. of Factors[a]	LCV Score	% Decrease from Best LCV Score
2	6	0.22987	2.16%
5	10	0.23341	0.65%
10	10	0.23370	0.53%
15[b]	10	0.23469	0.11%
20	10	0.23452	0.18%
25	10	0.23470	0.10%
30	10	0.23485	0.04%
35	10	0.23494	0.00%
40	10	0.23488	0.03%
45	10	0.23490	0.02%
50	10	0.23481	0.06%

Note. LCV = likelihood cross-validation.

[a]The number of factors with maximum LCV score for 0-15 factors.

[b]The first local maximum for reported LCV scores in the number of folds, which is used in subsequent analyses.

Table 3

Impact of the Number of Factors

No. of Factors	LCV Score[a]	Decrease from Best LCV Score	Decrease in Degrees of Freedom	LCV Ratio Test P-Value[b]
0	0.20987	10.57%	650	<0.01
1	0.22627	3.59%	585	<0.01
2	0.22949	2.21%	520	<0.01
3	0.23085	1.63%	455	<0.01
4	0.23150	1.36%	390	<0.01
5	0.23306	0.69%	325	0.63
6	0.23337	0.56%	260	0.56
7	0.23413	0.24%	195	1.00
8	0.23447	0.09%	130	1.00
9	0.23430	0.16%	65	0.18
10	0.23469	0.00%	0	--
11	0.23445	0.10%	--	--
12	0.23403	0.28%	--	--
13	0.23368	0.43%	--	--
14	0.23338	0.56%	--	--
15	0.23276	0.82%	--	--
16	0.23252	0.92%	--	--
17	0.23214	1.08%	--	--
18	0.23140	1.40%	--	--
19	0.23117	1.50%	--	--
20	0.23061	1.74%	--	--
UN[c]	0.22190	5.45%	--	--

Note. LCV = likelihood cross-validation; UN = unstructured.

[a]Computed using 15 folds.

[b]The p-value for the χ^2 test for the difference of $-2\cdot\log(\mathrm{LCV}^m)$ values for the given number of factors compared to 10 factors, where m is the total number of measurements (item values) for all subjects (mothers). The associated degrees of freedom parameter for this test is given by the difference in the number of degrees of freedom for the two factor settings which equals the difference in the number of factors times 65 because removing a factor reduces the number of loadings by the number of items.

[c]The multivariate normal model for the items with general, unstructured covariance matrix.

Table 4

Comparison of Unweighted Scales Suggested by Rotations[a]

Rotation Approach	LCV Score[b]	% Decrease from Best LCV Score
Orthogonal		
Biquartimax	0.22861	0.21%
Equamax	0.22892	0.07%
Factor Parsimax	0.22905	0.02%
Parsimax	0.22905	0.02%
Quartimax	0.22785	0.54%
Varimax	0.22909	0.00%
Oblique		
Biquartimin	0.22818	0.40%
Covarimin	0.22839	0.31%
Harris-Kaiser	0.22630	1.22%
Oblimin	0.22842	0.29%
Oblique version of		
Biquartimax	0.22818	0.40%
Equamax	0.22848	0.27%
Factor Parsimax	0.22813	0.42%
Parsimax	0.22848	0.27%
Quartimax	0.22842	0.29%
Varimax	0.22837	0.31%
Promax starting from		
Biquartimax	0.22880	0.13%
Equamax	0.22814	0.41%
Factor Parsimax	0.22848	0.27%
Parsimax	0.22868	0.18%
Quartimax	0.22850	0.26%
Varimax	0.22882	0.12%
None	0.22533	1.64%

Note. LCV = likelihood cross-validation; EFA = exploratory factor analysis; CFA = confirmatory factor analysis.

[a]An item is allocated to the factor for which it achieves its maximum absolute rotated loading for the 10-factors EFA model. The sign of that loading determines if the item is positively or negatively coded in the associated unweighted scale. All items are allocated, each to a single factor.

[b]Computed using 15 folds for the CFA model corresponding to the associated unweighted scales with items assigned to unique factors with unit loadings (±1 depending on codings) and with all factors treated as correlated with each other.

Table 5

Adjustments to the Item-Factor Allocation Suggested by Varimax Rotation

Step	Adjustment[a]	LCV Score[b]
0	--	0.23350[c]
1	item 5 to factor 9	0.23357
2	item 46 to factor 6	0.23362
3	item 6 to factor 7	0.23365
4	item 3 to factor 5	0.23368
5	item 39 to factor 6	0.23370
6	item 36 removed	0.23370[d]
7	item 42 to factor 5	0.23371
8	item 6 to factor 9	0.23371[e]

Note. LCV = likelihood cross-validation; CFA = confirmatory factor analysis

[a]The adjustment to the current allocation with the best LCV score considering each of the 10 items with the weakest standardized loadings in the current CFA model shifted to each of the other 9 possible factors or removed from the model (i.e., by setting the item's loadings for all factors to zero). Multiple removed items are treated as having an unstructured covariance matrix.

[b]Computed using 15 folds for associated CFA models with estimated loadings and estimated inter-factor correlations.

[c]For the item-allocation based on varimax rotation given in Table 4.

[d]Larger scores than for the previous step, but the same after rounding, are produced by moving item 36 to factor 2 or to factor 6 as with it removed, and so item 36 is removed as ineffective. If item 36 had been reallocated to factor 6, the reallocation process would have terminated at the next step.

[e]The next best shift would be to reverse step 8 generating a lower score which rounds to the same amount, and so the search process is terminated. The LCV score for this final adjusted allocation is only negligibly (0.09%) larger than the LCV score for the initial allocation.

Table 6

Combining Factors of the Adjusted Item-Factor Allocation

Step	Combined Factors[a]	No. Items	LCV Score[b]	Alpha[c]
0	--	--	0.23371	--
1	4 + 8	11	0.23348	0.72
2	9 + 10[d]	7	0.23315	0.70
3	3 + 5[e]	10	0.23266	0.69

Note. LCV = likelihood cross-validation; CFA = confirmatory factor analysis.

[a]At each step, the two factors generating the best score are combined, considering only factors with associated scales having reliabilities less than 0.70. After each combination, item-factor allocations are adjusted using the same process as for Table 5.

[b]Computed using 15 folds for associated CFA models with estimated loadings and estimated inter-factor correlations.

[c]Cronbach's alpha for the scale obtained by summing up associated items after reverse coding negatively coded items.

[d]Item 3 is shifted as well from factor 5 to factor 6 increasing the scale reliability scores for factor 5 to 0.64 and for factor 6 to 0.57. This is the only step affected by the adjustment process.

[e]The combined factor 3 + 5 (with item 3 removed) and factor 6 (with item 3 added) are the two remaining factors with associated scales having reliability less than 0.70. When these are combined, the reliability is 0.69, still less than 0.70, even though based on 20 items. So the combination process is terminated after the third step, and the combined factor 3 + 5 is not combined with factor 6.

Table 7

Four Most Important Items for Factors in the Final Allocation

Factor	Standardized Loading	Item Wording
1. Difficulty		
	17.21	A condition like the one our child has makes it very difficult to lead a normal life.
	17.17	A condition like the one our child has makes family life very difficult.
	16.55	Dealing with our child's condition makes family life more difficult.
	15.34	We are sometimes undecided about how to balance the condition and family life.
2. Mutuality		
	17.67	I am pleased with how my partner and I work together to manage our child's condition.
	15.90	My partner and I support each other in taking care of our child's condition.
	–15.08	I am unhappy about the way my partner and I share the management of our child's condition
	10.37	My partner and I have similar ideas about how we should be raising our child.
3. Ability		
	–11.77	We often feel unsure about what to do to take care of our child's condition.
	11.66	When something unexpected happens with our child's condition, we usually know how to handle it.
	9.43	We are looking forward to a happy future for our child
	8.36	In the future we expect our child to take care of the condition.
4. Impact		
	10.34	We think about our child's condition all the time.
	10.11	Because of the condition, we worry about our child's future.
	9.64	Our child's condition will be harder to take care of in the future.
	–7.34	People with our child's condition have a normal length of life.
5. Effort		
	–14.15	Our child's condition doesn't take a great deal of time to manage
	13.98	It takes a lot of organization to manage our child's condition
	12.15	Our child's condition is like a roller coaster with lots of ups and downs.
	8.87	Our child's condition requires frequent visits to the clinic.
6. Child		
	–16.18	Our child's friendships are different because of the condition.
	12.73	Our child's everyday life is similar to that of other children his/her age.
	–12.71	Our child is different from other children his/her age because of the condition.
	–10.26	Our child enjoys life less because of the condition.