Published in final edited form as:

Qual Life Res. 2014 March ; 23(2): 509–513. doi:10.1007/s11136-013-0474-x.

Use of factor analysis models to evaluate measurement invariance property of the Asthma Control Questionnaire (ACQ)

Yan Yan,

Washington University Medical School, St Louis, MO, USA

Wei Wu,

University of Kansas, Lawrence, KS, USA

Robert. C. Strunk, and

Washington University Medical School, St Louis, MO, USA

Jane Garbutt

Washington University Medical School, St Louis, MO, USA

Yan Yan: yany@wudosis.wustl.edu

Abstract

Purposes—To demonstrate the assessment of measurement invariance property in a health status instrument and to increase the awareness of its importance, we evaluate the measurement invariance of the Asthma Control Questionnaire (ACQ) across age and gender subgroups.

Methods—Data are obtained from children 7–12 years of age at entry into a randomized trial, which evaluates the effect of a telephone coaching program on improving asthma outcome. Multi-group confirmatory factor analysis is used to assess the comparability of factor loadings and intercepts across age and gender subgroups. Since age is a continuous variable, two different categorizations (7–10 vs 11–12 and 7–9 vs 10–12) are analyzed.

Results—The factor loadings and intercepts of all six items in ACQ are comparable across gender subgroups. Although the factor loadings are comparable across age 7–10 and 11–12 subgroups, one intercept is statistically but not practically different. For age 7–9 versus 10–12 subgroup comparison, the factor loadings are not comparable.

Conclusion—In children, the ACQ can be used to compare asthma control construct between boys and girls and between age 7–10 and 11–12 subgroups. Measurement invariance is an important property that should be examined when the latent construct(s) are compared across different subgroups.

Keywords

Asthma Control Questionnaire; Measurement invariance; Factor analysis

Introduction

In many clinical and epidemiological studies, quality of life and health status outcomes are latent constructs measured by multiple item instruments. Several criteria have been proposed to examine the measurement properties of these instruments [1–3], including content validity, criterion validity, construct validity, internal consistency, reproducibility,

responsiveness, and interpretability. When the latent construct(s) are to be compared across subgroups, demonstration of an instrument's measurement invariance is necessary to ensure meaningful comparisons. The Asthma Control Questionnaire (ACQ) is the most commonly used instrument for measuring asthma control in children [4–6]. A six-item version of the ACQ is used in our large-scale clinical trial on asthma outcomes after a coaching intervention. Each item is measured on a 7-point scale, with zero being no impairment and six being maximum impairment. Although age and gender are common demographic characteristics for subgroup comparisons, the measurement invariance properties of the ACQ are unknown for these subgroups.

Methods

Study subjects and data

Study subjects of this study are children with asthma who participated in a randomized controlled trial (RCT) to evaluate a 12-month telephone asthma coaching program. The trial design and implementation are reported elsewhere [7]. In this study, we use data from the baseline assessment of the ACQ, completed during a telephonic interview with the child. The study is approved by the Washington University Human Research Protection Office.

Statistical data analysis

Confirmatory factor analysis (CFA) is a useful method to test the measurement structure of a scale [8], particularly when making cross-group comparisons because multiple-group CFA allows for simultaneous estimation and direct statistical comparisons of the measurement model parameters across the groups [8]. The fit of CFA models is evaluated using the χ^2 goodness-of-fit statistic, absolute fit indices [e.g., the root mean square error of approximation (RMSEA), standardized root mean squared residual (SRMR)], and relative fit indices [e.g., the Tucker–Lewis index (TLI) and Comparative Fit index (CFI)] [9].

The graphic illustration of the multiple-group CFA for gender is presented in Fig. 1. The degree of measurement invariance can be described as a hierarchical sequence from the weakest form to the strongest form [8, 10]. The first level—*configural* invariance only requires that the two groups have a similar basic factor structure. The second level—*weak* factorial invariance requires similarity in the factor loadings, and the third level—*strong* factorial invariance further requires similarity in both factor loadings and intercepts in the two groups. In many situations where *strong* factorial invariance is untenable, not all of the intercepts are equivalent, giving rise to partial invariance [11].

Following Byrne [11] and Little [8], we establish the different levels of ACQ measurement invariance using the χ^2 difference test statistic. We start with fitting the *configural* model (Model 0) with all factor loading and intercept parameters freely estimated across two groups. Some of the residuals are allowed to be correlated in the model as their correlations cannot be completely explained by the underlying factor. Then, we fit a *weak invariance* model (Model 1) with the factor loadings equated across the two groups. If the factor loading invariance holds (a non-significant χ^2 difference test), we continue to next step, fitting a *strong invariance* model (Model 2) with both factor loading and intercept parameters equated across two groups. If a significant χ^2 difference test rejects Model 2 at this step, indicating partial invariance, we then free the constraint on one intercept at a time to identify the item(s) responsible for partial invariance. The order in which the intercept parameters are freed is determined by the modification index (MI) and the standardized residuals [12]. Age is a continuous variable that can be categorized in different ways. Based on the previous work, we first divide the children into two age groups: 7–10 and 11–12 years [5]. We also repeat the analysis using age groups 7–9 versus 10–12 years. SAS

software v9.3 [13] is used to prepare the data and to perform simple descriptive and comparative analyses. Mplus software v6.12 [14] is used to perform multiple-group CFA.

Results

Description of study subjects

A total of 462 children, who are 7–12 years old and diagnosed with asthma, have completed the ACQ. Among them, 72.3 % (334) are 10 years old or younger, and 60.1 % (278) are boys. The means and standard deviations of the six items for the entire study sample and for the age and gender subgroups are presented in Table 1.

Multiple-group CFA

Table 2 presents the estimated factor loadings and intercepts of the multiple-group CFA models under Model 0 specification (configural invariance). The χ^2 goodness-of-fit statistics and other fit indices for configural invariance (M0), weak factorial invariance (M1), strong factorial invariance (M2) models, along with the χ^2 difference test statistics between models, are presented in Table 3.

All the models fit these data reasonably well. All p values for the χ^2 goodness-of-fit statistics are larger than 5 % (first column of Table 3). The RMSEAs and SRMRs are all <5 %, and in addition, CFI/TLI are all greater than 98 %.

The χ^2 difference test statistics (the second column) between M0 and M1 for age and gender subgroups are 5.5 ($df = 6$, $p = 0.48$) and 11.3 ($df = 6$, $p = 0.08$), respectively, indicating the similarity of factor loadings between the age and gender subgroups. The χ^2 difference tests between M1 and M2 indicate that all intercepts are similar across gender subgroups ($p = 0.64$), but some are slightly different across age groups ($p = 0.05$).

To explore which of the six intercepts contributed to the age group difference, we performed partial invariance analyses. We first free the parameter of intercept for the 5th item (acq5) since it has the largest value of modification index, and the largest and the second largest value of standardized residuals in the two age groups. The resulting model has a χ^2 value of 30.5 with $df = 25$. The χ^2 difference between this model and Model 1 ($\chi^2 = 23.7$, $df = 20$) is 6.8 with $df = 5$ (p value = 0.23), indicating that allowing the intercept of the 5th item to be different across groups substantially improved model fit. In other words, the 5th item is responsible for the partial invariance.

When we use the different categorization for the age subgroups (7–9 vs 10–12 years), a significant χ^2 difference ($\chi^2 = 15.9$, $df = 6$, $p = 0.01$) is found between M0 and M1, providing statistical evidence of different factor loadings between children in these two age groups. Using chi-square difference test, we found that the difference comes from acq3 (0.648 vs 0.929) and acq4 (0.575 vs 0.803). Failure to achieve weak factorial invariance suggests that we should not use this age grouping to compare mean asthma control level.

Discussion

In this study, we have established strong factorial invariance of ACQ in gender subgroups, but partial strong factorial invariance in age 7–10 and age 11–12 subgroups. When age is categorized into 7–9 versus 10–12 years, we fail to establish weak factorial invariance in these two subgroups.

These findings have implications for the application of ACQ and perhaps other patient-reported outcome measurements. The strong invariance of ACQ between the gender

subgroups allows researchers to compare mean asthma control level between boys and girls. Although the factor loadings for the 7–10 versus 11–12 years subgroups are comparable, the intercept of acq5 is statistically different. Nevertheless, the magnitude of difference in the two intercepts (0.187) is small. As a result, the presence of this partial measurement invariance is unlikely to have a substantial impact on the latent mean comparison. However, the measurement invariance property of ACQ is sensitive to how the age groups are formed. Our study shows that it is not appropriate to make the latent mean comparison between the age 7–9 and age 10–12 subgroups since the factor loadings are not comparable. Thus, special attention is needed when comparison subgroups are formulated from a continuous variable.

The primary limitations of the study include the following: (1) we have only considered age and gender subgroups in this study. When other group differences are of substantive interest to researchers, measurement invariance should be established for those groups, and (2) we have used the chi-square difference test to compare the models with different levels of measurement invariance. A well-known problem of this test is its sensitivity to sample size. It tends to overreject the nested model with large sample size and underreject the nested model with small sample size. A recently developed method in the Bayesian analysis framework [15] and using change in fit indices such as CFI [16, 17] to establish measurement invariance might be able to solve this problem.

There is growing recognition that patient-reported outcomes play an important role in both clinical and epidemiological studies. When the intent is to compare the latent constructs measured by these instruments across subgroups, invariance property should be assessed in addition to validity, reliability, and responsiveness. We hope that our study will increase the awareness of this important measurement property and provide a useful template for those who wish to do so.

Acknowledgments

The authors would like to thank two anonymous reviewers for their very thoughtful comments, which help us to improve the paper. Drs Yan, Strunk, and Garbutt are partially funded by a grant from the National Heart, Lung, and Blood Institute (HL072919) at the National Institute of Health.

References

1. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *Journal of Chronic Diseases*. 1985; 38:27–36. [PubMed: 3972947]
2. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*. 2002; 11:193–205. [PubMed: 12074258]
3. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. 2007; 60:34–42. [PubMed: 17161752]
4. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR. Development and validation of a questionnaire to measure asthma control. *European Respiratory Journal*. 1999; 14:902–907. [PubMed: 10573240]
5. Juniper EF, Gruffydd-Jones K, Ward S, Svensson K. Asthma Control Questionnaire in children: Validation, measurement properties, interpretation. *European Respiratory Journal*. 2010; 36:1410–1416. [PubMed: 20530041]
6. Juniper EF, Svensson K, Mork AC, Stahl E. Measurement properties and interpretation of three shortened versions of the Asthma Control Questionnaire. *Respiratory Medicine*. 2005; 99:553–558. [PubMed: 15823451]

7. Garbutt JM, Highstein G, Yan Y, Strunk RC. Partner randomized controlled trial: Study protocol and coaching intervention. *BMC Pediatrics*. 2012; 12:42. [PubMed: 22469168]
8. Little, TD.; Slegers, DW. Factor analysis: Multiple groups. In: Everitt, BS.; Howell, DC., editors. *Encyclopedia of statistics in behavioral science*. Chichester: Wiley; 2005. p. 617-623.
9. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariate structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A multidisciplinary Journal*. 1999; 6:1–55.
10. Millsap, RE. *Statistical approaches to measurement invariance*. New York: Routledge; 2011.
11. Byrne BM, Shavelson RJ, Muthen BO. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*. 1989; 105:456–466.
12. Millsap RE, Kwok OM. Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*. 2004; 9:93–115. [PubMed: 15053721]
13. SAS. SAS Institute Inc. SAS/STAT software. Release 9.3 Release 9.1. Cary SAS Institute Inc; 2010.
14. Muthen, LK.; Muthén, BO. *Mplus user's guide*. 6th. Los Angeles, CA: Muthen & Muthen; p. 1998-2010.
15. Muthén B, Asparouhov T. BSEM measurement invariance analysis. 2013 Web note 17.
16. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255.
17. Meade AW, Johnson EC, Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*. 2008; 93:568–592. [PubMed: 18457487]

Fig. 1.

Illustration of multiple-group CFA for measurement invariance of ACQ instrument between boys ($x = 0$) and girls ($x = 1$). y_1 – y_6 represent acq1–acq6, the effect of asthma control construct on acq1–acq6 (factor loadings) is denoted as λ_{01} – λ_{06} for boys and λ_{11} – λ_{16} for girls. The mean levels of acq1–acq6 (intercepts) for boys and girls are denoted as a_{01} – a_{06} and a_{11} – a_{16} , respectively. To test the measurement invariance, we determine the extent to which $\lambda_{01} = \lambda_{11}, \dots, \lambda_{06} = \lambda_{16}$, and $a_{01} = a_{11}, \dots, a_{06} = a_{16}$.

Table 1
Mean and standard deviation of six items of ACQ instrument: entire sample and by age and gender subgroups

	Total N = 462	Age [7-10] n = 334	Age [11, 12] n = 128	Boys n = 278	Girls n = 184
acq1	0.53 (1.05)	0.53 (1.02)	0.54 (1.11)	0.57 (1.08)	0.47 (1.00)
acq2	0.88 (1.20)	0.92 (1.24)	0.78 (1.08)	0.91 (1.22)	0.84 (1.16)
acq3	0.83 (1.24)	0.87 (1.28)	0.73 (1.13)	0.90 (1.32)	0.73 (1.11)
acq4	1.08 (1.35)	1.04 (1.36)	1.17 (1.31)	1.15 (1.37)	0.97 (1.30)
acq5	0.82 (1.24)	0.77 (1.24)	0.95 (1.26)	0.87 (1.27)	0.73 (1.20)
acq6	0.83 (1.11)	0.85 (1.14)	0.77 (1.02)	0.90 (1.14)	0.71 (1.05)

Table 2
Factor loadings and intercepts of six items of ACQ instrument by age and gender subgroups

	Age [7–10]	Age [11, 12]	Boys	Girls
<i>Factor loading</i>				
acq1	0.609	0.707	0.618	0.574
acq2	0.751	0.875	0.848	0.654
zcq3	0.801	0.663	0.848	0.659
acq4	0.682	0.685	0.689	0.845
acq5	0.854	1.012	0.848	0.966
acq6	0.695	0.687	0.755	0.653
<i>Intercept</i>				
acq1	0.527	0.548	0.572	0.475
acq2	0.918	0.778	0.898	0.851
acq3	0.869	0.772	0.890	0.735
acq4	1.046	1.175	1.147	0.983
acq5	0.765	0.952	0.868	0.740
acq6	0.851	0.770	0.901	0.718

Table 3
Multiple-group CFA results: chi-square difference test statistics and CFI

	Chi-square	(df)/p value	Chi-square difference	(df)/p value	CFI/TLI	RMSEA (95 % CI)	SRMR
Age (7–10 vs 11 +)							
M0	18.2	(14)/0.20			0.994/0.998	0.036 (0.000–0.078)	0.023
M1	23.7	(20)/0.25	5.5	(6)/0.48	0.995/0.992	0.029 (0.000–0.067)	0.036
M2	36.3	(26)/0.09	12.6	(6)/0.05	0.986/0.984	0.042 (0.000–0.072)	0.041
Age (7–9 vs 10 +)							
M0	15.0	(15)/0.45			1.000/1.000	0.002 (0.000–0.063)	0.021
M1	30.9	(21)/0.07	15.9	(6)/0.01	0.984/0.978	0.046 (0.000–0.078)	0.048
M2	55.4	(27)/0.00	24.5	(6)/0.00	0.955/0.950	0.068 (0.042–0.094)	0.063
Gender group							
M0	13.7	(13)/0.39			0.999/0.998	0.015 (0.000–0.069)	0.021
M1	25.0	(19)/0.16	11.3	(6)/0.08	0.992/0.987	0.037 (0.000–0.073)	0.043
M2	29.3	(25)/0.25	4.3	(6)/0.64	0.994/0.993	0.027 (0.000–0.062)	0.050

The following correlations among residuals are estimated in the models for the different groups: r_{acq3} , $acq5$ and r_{acq1} , $acq2$ in age 11–12 group, r_{acq4} , $acq3$ in age 7–10 group, r_{acq4} , $acq3$ and r_{acq4} , $acq5$ in age 10–12 group, and r_{acq2} , $acq1$ in age 7–9 group; r_{acq6} , $acq3$ in the boys group, r_{acq2} , $acq1$, r_{acq2} , $acq4$, r_{acq3} , $acq1$, and r_{acq3} , $acq4$ in the girls group. We use modification index value of 5 or greater to aid our decision for including these correlations into the models

CFI/TLI comparative fit index/Tucker–Lewis fit index, RMSEA (95 % CI) root mean squared error of approximation (95 % confidence interval), SRMR standardized root mean squared residual