International Epidemiology of Child and Adolescent Psychopathology II: Integration and Applications of Dimensional Findings From 44 Societies

Leslie Rescorla, Ph.D., Masha Y. Ivanova, Ph.D., Thomas M. Achenbach, Ph.D., Ivan Begovac, M.D., Ph.D., Myriam Chahed, Ph.D., May Britt Drugli, Ph.D., Deisy Ribas Emerich, B.A., Daniel S.S. Fung, M.D., Mariam Haider, M.S., Kjell Hansson, Ph.D., Nohelia Hewitt, M.Sc., Stefanny Jaimes, Ps., Bo Larsson, M.D., Ph.D., Alfio Maggiolini, Ph.D., Jasminka Marković, M.D., Ph.D., Dragan Mitrović, Ph.D., Paulo Moreira, Ph.D., João Tiago Oliveira, M.S., Martin Olsson, Ph.D., Yoon Phaik Ooi, Ph.D., Djaouida Petot, Ph.D., Cecilia Pisa, M.D., Rolando Pomalima, M.D., Marina Monzani da Rocha, Ph.D., Vlasta Rudan, M.D., Ph.D., Slobodan Sekulić, Ph.D., Mimoza Shahini, M.D., Ph.D., Edwiges Ferreira de Mattos Silvares, Ph.D., Lajos Szirovicza, Ph.D., José Valverde, M.D., Luis Anderssen Vera, M.Sc., Maria Clara Villa, Ps., Laura Viola, M.D., Bernardine S.C. Woo, M.Med., Eugene Yuqing Zhang, Ph.D.

Objective: To build on Achenbach, Rescorla, and Ivanova (2012) by (a) reporting new international findings for parent, teacher, and self-ratings on the Child Behavior Checklist, Youth Self-Report, and Teacher's Report Form; (b) testing the fit of syndrome models to new data from 17 societies, including previously underrepresented regions; (c) testing effects of society, gender, and age in 44 societies by integrating new and previous data; (d) testing cross-society correlations between mean item ratings; (e) describing the construction of multisociety norms; (f) illustrating clinical applications. **Method:** Confirmatory factor analyses (CFAs) of parent, teacher, and self-ratings, performed separately for each society; tests of societal, gender, and age effects on dimensional syndrome scales, *DSM*-oriented scales, Internalizing, Externalizing, and Total Problems scales; tests of agreement between low, medium, and high ratings of problem items across societies. **Results:** CFAs supported the tested syndrome models in all societies according to the primary fit index (Root Mean Square Error of Approximation [RMSEA]), but less consistently according to other indices; effect sizes were small-to-medium for societal differences in scale scores, but very small for gender, age, and interactions with society; items received similarly low, medium, or high ratings in different societies; problem scores from 44 societies fit three sets of multisociety norms. **Conclusions:** Statistically derived syndrome models fit parent, teacher, and self-ratings when tested individually in all 44 societies according to RMSEAs (but less consistently according to other indices). Small to medium differences in scale scores among societies supported the use of low-, medium-, and high-scoring norms in clinical assessment of individual children. J. Am. Acad. Child Adolesc. Psychiatry; 2012; 51(12):1273-1283. **Key Words:** international, psychopathology, epidemiology, syndromes, confirmatory factor analysis (CFA).

This article is discussed in an editorial by Drs. Luis Augusto Rohde and Christian Kieling on page 1236.

Clinical guidance is available at the end of this article.

An interview with the author is available by podcast at www.jaacap.org or by scanning the QR code to the right.

Supplemental material cited in this article is available online.

Whereas Achenbach *et al.*[1] reviewed international findings from diagnostic interviews and dimensional rating scales published by other investigators, we present findings from new statistical analyses of international data obtained with the Child Behavior Checklist for Ages 6–18 (CBCL/6–18, hereafter CBCL), Youth Self-Report (YSR), and Teacher's Report Form (TRF).[2] These

dimensional instruments are scored on eight statistically derived syndrome scales, three broad-band scales (Internalizing, Externalizing, and Total Problems), and six *DSM*-oriented scales. Previous international comparisons of CBCLs from 31 societies, YSRs from 24 societies, and TRFs from 21 societies[3-8] with respect to factor structure, mean problem scale scores, age and gender patterns, and mean item ratings revealed considerable consistency but also some significant differences across societies.

The previous studies[3-8] were *etic*[9] in nature, meaning that the same standardized instruments were used in different societies. Although these studies identified many similarities across societies, they also identified some important differences, particularly with regard to overall levels of scores. These societal differences in score levels argued for different norms for low-, medium-, and high-scoring societies.[10]

The present article builds on the work of Achenbach *et al.*[1] by presenting new international findings from 44 societies derived from integrating previously analyzed data [3–8] with new data from North Africa, Asia, South America, and Europe. Data from 103 samples were analyzed (42 CBCL, 34 YSR, 27 TRF; see Table S1 and Supplement 1, available online).

PURPOSES

Our purposes were as follows: (a) to test how well the CBCL, YSR, and TRF syndrome models fit the data from 27 samples not included in previous CFA studies[3–5]; to test effects of society, age, and gender on CBCL, YSR, and TRF scores by integrating the 27 new samples with previous samples[6–8]; to test whether the same items received low, medium, or high ratings in different societies; to describe construction of multisociety norms; and to illustrate clinical applications.

METHOD

Samples

Table 1[11-20] describes the 27 samples used in the new CFAs (11 CBCL, 10 YSR, 6 TRF). For the international comparisons of scale scores and mean item ratings, we combined data from the 27 samples used for the CFAs with data from the samples used in previous international comparisons,[6–8] yielding CBCL samples of 69,866 children and adolescents aged 6 through 16 years from 42 societies, YSR samples of 38,070 youths aged 11 through 16 years from 34 societies, and TRF samples of 37,244 students aged 6 through 15 years from 27 societies. Conventions for obtaining informed consent required by each investigator's institution were followed.

Instruments

For non-Anglophone societies, a multistep process was used to create translations. This process typically included translation of the form into the foreign language, blind back-translation into English, review of the back-translation, and revision where the back-translation revealed problems with the translation. Translations were pilot tested with the intended kinds of informants and revised as needed. Each problem item was rated 0 = not true (as far as you know), 1 = somewhat or sometimes true, and 2 = very true or often true, based on the preceding 6 months (2 months for the TRF). Because the pre-2001 editions[21] were used for 31 CBCL samples, 20 YSR samples, and 18 TRF samples, the six items replaced on the CBCL and YSR (items 2, 4, 5, 28, 78, and 99) and the three replaced on the TRF (5, 28, 99) in 2001[2] were omitted from all analyses. The 2001 version of each scale was scored by summing the 0-1-2 ratings of the items comprising the scale.

Data Analysis

CFAs. The CBCL and YSR CFAs tested a correlated eight-syndrome model (Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior) (Figures S1 and S2, available online). The TRF CFAs tested a two-model structure[2] comprising a seven-syndrome model that excluded Attention Problems, and a hierarchical three-factor structure comprising the general Attention Problems syndrome plus Inattention and Hyperactivity–Impulsivity subsyndromes (Figure S3, available online). For the CBCL, YSR, and seven-factor TRF models, all items were assigned to only one syndrome. However, for the hierarchical Attention Problems TRF model, each item was assigned to both the general syndrome and one of the subsyndromes, with factor covariances set to zero.

Comparisons Across Societies. We combined new and previously analyzed samples in analyses of variance (ANOVAs) testing effects of society, gender, and age group on the 17 scales listed in Table 3. The complete samples (CBCL: N = 69,866, 42 societies; YSR: N = 38,070, 34 societies; TRF: N = 37,245, 27 societies) were used to compute the societal means for each scale as well as the mean item ratings for each sample. Because not all samples met our criterion of 80 or more participants in each age group tested, the CBCL ANOVAs included 38 societies for ages 6 to 11 years (n = 39,937), 33 societies for ages 12 to 16 years (n = 29,536), and 29