International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 19(Supplement 1): 1–3 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.316

Editorial

Screening for serious mental illness: methodological studies of the K6 screening scale

HANS-ULRICH WITTCHEN

Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany

Correspondence
Hans-Ulrich Wittchen, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany.
Telephone (+49) 351-463-36983
Fax (+49) 351-463-36984
Email: wittchen@psychologie.tu-dresden.de

Abstract

The K6 scale is a shortened version of the K10, a 10-question scale originally developed to provide an efficient population-level screen for serious mental disorders (SMI) in the USA. Evidence that the six-item shortened version performed as well as the original 10-item version, coupled with strong psychometric properties, led to rapid dissemination and replicated validation of the K6 in a number of other countries around the world. Based on these results, the K6 is now often included in large general-purpose government health tracking surveys in a number of different countries. Until now, though, the scoring rules for the K6 in these surveys were inconsistent. The first paper in this special issue introduces the K6 scale and summarizes the results of a series of investigations to resolve these inconsistencies by providing optimal scoring rules for the K6 in 14 countries. Subsequent papers explore the usefulness of the K6 to screen for serious emotional disturbance among adolescents and report findings from validation efforts based on independent diagnostic assessments as well as of other measures of impairment and disability (World Health organization Disability Assessment Scale). Finally a highly innovative analysis using a Bayesian multilevel modeling approach is presented, designed to estimate the prevalence of SMI in small areas, such as cities, states, or schools, from surveys carried out in a larger population that includes only relatively small samples of respondents in each of the areas in which prevalence estimates are to be made. Taken together, these studies demonstrate that the K6 is an efficient and useful screening tool. The psychometric and methodological explorations will hopefully stimulate additional interest in the use of short screening scales in large-scale general health surveys to supplement the more in-depth information obtained in periodic psychiatric epidemiological surveys on the basis of diagnostic interviews.

The six papers in this special issue present new and important data on the performance of a widely-used six-question screening scale for serious mental illness (SMI) developed by Kessler and his associates (Kessler *et al.*, 2002, 2003). This K6 scale is a shortened version of the K10, a 10-question scale originally developed to provide an efficient population-level screen for SMI in the USA. However, evidence that the six-item shortened version

performed as well as the original 10-item version, coupled with strong psychometric properties, led to rapid dissemination and replicated validation of the K6 in a number of other countries around the world. Based on these results, the K6 is now often included in large general-purpose government health tracking surveys in a number of different countries.

Until now, though, the scoring rules for the K6 in these surveys were inconsistent. The first paper by Kessler *et al.* (2010) helps to resolve this inconsistency by providing optimal scoring rules for the K6 in 14 countries based on analysis of data from epidemiological surveys carried out in these countries in conjunction with the World Health Organization (WHO) World Mental Health Survey Initiative (WMH). The paper shows that concordance of diagnoses of SMI based on the K6 with independent diagnoses based on in-depth fully structured diagnostic interviews is quite good in all the WMH surveys, providing an empirical justification for using the K6 as a screen for SMI in these countries.

Greif Green *et al.* (2010) in the second paper, present data on the usefulness of the K6 to screen for serious emotional disturbance (SED) among adolescents using data from a large epidemiological survey of adolescent mental health carried out in the USA. As shown there, the K6 is more limited in detecting SED than SMI because behavioral disorders are not included as part of the K6 screening questions. Greif Green *et al.* show, though, that augmentation of the K6 to include several marker items for behavioral disorders succeeds in improving the concordance of the augmented K6 with independent clinical diagnoses of SED among adolescents to a level similar to the level of concordance found between the K6 and independent clinical diagnoses of SMI among adults.

Colpe *et al.* (2010), in the third paper, provide an overview of an ambitious program of research launched by the Substance Abuse and Mental Health Services Administration (SAMHSA) to monitor trends in the prevalence and correlates of SMI over time in conjunction with the ongoing National Survey on Drug Use and Health (NSDUH). The first stage in this Mental Health Surveillance Study (MHSS) was to carry out methodological studies to calibrate screening data collected in the NSDUH, including both the K6 and measures of impairment caused by mental illness, to independent diagnoses of SMI based on clinical reappraisal interviews administered to a probability sample of NSDUH respondents.

Novak *et al.* (2010), in the fourth paper, report results from a preparatory phase of the MHSS study in which the WHO Disability Assessment Schedule (WHODAS; Von Korff *et al.*, 2008), one of the impairment measures considered for permanent inclusion in the NSDUH, was subjected to psychometric analysis to determine if a shortened version of the scale could be developed that retained the precision of the full scale in the range most relevant to defining SMI. As detailed by Novak *et al.* in the paper, sophisticated analyses using item response theory methods were able to produce an abbreviated WHODAS scale that retained virtually all the systematic variation found in the full scale, although meaningful variation in concordance was found across important socio-demographic segments of the population, especially those defined on the basis of gender and race/ethnicity.

Aldworth *et al.* (2010), in the fifth paper, report the results of the MHSS calibration study, which found that the reduced WHODAS scale developed by Novak *et al.* (2010) was, in fact, the optimal impairment scale to combine with the K6 to assess SMI in the NSDUH. Aldworth *et al.* did this by developing a scoring rule that combined scores on the K6 with scores on the reduced WHODAS to generate a dichotomous classification of the presence–absence of SMI for each respondent in the NSDUH clinical calibration sub-sample. A complexity in doing this was that different scoring rules were found to be optimal in different segments of the population. As noted by Kessler *et al.* (2010) in their earlier paper, though, an alternative approach is to use the results of a logistic regression equation in which screening scale scores, socio-demographics, and significant interactions between screening scale scores and socio-demographics are all used to predict SMI in a clinical reappraisal sub-sample to impute continuous predicted-probability-of-SMI scores to all survey respondents and to use those continuous scores to study time–space variation in the prevalence and correlates of SMI. As noted by Kessler *et al.*, this continuous scoring approach yields stronger concordance with independent clinical diagnoses of SMI than does the kind of dichotomous approach used by Aldworth *et al.* This is an encouraging observation because it means that that the good estimates of concordance reported by Aldworth *et al.* should be thought of as lower boundary estimates of the true concordance of the NSDUH screen with clinical diagnoses of SMI.

Li *et al.* (2010), in the final paper in the issue, present the results of a highly innovative analysis using a Bayesian multilevel modeling approach designed to estimate the prevalence of SMI in small areas, such as cities, states, or schools, from surveys carried out in a larger population that includes only relatively small samples of respondents in each of the areas in which prevalence estimates are to be made. This innovative approach to small-area estima-

tion is illustrated by Li *et al.* by estimating school-level prevalence of SED in the schools that participated in the large US epidemiological survey of adolescent mental health analysed by Greif Green *et al.* (2010). The methodology is much more broadly applicable, though, to the work of epidemiologist and mental health policy planners who want to use the results of epidemiological surveys to estimate the prevalence of SMI or SED to any small area. In the case of the work reported earlier by Colpe and her associates in the NSDUH, for example, it is of interest to estimate the prevalence of SMI in each of the counties and states that participated in the survey. Li *et al.* show that multilevel modeling can improve the accuracy of such estimates by using an extension of the individual-level clinical calibration approach featured in the earlier papers to include information about distributions of K6 scores in each small area to improve the accuracy of estimates of area-level SMI prevalence. Two remarkable results reported by Li *et al.* are especially noteworthy in this regard: first, that the association between aggregate scores on the K6 scale with prevalence of SMI in small areas using this approach is a good deal higher than the association between individual-level K6 scores and individual-level diagnoses of SMI; second, that optimal precision of small-area estimates using this approach can be obtained in large-scale surveys with within-area samples of no more than 200 respondents.

We hope that by bringing these important methodological studies of the K6 to the attention of the readers of this journal we will be able to stimulate additional interest in the use of short screening scales in large-scale general health surveys to supplement the more in-depth information obtained in periodic psychiatric epidemiological surveys. Certainly we would not expect short screening scales to provide the same kind of textured information that we obtain using in-depth epidemiological diagnostic interviews like the WHO Composite International Diagnostic Interview (CIDI), but neither is it realistic to think that expensive and lengthy CIDI surveys can be carried out on an ongoing basis in large-scale national tracking samples. Screening scales need to be thought of as a bridge between general-purpose health tracking surveys and in-depth clinical diagnostic surveys. The new and important results regarding measurement and statistical analysis of the K6 presented in this special issue help to strengthen that bridge in ways that we hope the journal readers will find useful.

Declaration of interests

The author has no competing interests.

References

Aldworth J., Colpe L.J., Gfroerer J.C., Novak S.P., Chromy J.R., Barker P.R., Barnett-Walker K., Karg R.S., Morton K.B., Spagnola K. (2010). The National Survey on Drug Use and Health's Mental Health Surveillance Study: calibration analysis. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 61–7.

Colpe L.J., Barker P.R., Karg R.S., Batts K.R., Morton K.B., Gfroerer J.C., Stolzenberg S.J., Cunningham D.B., First M.B., Aldworth J. (2010). The National Survey on Drug Use and Health Mental Health Surveillance Study: calibration study design and field procedures. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 36–48.

Greif Green J., Gruber M.J., Sampson N.A., Zaslavsky A.M., Kessler R.C. (2010). Improving the K6 short scale to predict serious emotional disturbance in adolescents in the USA. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 23–35.

Kessler R.C., Andrews G., Colpe L.J., Hiripi E., Mroczek D.K., Normand S.L., Walters E.E., Zaslavsky A.M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med*, **32**, 959–976.

Kessler R.C., Barker P.R., Colpe L.J., Epstein J.F., Gfroerer J.C., Hiripi E., Howes M.J., Normand S.L., Manderscheid R.W., Walters E.E., Zaslavsky A.M. (2003). Screening for serious mental illness in the general population. *Arch Gen Psychiatry*, **60**, 184–189.

Kessler R.C., Greif Green J., Gruber M.J., Sampson N.A., Bromet E., Cuitan M., Furukawa T.A., Gureje O., Hinkov H., Hu C., Lara C., Lee S., Mneimneh Z., Myer L., Oakley-Browne M., Posada-Villa J., Sagar R., Viana M.C., Zaslavsky A.M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) Survey Initiative. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 4–22.

Li F., Greif Green J., Kessler R.C., Zaslavsky A.M. (2010). Estimating prevalence of serious emotional disturbance in schools using a brief screening scale. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 88–98.

Novak S.P., Colpe L.J., Barker P.R., Gfroerer J.C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 49–60.

Von Korff M., Crane P.K., Alonso J., Vilagut G., Angermeyer M.C., Bruffaerts R., de Girolamo G., Gureje O., de Graaf R., Huang Y., Iwata N., Karam E.G., Kovess V., Lara C., Levinson D., Posada-Villa J., Scott K.M., Ormel J. (2008). Modified WHODAS-II provides valid measure of global disability but filter items increased skewness. *J Clin Epidemiol*, **61**, 1132–1143.