ORIGINAL ARTICLE

WILEY

Psychometric properties of the Kessler psychological scales in a Swiss young-adult community sample indicate poor suitability for community screening for mental disorders

Naweed Osman[1] | Winnie S. Chow[1] | Chantal Michel[2] | Eva Meisenzahl[1] |
Benno G. Schimmelmann[2,3] | Frauke Schultze-Lutter[1,2,4]

[1]Department of Psychiatry and
Psychotherapy, Medical Faculty, Heinrich-
Heine-University, Düsseldorf, Germany

[2]University Hospital of Child and Adolescent
Psychiatry and Psychotherapy, University of
Bern, Bern, Switzerland

[3]University Hospital of Child and Adolescent
Psychiatry, University Hospital Hamburg-
Eppendorf, Hamburg, Germany

[4]Department of Psychology, Faculty of
Psychology, Airlangga University, Surabaya,
Indonesia

Correspondence
Frauke Schultze-Lutter, Department of
Psychiatry and Psychotherapy, Medical
Faculty, Heinrich-Heine-University, Bergische
Landstraße 2, 40629 Düsseldorf, Germany.
Email: frauke.schultze-lutter@lvr.de

Abstract

Aim: The Kessler psychological distress scales (K10 and K6) are used as screening tools to assess psychological distress related to the likely presence of a mental disorder. Thus, we studied the psychometric properties of their German versions in a Swiss community sample to evaluate their potential usefulness to screen for mental disorders or relevant mental problems in the community and, relatedly, for low threshold transdiagnostic German-speaking services.

Methods: The sample consisted of 829 citizens of the Swiss canton Bern of age 19–43 years. K10/K6 were validated against mini-international neuropsychiatric interview (M.I.N.I.) diagnoses, questionnaires about health status and quality of life. Receiver operating characteristic (ROC) curve analyses were used to test for general discriminative ability and to select optimal cut-offs of the K10 and K6 for non-psychotic full-blown and subthreshold mental disorders.

Results: Cronbach's alphas were 0.81 (K10) and 0.70 (K6). ROC analyses indicated much lower optimal thresholds than earlier suggested; 10 for K10 and 6 for K6. At these thresholds, against M.I.N.I. diagnoses, Cohen's κ (≤0.173) and correspondence rates (≤58.14%) were insufficient throughout. Values were higher at the earlier suggested threshold, yet, at the cost of sensitivity that was below 0.5 in all but three, and below 0.3 in all but six cases.

Conclusions: For the lack of sufficient validity and sensitivity, respectively, our findings suggest that both K10 and K6 would only be of limited use in a low-threshold transdiagnostic mental health service for young adults in Switzerland and likely other German-speaking countries.

KEYWORDS
concurrent validity, general population, mental disorders, psychological distress, screening

1 | INTRODUCTION

Mental disorders are a main source of illness-related costs and burden, affecting 38% of the European population (Trautmann et al., 2016). Frequently, help for mental disorders is only sought when they have become severe and functionally disabling (Michel et al., 2018); this contributing to high rates of negative outcomes and high burden (Boerema et al., 2017; Trautmann et al., 2016). Consequently, from a clinical and public health perspective, an earlier detection and treatment of mental disorders is imperative (Trautmann et al., 2016). Therefore, valid, economical and easy-to-use screenings, which appropriately identify relevant mental problems/disorders in the community and primary health care, are of great public health importance (Michel, Schultze-Lutter, et al., 2014; Trautmann et al., 2016; Webb et al., 2016).

Today, the 10-item Kessler psychological distress scale (K10; Kessler et al., 2003) and its six-item short version (K6; Kessler et al., 2002) are frequently used for such a clinical screening as well as for monitoring of outcomes in primary mental health services (Cotton et al., 2021). In epidemiological research, K10/K6 are used to assess mental disorder caseness and non-specific psychological distress (Andrews & Slade, 2001; Cotton et al., 2021; Kessler et al., 2010). Since K10 was first developed with focus on disorders of the anxiety-depression spectrum, items mainly focus on signs of these (Kessler et al., 2003).

Good psychometric properties of K10/K6, including good validity for mental health problems/disorders, have been reported from several countries (Ferro, 2019; Batterham et al., 2016; Sampasa-Kanyinga et al., 2018; Stolk et al., 2014; Thelin et al., 2017). However, validity was mostly assessed against self-report questionnaires or fully-standardized lay-interviewer assessments and only rarely against clinical interviews conducted by mental health professionals (Furukawa et al., 2003; Sampasa-Kanyinga et al., 2018; Searle et al., 2017; Sunderland et al., 2011). Furthermore, despite their good research performance in epidemiological studies, inconsistent evidence for K6's/K10's cultural appropriateness in clinical settings, and a lack of clinical norms for different countries indicate the importance of further research into its use in clinical settings (Shon, 2020; Stolk et al., 2014). With regard to German-speaking countries, to the best of our knowledge, only one Austrian study has evaluated the psychometric properties of the German translation of K10 against the Brief Symptoms Inventory and State–Trait-Anxiety Inventory in a sample of psychotherapeutic outpatients and medical students (Giesinger et al., 2008). It found K10 to be a suitable measurement of unspecific psychological distress in clinical settings (Giesinger et al., 2008). A validation of the German K10/K6 against the gold-standard of clinician-assessed mental disorders in a less selected community sample, however, is still lacking.

Thus, we examined the psychometric properties of the German K10/K6 against clinician-assessed mental disorders in a young-adult community sample, thereby paying attention to the differential validity for different diagnostic categories. Given the scales' focus on depression and anxiety, we expected the best performance for depressive and anxiety disorders and problems.

2 | METHODS

2.1 | Study design and procedure

The sample consisted of 839 adults of age 19–45 years with main residency in the Swiss Canton Bern, who were assessed between 06/2015 and 03/2018 as part of the second wave (response rate: 66.4%) of the random-sampling Bern epidemiological at-risk (BEAR) community study and, at follow-up, oversampled for lifetime mental problems in terms of symptoms related to a clinical high risk of psychosis (Schultze-Lutter et al., 2018, 2021). Eligibility criteria included participation in the first wave (response rate: 63.4%) and agreement to be re-contacted given at baseline (provided by 97.9%) (Appendix S1). Five respondents had no data on K10/6 due to study-conform early termination of the interview because of development of a psychotic disorder (Schultze-Lutter et al., 2021), and another five broke off the assessment. Thus, complete data sets of 829 cases were analysed (Table 1).

All participants provided informed verbal consent on the phone. The Ethics Committee of the University of Bern had approved the study.

2.2 | Assessments

Items of K10/K6 are answered on a five-point Likert scale (Table S1), their total scores ranging from 10–50 and 6–30, resp. (Andrews & Slade, 2001; Kessler et al., 2002). The interpretation of total scores varies with the purpose of administration and the setting as well as between studies and cultures, with K10 ≥ 20 frequently being used as a threshold for a likely mental disorder (Table S2). In an extended version of K10, K10+ (Australian Government Department of Health, 2018), four add-on questions assess functioning and related factors (Table S1). We used the validated German translation by Giesinger et al. (2008).

Mental disorders were assessed by the mini international neuropsychiatric interview (M.I.N.I.), a brief semi-structured interview to reliably and validly assess mental disorders according to DSM-IV and ICD-10 (Lecrubier et al., 1997; Sheehan et al., 1997, 1998). The M.I.N.I. uses a two-step procedure: (1) screening questions and (2) full interview of disorders with affirmed screening question. Presence of any subthreshold mental problem that signals a need of professional assessment and, consequently, help-seeking was assumed when a screening question was affirmed (Alexander et al., 2008). For the recently reported low clinical relevance of specific phobia in the community when not accompanied by another mental disorder (Sancassiani et al., 2019), specific phobia, which was the most frequent mental disorder in our sample (11.8%), was not considered in the analyses.

Psychosocial functioning was estimated using the social and occupational functioning assessment scale (SOFAS; American Psychiatric Association, 1994) that has good psychometric properties incl. Good interrater-reliability and construct validity (Hilsenroth et al., 2000; Rybarczyk, 2011). SOFAS scores range from 0 (poor) to

TABLE 1 Sociodemographic and clinical characteristics of participants with and without a K10 score ≥ 20

	Total sample (N = 829)	K10 < 20 (n = 787)	K10 ≥ 20 (n = 42)	Statistics (χ^2 and U)
Age in years: M ± SD, median (range)	33.25 ± 7.77, 35 (19–45)	33.32 ± 7.73, 35 (19–45)	31.90 ± 8.58, 31.5 (19–44)	$U = 15\,151.0$, $p = .362$
Female sex: n (%)	441 (53.2%)	412 (52.4%)	29 (69.0%)	$\chi^2_{(1)} = 4.465$, $p = .035$
At least upper secondary education: n (%)	770 (92.9%)	732 (93.0%)	38 (90.5%)	$\chi^2_{(1)} = 0.388$, $p = .534$
Currently employed/in training in normal, non-sheltered setting: n (%)	797 (96.1%)	764 (97.1%)	33 (78.6%)	$\chi^2_{(1)} = 36.796$, $p < .001$
Any lifetime help-seeking for mental problems: n (%)	274 (33.1%)	242 (30.7%)	32 (76.2%)	$\chi^2_{(1)} = 37.207$, $p < .001$
Any non-psychotic M.I.N.I. mental disorder: n (%)	43 (5.2%)	24 (3.0%)	19 (45.2%)	$\chi^2_{(1)} = 144.303$, $p < .001$
Any non-psychotic M.I.N.I. mental problem: n (%)	268 (32.3%)	238 (30.2%)	30 (71.4%)	$\chi^2_{(1)} = 30.917$, $p < .001$
SOFAS score: mean ± SD, median (range)	84.43 ± 7.44, 88 (35–100)	85.19 ± 6.04, 88 (35–100)	70.29 ± 14.24, 71.5 (35–90)	$U = 5529.5$, $p < .001$
SOFAS < 71: n (%)	58 (7.0%)	37 (4.7%)	21 (50.0%)	$\chi^2_{(1)} = 125.738$, $p < .001$
EQ-5D score: mean ± SD, median (range)	5.35 ± 0.82, 5 (5–13)	5.31 ± 0.75, 5 (5–13)	6.19 ± 1.44, 6 (5–11)	$U = 23\,112.5$, $p < .001$
Subjective estimate of overall health: mean ± SD, median (range)	83.80 ± 12.88, 85 (10–100)	84.56 ± 11.82, 85 (20–100)	69.45 ± 21.23, 77.5 (10–95)	$U = 8486.5$, $p < .001$
K10 score: mean ± SD, median (range)	12.53 ± 3.73, 11 (10–43)	11.90 ± 2.37, 11 (10–19)	24.33 ± 4.71, 23.5 (20–43)	$U = 33\,054.0$, $p < .001$
K6 score: mean ± SD, median (range)	7.58 ± 2.37, 7 (6–27)	7.20 ± 1.59, 6 (6–14)	14.69 ± 3.11, 14.5 (11–27)	$U = 32\,830.5$, $p < .001$
K6 ≥ 19: n (%)	4 (0.5%)	0 (0.0%)	4 (9.5%)	$\chi^2_{(1)} = 75.316$, $p < .001$

Abbreviations: M.I.N.I., mini-international neuropsychiatric interview; SOFAS, social and occupational functioning assessment scale.

100 (excellent). The EQ-5D-3L questionnaire subjectively measures general health status (EuroQol Group, 1990). It consists of a descriptive (Table S3) and an evaluation part on which respondents rate their current overall health status on a scale from 0 = 'worst possible health' to 100 = 'best possible health'. Studies of the validity and reliability of the EQ-5D-3L reported good psychometric properties (Peasgood et al., 2012; Rabin & de Charro, 2001; Szende et al., 2007).

The brief multidimensional life satisfaction scale (BMLSS; Büssing et al., 2009) is a 10-item self-rating instrument for subjective satisfaction in five main and 10 single domains of life: intrinsic (myself, overall life), social (family life, friendships), external (work, residence area), perspective (financial situation, future prospects), and health (health situation, own abilities to cope with daily life). Each item is rated on a seven-point Likert-scale from 0 = 'terrible' to 6 = 'delighted'.

The questions of K10+, EQ-5D-3L and BMLSS were read to the interviewees, and answers were scored without answering follow-up questions or offering explanations, in order to stay close to their original self-rating format. Trained clinical psychologists carried out all assessments (Schultze-Lutter et al., 2018). Weekly supervision of all clinical interviews (including M.I.N.I.) was provided by CM and FSL to ensure high data quality.

2.3 | Statistical analysis

Persons with and without a likely mental disorder according to K10 were compared using χ^2 tests for categorical and Mann–Whitney U test for rank data (K10/K6, SOFAS, EQ-5D-3L, BMLSS) and non-normally distributed ratio data (age).

Internal consistency was examined by Cronbach's alpha. Convergent validity was tested against M.I.N.I., and discriminant (or divergent) validity against BMLSS domains (except health domain) and sum score of the four somatic items of the EQ-5D-3L. Validity was assessed by Cohen's κ correlations (κ) for dichotomized variables, and intra-class correlation coefficients (ICC) for continuous variables using the one-way random-effects model from single measurement (ICC1,1) (Koo & Li, 2016). Prevalence indices (PI) were also calculated as κ tends to be underestimated in case of low- or high-prevalence outcomes; in which case correspondence rates (CR) give a better estimation (Burn & Weir, 2011). The global diagnostic accuracy of K10/K6 was examined by receiver-operating characteristic (ROC) analyses, whose areas under the curve (AUCs) were used to select optimal cut-offs. Thereby, emphasis was put on high sensitivity (≥70%) as the most important diagnostic feature of a screener (Michel, Schultze-

Lutter, et al., 2014) while keeping specificity as high as possible. Additionally, positive and negative likelihood ratios (LR+, LR−) were calculated as conjoint estimations of sensitivity and specificity, i.e., of a test's ability to rule in or rule out a disorder (Jaeschke et al., 1994). Table S4 reports evaluation guidelines for AUC, ICC, κ and LR+ and LR−.

Throughout, we used SPSS version 22.0. The level of statistical significance was set at $p < .05$. Disorders with 0%-prevalence were excluded from the analyses.

3 | RESULTS

3.1 | Descriptive statistics

Mean K10/K6 scores were generally low (Figure 1), with only $n = 42$ (5.1%) of interviewees scoring ≥20 on K10 and only $n = 4$ (0.5%) scoring ≥19 on K6, in doing so, indicating some mental problems according to available interpretation guidelines (Table S1). While the prevalence of K10 ≥ 20 was roughly in line with the prevalence of functional impairment ($n = 58$; 7.0%) and non-psychotic mental disorders ($n = 43$; 5.2%); the prevalence of mental problems in terms of an affirmed M.I.N.I. screening question ($n = 268$; 32.3%) was clearly higher (Table 1). Persons with/without K10 ≥ 20 did not differ in age and educational level but in all other sociodemographic and clinical variables (Table 1).

3.2 | Internal consistency

Cronbach's alpha was 0.81 for K10 and 0.70 for K6, indicating good and satisfactory internal consistency, respectively (Bland & Altman, 1997).

3.3 | General discriminative ability and optimal cut-offs

The general ability of K10/K6 to discriminate between individuals with and without mental problems was insufficient (AUC = 0.650) but excellent (AUC = 0.822) for mental disorders (Table 2). Mood disorders/problems exhibited acceptable AUCs (0.782; 0.774); anxiety disorders had an excellent AUC (0.869); while AUCs for anxiety problems were insufficient (Table 2). AUCs were highest for the infrequent somatoform (0.892) and eating disorders (0.991) (Table 2).

Mostly, the optimal cut-off points to rate mental disorders and problems were >10 for K10 ($n = 486$, 58.6%) and >6 for K6 ($n = 429$, 51.7%) (Table 2, Table S5). At these thresholds, most sensitivities were between 60% and 90% with specificities >40% (Table 2, Table S5). Positive LRs indicated at most 'small and rarely important increases in pre-test-probability' of mental disorder and problems (1.03 ≤ LR+ ≤ 1.71) with some exceptions in rare disorders, while negative LRs generally performed better, sometimes even decreasing the pre-test probability to a large and often conclusive degree
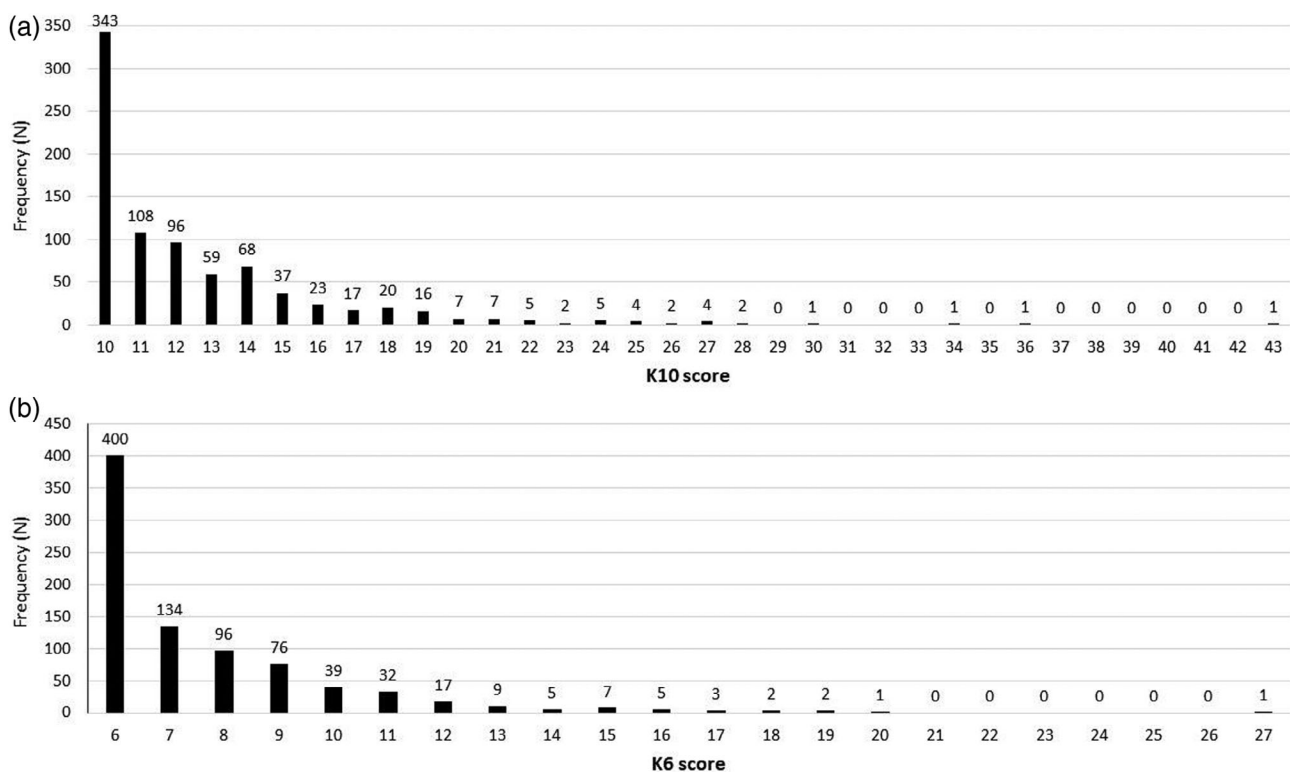
FIGURE 1 Distribution of (a) K10 scores and (b) K6 scores in the sample ($N = 829$). High scores never reached in the samples are not given

(LR− < 0.1) (Table 2). Overall, the K10 performed slightly better than the K6 (Table 2, Table S5).

3.4 | Convergent validity

Testing the accuracy of the K10 and K6 cut-offs suggested by Andrews and Slade (2001) and Kessler et al. (2003) (Table 3) against the M.I.N.I. results indicated poor convergent validity (Table 3) as most κ indicated no to minimal agreement (Table S4). This was against the background of unfavourable PIs. Thus, expectantly, ranging from 52.35% to 99.40%, CRs indicated better convergent validity. However, in particular for K6, sensitivities were small, rarely >40% (Table 3).

Using the lower cut-offs of the ROC analyses (Table 4), the convergent validity decreased, with values of CRs of 41.50–58.14%, and of κ of <0.001–0.177 (Table 4).

To assess the convergent validity of the K10+'s first additional question on the frequency of affirmed problems, responses were dichotomized: answers indicating lesser or similar presence of these feelings as well as absence of any such feelings were coded '0' (n = 660, 79.6%), and answers indicating an increase in these feelings were coded '1' (n = 169, 20.4%). κ and CR were 0.208 and 81.42% for any mental disorder, and 0.265 and 70.93% for any mental problem. Thus, the first additional question of the K10+ had a LR+ of 3.79 and a LR− of 0.40 for non-psychotic mental disorders (sensitivity = 0.67; specificity = 0.82), and a LR+ of 2.88 and LR− of 0.73 for non-psychotic mental problems (sensitivity = 0.37; specificity = 0.87).

3.5 | Discriminant validity

ICCs between the sum of the first four items of EQ-5D-3 L on somatic problems and K10/K6 sum scores were − 0.678 (p = 1.0) and − 0.407 (p = 1.0), respectively. ICCs between the sum scores of K10/K6, and the sum score of the four non-health-related BMLSS domains were −0.942 (p = 1.0) for K10 and −0.964 (p = 1.0) for K6. These results indicated poor discriminant validity of both Kessler scales.

4 | DISCUSSION

To enhance the early transdiagnostic detection of mental disorders at community and primary care level, easy-to-use screenings of good psychometric properties are required (Michel, Schultze-Lutter, et al., 2014; Trautmann et al., 2016; Webb et al., 2016). One suggested screener is the originally English K10 and its short-version, K6 (Cotton et al., 2021). Yet, validity of K10/K6 was reported to strongly depend on subtleties of wording and, thus, to vary between languages and countries (Stolk et al., 2014). Therefore, using clinician-assessment of the M.I.N.I. as a gold standard assessment for the first time in a German-speaking country, we examined the diagnostic validity of the German translation of K10/K6 for non-psychotic mental

disorders/problems. Overall, both scales showed poor discriminant validity, and failed to either sufficiently rule in or to sufficiently rule out mental disorders and relevant mental problems; this depending on the chosen threshold. Since their validity might be higher in clinical samples, i.e., among persons seeking help for mental problems, we re-ran the analyses separately for participants with and without lifetime help-seeking for mental disorders at points-of-contact other than family and friends (see Tables S6–S8). Although, agreement was slightly better in help-seekers, in which the K10 cut-off suggested by Andrews and Slade (2001) demonstrated weak agreement between the K10 at for mental disorders (κ = 0.483), mood problems (κ = 0.465) and anxiety disorders (κ = 0.456), sensitivities for these were still poor (≤57%). Thus, our results discourage from using K10/K6 to screen transdiagnostically for mental disorders in the community or in community-based low-threshold adult services in German-speaking countries. On the whole, this was also true for disorders/problems related to depression and anxiety for whose assessment K10/K6 had been developed and for that good validity had been reported in earlier studies (Cairney et al., 2007; Ferro, 2019; Furukawa et al., 2003; Kessler et al., 2002, 2003; Sakurai et al., 2011; Sampasa-Kanyinga et al., 2018). Only the additional question of the K10+ on frequency was partly exhibiting acceptable diagnostic accuracy and validity in terms of CR.

Interestingly, our cut-offs for K10/K6 were much lower than the ones earlier generated for K10 by Andrews and Slade (2001), and for K6 by Kessler et al. (2003). This might be due to different rates of mental disorders. Our point-prevalence of 5.2% was lower than the 13.1% prevalence reported by Andrews and Slade (2001) and the 23.1% unweighted 12-months prevalence rate reported by Kessler et al. (2003). Yet, already at the cut-offs suggested by these earlier studies (Andrews & Slade, 2001; Kessler et al., 2003), sensitivities were lower than specificities (K10: 66% vs. 96%, and K6: 36% vs. 96%). These discrepancies between specificity and sensitivity were enlarged in our sample (K10: 40% vs. 98%, and K6: 5% vs. 100%), indicating a poor ability to identify true positive cases by these earlier suggested cut-offs in our sample. Additionally, differences in cut-offs likely resulted from the different emphasis put on sensitivity and specificity. While these were balanced by Andrews and Slade (2001), and Kessler et al. (2003), we had put emphasis on sensitivity, because a screener—generally a first step in diagnosis—should not miss positive cases (i.e., possess excellent sensitivity), while ruling out as many negative cases as possible (Michel, Schultze-Lutter, et al., 2014).

Cultural impacts that lower affirmation on the items of the Kessler scales (Stolk et al., 2014) might be another reason for their poorer performance in our study. Significant variations in the form and symptomatic expression of mood and anxiety disorders, including obsessive compulsive and posttraumatic stress disorders, and somatization disorders were described across cultures (Kirmayer & Ryder, 2016). Thus, the selection of items in US populations assessed in the 1990s (Kessler et al., 2002) and their verbatim translation into German (Giesinger et al., 2008) may have only insufficiently reflected the subjective experiences of distress by mental disorders in the Swiss interviewees 20 years later. In view of other cultural adaptations of

TABLE 2   Diagnostic accuracy of the K10 and K6 at optimal cut-offs in identifying people of the community with any mental disorders, any disorders of a main diagnostic category or any mental problems related to a main diagnostic category (N = 829)

Presence of ...	n (%) of any disorders/ any problem	Kessler psychological distress scale—K10						Kessler psychological distress scale—K6					
		Optimal cut-off	AUC[a]	Sensitivity	Specificity	LR+[b]	LR−[b]	Optimal cut-off	AUC[a]	Sensitivity	Specificity	LR+[b]	LR−[b]
Non-psychotic axis-I disorders	43 (5.2%)	10	0.822***	0.91	0.43	1.59	0.22	6	0.790***	0.84	0.50	1.67	0.33
Non-psychotic axis-I problems	268 (32.3%)	10	0.650***	0.73	0.48	1.41	0.56	6	0.635***	0.65	0.55	1.44	0.63
Mood disorders	20 (2.4%)	10	0.782***	0.85	0.42	1.47	0.36	6	0.734***	0.75	0.49	1.46	0.51
Mood problems	67 (8.1%)	10	0.774***	0.87	0.44	1.54	0.31	6	0.751***	0.81	0.51	1.64	0.38
Anxiety disorders	29 (3.5%)	10	0.869***	0.93	0.43	1.62	0.16	6	0.864***	0.93	0.50	1.85	0.14
Anxiety problems	160 (19.3%)	10	0.663***	0.76	0.45	1.38	0.54	6	0.646***	0.68	0.52	1.42	0.61
Obsessive–compulsive disorder	4 (0.5%)	10	0.709	0.75	0.42	1.28	0.60	6	0.605	0.50	0.48	0.97	1.04
		12	0.709	0.75	0.66	2.22	0.38	11	0.605	0.50	0.94	8.20	0.53
Obsessive–compulsive problems	30 (3.6%)	10	0.705***	0.80	0.42	1.38	0.47	6	0.663**	0.70	0.49	1.37	0.61
Posttraumatic stress disorder	1 (0.1%)	10	0.479	1.00	0.41	1.71	0.00	6	0.563	1.00	0.48	1.93	0.00
Posttraumatic stress problems	25 (3.0%)	10	0.646*	0.80	0.42	1.38	0.48	6	0.602	0.64	0.49	1.25	0.74
Somatoform disorders	5 (0.6%)	10	0.892**	1.00	0.42	1.71	0.00	6	0.901**	1.00	0.49	1.94	0.00
		12	0.892**	1.00	0.66	2.98	0.00	7	0.901**	1.00	0.65	2.84	0.00
Somatoform problems	78 (9.4%)	10	0.622***	0.77	0.43	1.36	0.53	6	0.628***	0.73	0.51	1.48	0.53
Eating disorders	1 (0.1%)	10	0.991	1.00	0.41	1.71	0.00	6	0.992	1.00	0.48	1.93	0.00
		26	0.991	1.00	0.99	90.91	0.00	16	0.992	1.00	0.99	100.00	0.00
Eating problems	26 (3.1%)	10	0.674**	0.81	0.42	1.40	0.46	6	0.679**	0.77	0.49	1.51	0.47
Substance misuse/dependence	19 (2.3%)	10	0.679**	0.79	0.42	1.36	0.50	6	0.669*	0.74	0.49	1.44	0.54

Abbreviations: AUC, areas under the curve; LR, likelihood ratio; ROC, receiver operating characteristic.

*<.05.

**<.01.

***<.001.

[a]AUC, area under the ROC curve.

[b]LR, likelihood ratio; for interpretation of AUC and LRs see Table S4.

TABLE 3 Agreement between K10/K6 scores dichotomized according to the cut-off scores suggested by Andrews and Slade (2001) for K10, and Kessler et al. (2003) for likely presence of a mental disorder of different M.I.N.I. diagnostic categories (N = 829)

Presence of ...	Correspondence rate (%)		Cohen's κ		Prevalence index (PI)		Sensitivity		Specificity		LR+		LR−	
	K10 ≥ 20	K6 ≥ 19	K10 ≥ 20	K6 ≥ 19	K10 ≥ 20	K6 ≥ 19	K10 ≥ 20	K6 ≥ 19	K10 ≥ 20	K6 ≥ 19	K10 ≥ 20	K6 ≥ 19	K10 ≥ 20	K6 ≥ 19
Non-psychotic axis-I disorders	94.33	95.30	0.417	0.163	0.897	0.943	0.40	0.05	0.98	1.00	17.17	0.00	0.62	0.95
Non-psychotic axis-I problems	69.84	68.15	0.116	0.020	0.626	0.672	0.10	0.01	0.98	1.00	6.06	0.00	0.92	0.99
Mood disorder	94.69	97.59	0.266	0.160	0.925	0.971	0.40	0.05	0.97	1.00	12.12	50.00	0.62	0.95
Mood problems	92.16	92.40	0.364	0.105	0.869	0.914	0.28	0.03	0.98	1.00	13.52	0.00	0.73	0.97
Anxiety disorders	95.05	96.98	0.398	0.236	0.914	0.960	0.48	0.07	0.97	1.00	18.58	0.00	0.53	0.93
Anxiety problems	81.91	81.18	0.193	0.040	0.756	0.802	0.14	0.01	0.98	1.00	8.00	0.00	0.87	0.99
Obsessive Compulsive disorder	94.93	99.03	0.079	−0.005	0.945	0.990	0.50	0.00	0.96	1.00	12.50	0.00	0.52	1.00
Obsessive Compulsive problems	93.00	96.38	0.159	−0.009	0.913	0.954	0.20	0.00	0.96	1.00	5.56	0.00	0.83	1.00
Posttraumatic stress disorder	94.81	99.40	−0.002	−0.002	0.948	0.994	0.00	0.00	0.96	1.00	0.00	0.00	1.04	1.00
Posttraumatic stress problems	92.40	96.50	0.023	−0.008	0.919	0.965	0.04	0.00	0.96	1.00	0.95	0.00	1.00	1.00
Somatoform disorder	95.05	99.40	0.118	0.441	0.943	0.989	0.60	0.40	0.96	1.00	15.38	0.00	0.42	0.60
Somatoform problems	86.97	90.59	0.037	0.400	0.855	0.901	0.05	0.03	0.96	1.00	1.24	0.00	0.99	0.97
Eating disorder	95.05	99.40	0.044	−0.002	0.948	0.994	1.00	0.00	0.96	1.00	24.39	0.00	0.00	1.00
Eating problems	93.00	96.86	0.113	0.126	0.918	0.964	0.19	0.08	0.96	1.00	5.19	0.00	0.84	0.92

Note: Evaluation guidelines for κ (see also Table S4): <0.20 = no; 0.21–0.39 = minimal; 0.40–0.59 = weak; and 0.60–0.79 = moderate agreement.
Abbreviations: LR, likelihood ratio; M.I.N.I., mini-international neuropsychiatric interview.

TABLE 4 Agreement between K10 and K6 (cut-off score of 10 and 6 based on ROC analyses) and M.I.N.I. (N = 829)

Presence of ...	Correspondence rate (%)		Cohen's κ (k)[a]		Prevalence index (PI)[b]	
	K10 ≥ 10	K6 ≥ 6	K10 ≥ 10	K6 ≥ 6	K10 ≥ 10	K6 ≥ 6
Non-psychotic axis-I disorders	45.60	51.75	0.058	0.064	0.362	0.431
Non-psychotic axis-I problems	56.33	58.14	0.177	0.173	0.090	0.159
Mood disorders	43.06	49.46	0.022	0.022	0.390	0.458
Mood problems	47.29	53.20	0.079	0.091	0.333	0.402
Anxiety disorders	44.39	51.27	0.042	0.056	0.379	0.448
Anxiety problems	51.27	55.25	0.119	0.124	0.221	0.290
Obsessive compulsive disorder	41.62	48.25	0.003	0.000	0.409	0.478
Obsessive compulsive problems	43.55	49.70	0.027	0.026	0.378	0.446
Posttraumatic stress disorder	41.50	48.37	0.002	0.002	0.413	0.481
Posttraumatic stress problems	43.18	49.10	0.022	0.014	0.384	0.452
Somatoform disorders	41.98	48.85	0.009	0.011	0.408	0.476
Somatoform problems	46.44	52.59	0.060	0.078	0.320	0.388
Eating disorder	41.50	48.37	0.002	0.002	0.413	0.481
Eating problems	43.31	49.94	0.024	0.031	0.382	0.451

Note: Somatization disorder and Anorexia Nervosa were not reported and thus not considered here.

Abbreviations: M.I.N.I., mini-international neuropsychiatric interview; ROC, receiver operating characteristic.

[a]K10 Evaluation guidelines for Cohen's κ (K): K ≥ 0.81 = excellent agreement; 0.61–0.80 = substantial; 0.41–0.60 = moderate; 0.21–0.40 = fair; 0.10–0.20 = slight; K < 0.10 = no agreement.

[b]Prevalence index values between −1 and 1 and is 0 when both responses are equally probable, (i.e., their prevalence is 50%). A high prevalence index = a low prevalence rate; a low prevalence index = a high prevalence rate.

the Kessler scales (e.g., Brinckley et al., 2021; Hajebi et al., 2018), future qualitative studies might increase the psychometric properties of the Kessler scales in the Swiss community by changes in wording and/or item selection.

4.1 | Limitations and strengths

The study may have some potential limitations. First, the Swiss community sample may not necessarily represent all German-speaking communities. However, Switzerland has one of the third largest community of German speakers. Second, psychotic disorders (n = 5) were not included. Third, the point-prevalence of mental disorders (5.2%) was rather low, with only low rates of some diagnoses (mania, GAD, dysthymia, anorexia and bulimia nervosa), which, however, were comparable to the rates of these disorders in the community (Ritchie & Roser, 2018). Fourth, before conducting the BEAR study, the validity of telephone assessments against face-to-face assessments had only been assured for clinical high-risk symptoms/criteria (Michel, Schimmelmann, et al., 2014) but not examined for M.I.N.I. diagnoses. However, assessing mental disorders/symptoms over the phone has been considered sufficiently comparable to face-to-face interviews with added positive effect on the disclosure of personal/intimate data (Azad et al., 2021; Muskens et al., 2014; Smith et al., 2009; Zhang et al., 2017). Thus, the low prevalence rate of mental disorders is unlikely caused by the telephone assessment that, compared to face-to-face assessments, has to be considered a much more valid

assessment of mental disorders/problems than questionnaire assessments (Zhang et al., 2017) that were mainly used in studies on the Kessler scales' validity. Furthermore, the far higher frequency of mental problems (32.3%) did not substantially increase validity measures.

Despite the limitations, strengths of this study include validation of K6/10 against a reliable outcome, i.e., clinician-assessment of mental disorders and problems by M.I.N.I. as 'gold standard', the use of well-trained interviewers, the relatively large community sample size, the high response rate and the excellent representativeness of the baseline sample (Schultze-Lutter et al., 2018). Another strength links to the identification of cut-offs for both screeners and for estimation of the prevalence of mood and anxiety disorders. Lastly, besides using sensitivity, specificity, LRs were also calculated to offer additional useful diagnostic information to ensure a more accurate validation of screeners as each parameter offers a specific characteristic of the diagnostic test (Parikh et al., 2009).

5 | CONCLUSION

Given that integrated pre-clinical and primary care youth health services are becoming increasingly widespread around the world (Hetrick et al., 2017), reliable and valid screenings that accurately identify mental health problems/disorders are more important than ever to ensure effective and targeted intervention in high-risk populations. Our findings, however, suggest that K10/K6 should not be recommended to use as main screening tools without additional clinical assessment.

Although the focus of K10/K6 is mainly on depression and anxiety disorders, our findings illustrate that the screening performance for these most common mental disorders was nevertheless poor. The present findings speak to the vital need for validating a screening tool in both German-speaking community and clinical settings as symptoms might also be expressed differently between these. Further validation studies are required to conclude whether K10/K6 make an adequate initial screening tool in community settings, but also in very low-thresholds programs.

CONFLICT OF INTEREST
Drs. Chow, Michel, Meisenzahl and Schultze-Lutter, and Mr. Osman report no competing interests. Dr Schimmelmann reports serving as a consultant and/or advisor to, or receiving honoraria from Takeda and Shire outside the reported work.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Chantal Michel 🄳 https://orcid.org/0000-0003-1165-6681
Frauke Schultze-Lutter 🄳 https://orcid.org/0000-0003-1956-9574

REFERENCES
Alexander, M. J., Haugland, G., Lin, S. P., Bertollo, D. N., & McCorry, F. A. (2008). Mental health screening in addiction, corrections and social service settings: Validating the MMS. *International Journal of Mental Health Addiction*, 6, 105–119. https://doi.org/10.1007/s11469-007-9100-x

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV* (4th ed.). APA.

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler psychological distress scale (K10). *Australian and New Zealand Journal of Public Health*, 25, 494–497. https://doi.org/10.1111/j.1467-842x.2001.tb00310.x

Australian Government Department of Health. (2018). Primary Mental Health Care Minimum Data Set Scoring the Kessler-10 Plus. Retrieved from http://pmhc-mds.com/doc/pmhc-scoring-k10p.pdf. Accessed on 05.03.2021.

Azad, A., Sernbo, E., Svärd, V., Holmlund, L., & Björk Brämberg, E. (2021). Conducting in-depth interviews via mobile phone with persons with common mental disorders and multimorbidity: The challenges and advantages as experienced by participants and researchers. *International Journal of Environmental Research and Public Health*, 18, 11828. https://doi.org/10.3390/ijerph182211828

Batterham, P. J., Sunderland, M., Carragher, N., Calera, A. L., Mackinnon, A. J., & Slade, T. (2016). The distress questionnaire-5: Population screener for psychological distress was more accurate than the K6/K10. *Journal of Clinical Epidemiology*, 71, 35–42. https://doi.org/10.1016/j.jclinepi.2015.10.005

Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *The BMJ*, 314, 572. https://doi.org/10.1136/bmj.314.7080.572

Boerema, A. M., Ten Have, M., Kleiboer, A., De Graaf, R., Nuyen, J., Cuijpers, P., & Beekman, A. T. F. (2017). Demographic and need factors of early, delayed and no mental health care use in major depression: A prospective study. *BMC Psychiatry*, 17, 367. https://doi.org/10.1186/s12888-017-1531-8

Brinckley, M.-M., Calabria, B., Walker, J., Thurber, K. A., & Lovett, R. (2021). Reliability, validity, and clinical utility of a culturally modified Kessler scale (MK-K5) in the aboriginal and Torres Strait islander population. *BMC Public Health*, 21, 1111. https://doi.org/10.1186/s12889-021-11138-4

Burn, C. C., & Weir, A. A. (2011). Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers. *Veterinary Journal*, 188, 166–170. https://doi.org/10.1016/j.tvjl.2010.04.021

Büssing, A., Fischer, J., Haller, A., Heusser, P., Ostermann, T., & Matthiessen, P. F. (2009). Validation of the brief multidimensional life satisfaction scale in patients with chronic diseases. *European Journal of Medical Research*, 14, 171–177. https://doi.org/10.1186/2047-783X-14-4-171

Cairney, J., Veldhuizen, S., Wade, T. J., Kurdyak, P., & Streiner, D. L. (2007). Evaluation of 2 measures of psychological distress as screeners for depression in the general population. *Canadian Journal of Psychiatry*, 52, 111–120. https://doi.org/10.1177/070674370705200209

Cotton, S. M., Menssink, J., Filia, K., Rickwood, D., Hickie, I. B., Hamilton, M., Hetrick, S., Herrman, H., McGorry, P. D., & Gao, C. (2021). The psychometric characteristics of the Kessler psychological distress scale (K6) in help-seeking youth: What do you miss when using it as an outcome measure? *Psychiatry Research*, 305, 114182. https://doi.org/10.1016/j.psychres.2021.114182

EuroQol Group. (1990). EuroQol – A new facility for the measurement of health-related quality of life. *Health Policy*, 16, 199–208. https://doi.org/10.1016/0168-8510(90)90421-9

Ferro, M. A. (2019). The psychometric properties of the Kessler psychological distress scale (K6) in an epidemiological sample of Canadian youth. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, 64, 647–657. https://doi.org/10.1177/0706743718818414

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of mental health and well-being. *Psychological Medicine*, 33, 357–362. https://doi.org/10.1017/s0033291702006700

Giesinger, J., Rumpold, G., & Schüßler, G. (2008). Die K10-Screening-Skala für unspezifischen psychischen Distress. *Psychosomatik und Konsiliarpsychiatrie*, 2, 104–111. https://doi.org/10.1007/s11800-008-0100-x

Hajebi, A., Motevalian, A., Amin-Esmaeili, M., Rahimi-Movaghar, A., Sharifi, V., Hoseini, L., Shadloo, B., & Mojtabai, R. (2018). Adaptation and validation of short scales for assessment of psychological distress in Iran: The Persian K10 and K6. *International Journal of Methods in Psychiatric Research*, 27, e1726. https://doi.org/10.1002/mpr.1726

Hetrick, S. E., Bailey, A. P., Smith, K. E., Malla, A., Mathias, S., Singh, S. P., O'Reilly, A., Verma, S. K., Benoit, L., Fleming, T. M., Moro, M. R., Rickwood, D. J., Duffy, J., Eriksen, T., Illback, R., Fisher, C. A., & McGorry, P. D. (2017). Integrated (one-stop shop) youth health care: Best available evidence and future directions. *The Medical Journal of Australia*, 207, S5–S18. https://doi.org/10.5694/mja17.00694

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R., Smith, S. R., Price, J. L., Smith, C. L., Heindselman, T. L., Mount, M. K., & Holdwick, D. J. (2000). Reliability and validity of DSM-IV axis V. *The American Journal of Psychiatry*, 157, 1858–1863. https://doi.org/10.1176/appi.ajp.157.11.1858

Jaeschke, R., Guyatt, G. H., & Sackett, D. L. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients?

The evidence-based medicine working group. *JAMA*, *271*, 703–707. https://doi.org/10.1001/jama.271.9.703

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*, 959–976. https://doi.org/10.1017/s0033291702006074

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L. T., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, *60*, 184–189. https://doi.org/10.1001/archpsyc.60.2.184

Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., Furukawa, T. A., Gureje, O., Hinkov, H., Hu, C.-Y., 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. *International Journal of Methods in Psychiatric Research*, *19*(Suppl. 1), 4–22. https://doi.org/10.1002/mpr.310

Kirmayer, L. J., & Ryder, A. G. (2016). Culture and psychopathology. *Current Opinion in Psychology*, *8*, 143–148. https://doi.org/10.1016/j.copsyc.2015.10.020

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lecrubier, Y., Sheehan, D., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., Janavs, J., & Dunbar, G. (1997). The MINI international neuropsychiatric interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, *12*, 224–231. https://doi.org/10.1016/S0924-9338(97)83296-8

Michel, C., Schimmelmann, B. G., Kupferschmid, S., Siegwart, M., & Schultze-Lutter, F. (2014). Reliability of telephone assessments of at-risk criteria of psychosis: A comparison to face-to-face interviews. *Schizophrenia Research*, *153*, 251–253. https://doi.org/10.1016/j.schres.2014.01.025

Michel, C., Schnyder, N., Schmidt, S. J., Groth, N., Schimmelmann, B. G., & Schultze-Lutter, F. (2018). Functioning mediates help-seeking for mental problems in the general population. *European Psychiatry*, *54*, 1–9. https://doi.org/10.1016/j.eurpsy.2018.06.009

Michel, C., Schultze-Lutter, F., & Schimmelmann, B. G. (2014). Screening instruments in child and adolescent psychiatry: General and methodological considerations. *European Child & Adolescent Psychiatry*, *23*, 725–727. https://doi.org/10.1007/s00787-014-0608-x

Muskens, E. M. H., Lucassen, P., Groenleer, W., van Weel, C., Oude Voshaar, R., & Speckens, A. (2014). Psychiatric diagnosis by telephone: Is it an opportunity? *Social Psychiatry and Psychiatric Epidemiology*, *49*, 1677–1689. https://doi.org/10.1007/s00127-014-0861-9

Parikh, R., Parikh, S., Arun, E., & Ravi, T. (2009). Likelihood ratios: Clinical application in day-to-day practice. *Indian Journal of Ophthalmology*, *57*, 217–221. https://doi.org/10.4103/0301-4738.49397

Peasgood, T., Brazier, J., & Papaioannou, D. (2012) A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety. *HEDS discussion paper 12/15*. https://eprints.whiterose.ac.uk/74659/

Rabin, R., & De Charro, F. (2001). EQ-5D: A measure of health status from the EuroQol group. *Annals of Medicine*, *33*, 337–343. https://doi.org/10.3109/07853890109002087

Ritchie, H. & Roser, M. (2018). Mental health. *Our World in Data*. Retrieved from https://ourworldindata.org/mental-health. Accessed on 10.04.2021.

Rybarczyk, B. (2011). Social and occupational functioning assessment scale (SOFAS). In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 2313–2313). Springer. https://doi.org/10.1007/978-0-387-79948-3_428

Sakurai, K., Nishi, A., Kondo, K., Yanagida, K., & Kawakami, N. (2011). Screening performance of K6/K10 and other screening instruments for mood and anxiety disorders in Japan. *Psychiatry and Clinical Neurosciences*, *65*, 434–441. https://doi.org/10.1111/j.1440-1819.2011.02236.x

Sampasa-Kanyinga, H., Zamorski, M. A., & Colman, I. (2018). The psychometric properties of the 10-item Kessler psychological distress scale (K10) in Canadian military personnel. *PLoS One*, *13*, e0196562. https://doi.org/10.1371/journal.pone.0196562

Sancassiani, F., Romano, F., Balestrieri, M., Caraci, F., Di Sciascio, G., Drago, F., Hardoy, M. C., Moro, M. F., Roncone, R., Piras, M., Preti, A., Dell'Osso, L., Faravelli, C., & Carta, M. (2019). The prevalence of specific phobia by age in an Italian Nationwide survey: How much does it affect the quality of life? *Clinical Practice & Epidemiology in Mental Health*, *15*, 30–37. https://doi.org/10.2174/1745017901915010030

Schultze-Lutter, F., Michel, C., Ruhrmann, S., & Schimmelmann, B. G. (2018). Prevalence and clinical relevance of interview-assessed psychosis-risk symptoms in the young adult community. *Psychological Medicine*, *48*, 1167–1178. https://doi.org/10.1017/S0033291717002586

Schultze-Lutter, F., Schimmelmann, B. G., & Michel, C. (2021). Clinical high-risk of and conversion to psychosis in the community: A three-year follow-up of a cohort study. *Schizophrenia Research*, *228*, 616–618. https://doi.org/10.1016/j.schres.2020.11.032

Searle, A. K., Van Hooff, M., McFarlane, A. C., Davies, C. E., Tran, T., Hodson, S. E., Benassi, H. P., & Steele, N. M. (2017). Screening for depression and psychological distress in a currently serving military population: The diagnostic accuracy of the K10 and the PHQ9. *Assessment*, *26*, 1411–1426. https://doi.org/10.1177/1073191117745124

Sheehan, D., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M., & Dunbar, G. (1997). The validity of the MINI international neuropsychiatric interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, *12*, 232–241. https://doi.org/10.1016/S0924-9338(97)83297-X

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59*(Suppl. 20), 22–33.

Shon, E.-J. (2020). Measurement equivalence of the Kessler 6 psychological distress scale for Chinese and Korean immigrants: Comparison between younger and older adults. *International Journal Methods Psychiatric Research*, *29*, e1823. https://doi.org/10.1002/mpr.1823

Smith, A., Lyons, A., Pitts, M., Croy, S., Ryall, R., Garland, S., Wong, M. L., & Tay, E. H. (2009). Assessing knowledge of human papillomavirus and collecting data on sexual behavior: Computer assisted telephone versus face to face interviews. *BMC Public Health*, *9*, 429. https://doi.org/10.1186/1471-2458-9-429

Stolk, Y., Kaplan, I., & Szwarc, J. (2014). Clinical use of the Kessler psychological distress scales with culturally diverse groups. *International Journal of Methods in Psychiatric Research*, *23*, 161–183. https://doi.org/10.1002/mpr.1426

Sunderland, M., Slade, T., Stewart, G., & Andrews, G. (2011). Estimating the prevalence of DSM-IV mental illness in the Australian general population using the Kessler psychological distress scale. *Australian and New Zealand Journal of Psychiatry*, *45*, 880–889. https://doi.org/10.3109/00048674.2011.606785

Szende, A., Oppe, M., & Devlin, N. (Eds.). (2007). *EQ-5D value sets: Inventory, comparative review and user guide*. Springer Netherlands. https://doi.org/10.1007/1-4020-5511-0

Thelin, C., Mikkelsen, B., Laier, G., Turgut, L., Henriksen, B., Olsen, L., Larsen, J., & Arnfred, S. (2017). Danish translation and validation of

Kessler's 10-item psychological distress scale – K10. *Nordic Journal of Psychiatry*, 71, 411–416. https://doi.org/10.1080/08039488.2017.1312517

Trautmann, S., Rehm, J., & Wittchen, H.-U. (2016). The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO Reports*, 17, 1245–1249. https://doi.org/10.15252/embr.201642951

Webb, M. J., Kauer, S. D., Ozer, E. M., Haller, D. M., & Sanci, L. A. (2016). Does screening for and intervening with multiple health compromising behaviours and mental health disorders amongst young people attending primary care improve health outcomes? A systematic review. *BMC Family Practice*, 17, 104. https://doi.org/10.1186/s12875-016-0504-1

Zhang, X., Kuchinke, L., Woud, M. L., Velten, J., & Margraf, J. (2017). Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ. *Computers in Human Behavior*, 71, 172–180. https://doi.org/10.1016/j.chb.2017.02.006

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.