

Hierarchical screening for multiple mental disorders

Philip J. Batterham ¹, Alison L. Calear ¹, Matthew Sunderland ², Natacha Carragher ², Helen Christensen ³, Andrew J Mackinnon ⁴

¹ Centre for Mental Health Research, The Australian National University, Canberra, Australian Capital Territory, Australia

² National Drug and Alcohol Research Centre, The University of New South Wales, Sydney, New South Wales, Australia

³ Black Dog Institute, The University of New South Wales, Sydney, New South Wales, Australia

⁴ Orygen Youth Health Research Centre, The University of Melbourne, Melbourne, Victoria, Australia

Correspondence: Philip Batterham
Centre for Mental Health Research
Building 63, Eggleston Road
The Australian National University
Canberra 0200 Australia
Ph: +61 2 6125 1031 Fax: +61 2 6125 0733
Email: philip.batterham@anu.edu.au

Word count: 4730

Abstract

Background: There is a need for brief, accurate screening when assessing multiple mental disorders. Two-stage hierarchical screening, consisting of brief pre-screening followed by a battery of disorder-specific scales for those who meet diagnostic criteria, may increase the efficiency of screening without sacrificing precision. This study tested whether more efficient screening could be gained using two-stage hierarchical screening than by administering multiple separate tests.

Method: Two Australian adult samples (N=1990) with high rates of psychopathology were recruited using Facebook advertising to examine four methods of hierarchical screening for four mental disorders: major depressive disorder, generalized anxiety disorder, panic disorder and social phobia.

Results: Using K6 scores to determine whether full screening was required did not increase screening efficiency. However, pre-screening based on two decision tree approaches or item gating led to considerable reductions in the mean number of items presented per disorder screened, with estimated item reductions of up to 54%. The sensitivity of these hierarchical methods approached 100% relative to the full screening battery.

Limitations: Further testing of the hierarchical screening approach based on clinical criteria and in other samples is warranted.

Conclusions: The results demonstrate that a two-phase hierarchical approach to screening multiple mental disorders leads to considerable increases efficiency gains without reducing accuracy. Screening programs should take advantage of prescreeners based on gating items or decision trees to reduce the burden on respondents.

Key words: depression, anxiety disorders, screening

There is a demand for fast and accurate mental health screening in a range of general population settings where time and resource restrictions preclude administration of comprehensive clinical measures. Mental health screening has application to research and to clinical settings, including virtual clinics (Cuijpers et al., 2009; Donker et al., 2009), primary care (Spitzer et al., 1999) and schools (Husky et al., 2011; Weist et al., 2007). Although there are accurate and brief self-report scales available for assessing many specific mental disorders, there is little knowledge about the most efficient methods of screening for multiple disorders (Donker et al., 2009). We propose a new method of screening for multiple mental disorders that combines the strong psychometric characteristics of existing epidemiological screening scales with shorter mean administration time. Specifically, hierarchical screening refers to a multi-phase screening process involving presentation of a very brief screener for general psychological distress, followed by more lengthy screening for respondents who meet specified criteria on the brief screener. This hierarchical selection process has the potential to increase efficiency through the reduction of mean administration time, without loss of precision. Furthermore, hierarchical screening across multiple disorders may generate efficiency gains as a result of the high rates of comorbidity seen among mental disorders (Kessler et al., 2005), which may be overlooked by serial screening methods.

Two-phase screening has been previously used to screen for a single disorder (Clover et al., 2009). The two-phase approach is also cognate to two-phase clinical diagnosis, which is often used in national mental health surveys (e.g., Sunderland et al., 2012), although this approach is not true multi-phase screening as the second phase of assessment is a lengthy diagnostic interview. One previous study developed this joint screener-diagnosis process further, providing a psychometric evaluation of a gating approach to screen for multiple disorders using the DISC Predictive Scales screen (DPS screen; Lucas et al., 2001). This evaluation found greater screening efficiency through the use of gating items followed by a

brief screener for each disorder. However, the DPS screen was developed for children and adolescents and requires interviewer administration. In addition, the sensitivity of the screeners was not high for several of the disorders, resulting in missed cases. There has been no further development of this gating approach for the self-screening of community-based populations on the basis of multiple disorders. Moreover, a number of alternative optimisation strategies are yet to be tested, with potential to further decrease total response burden without reducing the sensitivity of the existing measures.

In the current study, four methods of hierarchical screening were tested in two online population samples displaying high rates of psychopathology. Participants in both samples completed all mental health measures. The hierarchical screening methods were therefore developed by estimating the mean number of items that would have been presented to participants, simulated based on the criteria for each hierarchy. The screening methods tested were: (i) no hierarchy (control), (ii) a hierarchy based on K6 scores (Kessler et al., 2002), (iii) a hierarchy based on a decision tree using general psychological distress items, (iv) a hierarchy based on a decision tree using items from disorder-specific scales, and, (v) a hierarchy using gating questions from each of the disorder-specific scales.

The K6 tends to be used to identify those at high risk of disorder. However, there has been little examination of whether the K6 can identify individuals at low risk of disorder, thereby reducing their need to respond to a broader battery of screening measures. To identify a more tailored set of items identifying low risk of disorder, two decision tree approaches were tested. The decision tree is a method to identify subgroups at high risk of a specified, known outcome (see Batterham et al., 2009). This method has not previously been applied to screening, but has the potential to optimally distinguish absence of disorder from presence of disorder by identifying a set of screening items tailored to an individual's risk for disorder. Similar to the DPS screen approach, a gating item approach was also tested using existing

measures that incorporate gating in the identification of disorder. Although the present study focused on four common internalizing disorders, the study aimed to lay the foundation for the development of more comprehensive screeners including additional disorders. It was hypothesized that each of the hierarchical screening methods would lead to a reduced mean number of items presented, with negligible reduction in sensitivity.

Method

Participants and procedure

Australian adults were recruited for an online survey using Facebook advertising during July 2012. The survey was hosted on a secure server at the Australian National University. Surveys were completed in approximately 20-30 minutes and included online informed consent, the screening measures, a number of other mental health measures, sociodemographic characteristics, and concluded with service referral options. For this initial version of the survey (Sample 1), a total of 1360 surveys were completed over two months. This survey was advertised in Facebook as a “mental health” survey. Clicks on the advertisement directly referred individuals to the secure survey website.

A second online survey was conducted in October 2012 to examine additional mental health screening measures. The survey was a similar length to the original survey but replaced a number of suicide measures with additional items assessing general psychological distress. The survey also differed in its use of an internal Facebook advertisement for recruitment, that is, the advertisement linked to a Facebook page that contained prominent links to the survey rather than linking directly to the survey (this recruitment method was less costly). This survey (Sample 2) had 630 respondents, yielding a total of 1990 participants across both versions of the survey. Further information about the recruitment process is

described elsewhere (Batterham, in press). Human research ethics approval (protocol #2012/310) was obtained from the Science and Medical Delegated Ethics Review Committee at the Australian National University.

Measures

The presence of four common mental disorders—Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), Panic Disorder (PD), and Social Phobia (SP)—was assessed using validated self-report measures. The four measures acted as the standard for identifying criteria for these disorders in the sample. MDD was assessed using the 9-item Patient Health Questionnaire-9 (PHQ-9; Spitzer et al., 1999), with presence of MDD defined based on the algorithm identified by the authors of the scale. Specifically, respondents who endorsed the presence of five or more symptoms in the last two weeks (rated as ‘more than half the days’ or higher; ‘several days’ or higher for the suicidal ideation symptom), with at least one of these symptoms being anhedonia or depressed mood were classified as experiencing current MDD. The other seven items of the PHQ-9 assess sleep disturbance, fatigue, appetite changes, guilt, difficulty concentrating, motor retardation/agitation, and suicidal ideation. The criterion and construct validity of the PHQ-9 have previously been demonstrated, with 73% sensitivity and 98% specificity in detecting major depression compared to clinician-based assessment of DSM-IV criteria and strong relationships with measures of functional impairment and health care use (Spitzer et al., 1999). The measure has demonstrated excellent internal consistency (Cronbach $\alpha > 0.85$ in multiple samples) and 48-hour test-retest reliability of 0.84 (Kroenke et al., 2001). GAD was assessed using the 7-item GAD-7 scale (Spitzer et al., 2006), which was also scored using the authors’ diagnostic algorithm (see Spitzer et al., 1999) based on ratings of “more than half the days” or “nearly every day” on the first item and at least three subsequent items, assessed over the previous

two weeks. The items assess feelings of anxiety, inability to control worrying, excessive worrying, trouble relaxing, restlessness, irritability and feeling afraid. Presence of PD was based on the 5-item Patient Health Questionnaire-Panic scale (PHQ-panic; Spitzer et al., 1999) assessed over the previous four weeks. Endorsement of all five items (presence of panic attack, previous panic attack, sudden attack, worry about attacks, multiple symptoms present) indicated presence of PD.

SP was assessed using a 5-item screening scale based on DSM-IV criteria (social phobia screen, SOPHS). The items were: 1) “To what extent have you felt fearful or embarrassed of any social situations during the past month?”, 2) “Was the fear or embarrassment you experienced during the past month excessive or unreasonable?”, 3) “During the past month, have you avoided any social situations because of your fear or embarrassment?”, 4) “During the past month, how much have you suffered through any social situations because of your fear or embarrassment?”, and 5) “During the past month, how much has your work, home or social life been disrupted because of your fear or embarrassment?”. Each item was rated on a 5-point Likert scale: not at all (0), a little (1), moderately (2), severely (3), extremely (4). Presence of SP was based on ratings higher than 2 on items 1, 2 and 5, and a rating higher than 2 on either item 3 or 4. In the present study, the SOPHS had 85.4% sensitivity and 77.2% specificity in comparison to the Social Phobia Scale (Mattick and Clarke, 1998).

In addition, general psychological distress was assessed using the Kessler-6 (K6; Kessler et al., 2002). In the second survey, supplementary psychological distress items were also included. These consisted of the ten items from the Kessler-10 (K10; Kessler et al., 2002): depressed, unceasingly depressed, hopeless, restless, unceasingly restless, tired out, everything effortful, worthless, nervous, unceasingly nervous, along with ten additional items validated in the development of the K10 (“in a really good mood”, “irritable”, “happy”,

“keyed up or on edge”, “worried about things that were not really important”, “that nothing was worthwhile anymore”, “physically tense or shaky”, “angry”, “that you would be better off dead”; Kessler et al., 2002) and the 14 items from the Warwick-Edinburgh Mental Well-being Scale (WEMWBS; Tennant et al., 2007): optimistic, useful, relaxed, interested in people, energetic, solving problems, clear thinking, feeling good, feeling close, confident, decision making, feeling loved, interesting in things and cheerful.

Analysis

The mean number of items that would be presented to respondents was calculated for each of four simulated screening methods. Respondents completed all of the screening scale items, so that their responses could be used to assess each of the five methods.

- 1) Participants receive the four screening scales (PHQ-9, GAD-7, PHQ-panic, SOPHS) with no hierarchy (control method, “no hierarchy”)
- 2) Participants receive the K6, followed by the four screening scales for those who meet criteria (“K6 hierarchy”)
- 3) Participants receive an adaptive set of psychological distress items, followed by the four screening scales for those who meet pre-determined criteria on the initial distress items (“Psychological Distress hierarchy”)
- 4) Participants receive an adaptive set of items from the four screening scales, followed by the remainder of items from the four screening scales for those who meet pre-determined criteria on the initial items (“Disorder-based hierarchy”)
- 5) Participants receive the first two gating items of the PHQ-9 and the first gating items of the GAD-7, PHQ-panic and SOPHS, followed by the remainder of items from the each of the screening scales for those who meet criteria on the initial items (“Gating hierarchy”)

The adaptive decision tree methods (methods 3 and 4) tailor items presented based on previous responses (rather than items being presented serially), such that different respondents may receive a different set of questions. The Gating hierarchy criteria were based on the scoring rules for each of the scales, specifically, responses on at least one of the first two PHQ-9 items being “more than half the days” or higher, response of “more than half the days” or higher on the first GAD7 item, response of “yes” on the first PHQ-panic item and response higher than “moderately” on the first SOPHS item. The four hierarchical methods (methods 2-5) are presented in Figures 1-4. For each of the hierarchical methods, criteria for full screening were set to maximise sensitivity for identifying any of the specific disorders, although with a recognition of the need to exclude a sufficient proportion of the sample to make the inclusion of the additional hierarchy items worthwhile. The cut-off on the K6 was based on sensitivity closest to 95%, with sensitivity calculations made with reference to administration of the four disorder-specific scales. By definition, the specificity for all hierarchical methods relative to the control method was 100%, as no additional participants were identified as meeting disorder criteria.

The Psychological Distress hierarchy method identified the psychological distress items that best distinguished the presence of one or more disorders from absence of all disorders using a decision tree approach. The decision tree was constructed using the *treedisc* macro in SAS v9.1.3, which is similar to the Chi-square Automatic Interaction Detector method described by Kass (1980). The *treedisc* macro selects items on the basis of the minimum p-value from the chi-square statistic comparing individuals with and without disorder. This branching is repeated to build a tree that represents the ordering of items to be presented, with these items designed to optimally distinguish absence of disorder from presence of any disorder. The first item was chosen as the one that best distinguished absence of disorder, while subsequent items were chosen to optimally distinguish absence of disorder

at each response level of the previous item (see Figures 2 and 3 for examples). Branching was stopped when there were no variables with a p-value less than 0.2 for division. This highly liberal p-value was used for maximal branching to ensure maximum sensitivity, that is, to ensure that individuals who met criteria for a disorder would not be excluded from second-phase screening. The minimum sample size for each leaf (node) was specified as $n = 20$, and branching was limited to six levels to ensure the initial screener would be brief. The same process was used to select items for the Disorder-based hierarchy.

Candidate items for the Psychological Distress hierarchy were the 20 items from the K10 validation (Kessler et al., 2002) and the 14 items of the WEMWBS (Tennant et al., 2007). Candidate items for the Disorder-based hierarchy were the 26 items from the four disorders-specific screening scales. The Psychological Distress hierarchy was only tested in Sample 2 ($n = 630$) who completed the version of the survey that contained the extended set of psychological distress items. The Disorder-based hierarchy was developed using the Sample 1 ($n = 1360$) and then validated in the Sample 2 ($n = 630$).

Results

Characteristics of the two survey samples are shown in Table 1. The majority of the missingness on the demographic variables was attributable to these items being presented at the end of the survey. Sample 1 was recruited using an advertisement that directed respondents directly to the online survey. Sample 2 was recruited with an advertisement that linked to an internal Facebook page that included prominent links to the survey and allowed sharing of the page and link across individuals' social networks. Compared to Sample 1, Sample 2 were more likely to be older, female, married or separated/divorced, have lower education, be English speaking only, have MDD, have GAD, have PD and not have SP. Both samples had high rates of psychopathology, far exceeding the rates seen in the general

Australian population at large (Slade et al., 2009), and with high rates of comorbidity. As a result, to estimate the efficiency of hierarchical screening in the population, the analyses included a projection of the mean number of items presented based on a 50% increase in the proportion of the sample that did not require full screening.

Development of the hierarchies

The four hierarchies that were tested are presented in Figures 1-4. The cut-off of 4 on the K6 (shown in Figure 1) distinguished between cases and non-cases with 95.7% sensitivity. Participants who scored 0-4 (n = 625; 31.4%) were therefore categorised as being eligible to receive only the K6 items without further screening, while the remainder (n = 1365) were categorised as requiring full screening.

The decision tree for the Psychological Distress hierarchy is shown in Figure 2. The tree displays six levels of item branching, such that respondents may receive up to six psychological distress items before determination of whether administration of the full screening battery (PHQ-9, GAD-7, PHQ-panic and SOPHS) is then required. For example, all respondents would receive item K7, “depressed” first (this item comes from the 20 items used to validate the K10). Individuals who respond 0 (“none of the time”) for this item would then receive item 9 from the WEMWBS (“close to others”), those who respond 1 or 2 (“a little” or “some” of the time) would then receive item 18 (“tense/shaky”), and those who respond 3 or 4 (“most” or “all” of the time) would continue directly to the full screening battery. This process would continue until a termination decision point is reached, which is either to administer full screening, or stop assessment and identify that individual as not meeting criteria for any disorder. This hierarchy categorised 180 of 630 participants as not requiring further screening. As 310 participants in this sample did not meet criteria for any

disorder, the decision tree screening process identified 58% of these participants using six or fewer items.

The Disorder-based decision tree in Figure 3 was developed in the same way as the Psychological distress decision tree. Individuals respond to selected items from the four screening measures (PHQ-9, GAD-7, PHQ-panic and SOPHS) to determine whether full screening is required. A key difference of this hierarchical screening process is that fewer items are required in the secondary (full) phase of screening, as the items included in the first phase need not be repeated. The Disorder-based decision tree categorised 666 participants (of 1360) as not requiring further screening, representing 78% of participants who did not meet criteria for any of the four disorders. This tree required a maximum of only five items to reach a termination decision.

The Gating hierarchy presented in Figure 4 relied on the scoring algorithms of the four scales to separate out one or two screening items per scale for the first phase of screening. This results in presentation of only five items to those respondents who do not meet criteria for any disorder, without excluding any respondent who may meet criteria from full screening. This method gains further efficiency by presenting only the selected full scales for which an individual meets screening criteria in the first phase, rather than requiring presentation of all four scales.

Efficiency gains of the hierarchies

The efficiency of each screening method, assessed as the mean number of items presented to respondents for each method, is shown in Table 2. Methods 1, 2 and 5 were assessed in the combined samples from both surveys. Method 3 was only assessed among the participants in Sample 2, as the full battery of psychological distress items was not assessed in Sample 1. To test the validity of the decision tree methods, the Disorder-based decision

tree was developed using Sample 1 and then validated using Sample 2. In the control condition (no hierarchy), items on the PHQ-panic and SOPHS were gated for all respondents, such that a negative response to the first item of each scale resulted in the remaining items in that scale being skipped. Consequently, the mean number of items presented for this method was less than the 26 items from the four scales.

The table indicates that there was only a 0.9% efficiency gain from the K6 hierarchy, projected to be 17.3% among general population samples with lower rates of psychopathology. The efficiency gain was greater for the Psychological Distress hierarchy at 16.4% (projected to 26.4%). A further efficiency gain was found for the Disorder-based hierarchy, at 37.3% (43.6% projected) within Sample 1, which was used to develop the decision tree, and 25.9% (30.9% projected) in Sample 2 where the decision tree was validated. Across both samples, the Disorder-based hierarchy resulted in a mean of 14.6 items being presented, a gain of 33.6% in efficiency (projected to 13.3 items, 39.5% gain). The Gating hierarchy had the highest efficiency gain of 46.4% (projected to 54.1%). The sensitivities of the Psychological Distress, Disorder-based and Gating hierarchies were far superior to the K6 hierarchy, approaching 100% sensitivity relative to use of no hierarchy. The application of the Disorder-based hierarchy to Sample 2 also demonstrated that the decision tree was valid across samples, with no loss of sensitivity.

Discussion

Typically, screening for multiple disorders has relied on serial presentation of a range of scales that are either brief but imprecise (e.g., Donker et al., 2009) or more accurate but somewhat lengthier (e.g., Kessler et al., 2012; Spitzer et al., 1999). The results of the present study indicate that a hierarchical screening approach can substantially increase screening efficiency without sacrificing precision. The use of gating items was shown to be the most

successful method of reducing response burden. The use of a decision tree to identify the few items that best distinguished participants who were least at risk of meeting criteria for disorder also led to considerable efficiency gains. Using the gating or decision tree methods, the respondents who did not meet criteria for any of the four disorders could predominantly be screened using six or fewer items. Overall, the efficiency of these methods was reflected in the reduction of items presented by up to 54%. The efficiency gains are likely to be even greater in population samples with lower rates of psychopathology. The hierarchical methods resulted in high sensitivity relative to administering the entire battery of screening measures. Population screening settings that require highly sensitive measures to minimise the costs of false positives, such as primary care and school screening programs, may derive particular benefit from use of this efficient approach to screening.

The hierarchy using the K6, an existing measure of psychological distress, did not perform efficiently. The K6 was designed to identify individuals at high risk of meeting criteria for a psychological disorder (Kessler et al., 2002). That is, the K6 was designed as a screener to “rule in” possible psychopathology, rather than ruling it out. Therefore its use as a gateway or preliminary screener to reduce the response burden of individuals without psychopathology may be limited. Moreover, the majority of items identified as optimal in the Psychological distress decision tree were not derived from the K6.

It is important to note that the reductions in response burden achieved by hierarchical screening are gained from individuals at very low risk of psychopathology. Existing measures are generally designed with the intent of identifying individuals at the upper end of the liability spectrum. The present study included additional items that assessed both distress and well-being in an effort to better identify low-risk individuals. Using the decision tree approach, more optimal sets of distress items were selected for preliminary screening, resulting in efficiency gains.

The two-phase hierarchical screening tested in the present study may be extended to multiple-phase screening for a broader range of mental health problems. For example, a three-phase screener may consist of assessing general psychological distress, followed by separate pre-screeners for internalizing and externalizing disorders and then screeners for specific mental disorders. Further research in larger population-based samples may bear out the feasibility of a higher-order hierarchy. Furthermore, the efficiencies of hierarchical screening may be incorporated into adaptive screening programs, where presentation of each item is contingent on previous responses. For example, computer adaptive scales for assessing depression and anxiety, using Item Response Theory algorithms to select items, are highly efficient and precise in assessing severity on these traits (Pilkonis et al., 2011). Our research team intends to test embedding these adaptive measures into a hierarchical screening framework, along with developing adaptive screening measures for other mental disorders.

Although this is the first study to systematically examine hierarchical screening, there are limitations to be acknowledged. No clinical assessment was made, so the sensitivities of the hierarchical methods were benchmarked against administration of the complete scales. Research is planned to further test hierarchical screening by comparison of existing and new screening measures with clinical diagnosis. However, given previous evidence that the individual disorder-specific screeners are psychometrically sound relative to diagnosis, the implementation of the hierarchy as described would not damage the existing psychometric properties of the scales, as sensitivity relative to the scales approached 100% and specificity by definition was 100%. As participants included in the analyses received all screening items, the calculation of mean items received for each of the screening methods reflected simulated estimates that may differ when individuals are involved in a true hierarchical screening process. As both surveys were advertised as “mental health surveys”, interest in completing the surveys appeared to be higher among individuals with direct or indirect experience of

mental health problems, resulting in samples with high rates of psychopathology. These high rates of psychopathology in the samples likely resulted in underestimates of the efficiency gains of hierarchical screening. True, randomly selected and nationally representative population samples are likely to result in greater efficiencies gained through hierarchical screening, as suggested by the conservatively projected estimates presented in Table 2. In the Gating hierarchy, the efficiency gained was a result of the scoring algorithms of the scales, which require particular responses on one or two items for an individual to meet criteria for disorder. While some of the efficiency gains of the Disorder decision tree hierarchy were also related to the scoring algorithms, the decision tree approach has more flexibility than the gating approach for application to other scales. Indeed, additional testing of the decision tree method using other epidemiological scales found similar efficiency gains that would not be realised using a gating method [analyses not presented, based on the Centers for Epidemiologic Studies Depression scale (Radloff, 1977) and Social Phobia Scale, (Mattick and Clarke, 1998) in place of the PHQ-9 and SOPHS].

A related limitation is that hierarchical screening may not be appropriate for all screening situations. Epidemiological studies and clinical trials may require complete data on all mental health measures, which would not be compatible with a hierarchical approach. The researcher or clinician should weigh up the need for brevity with the need for complete data. Similarly, the four screening scales (PHQ-9, GAD-7, PHQ-panic, SOPHS) are easily presented in a paper format or programmed in a computer-based survey. Although it would not be difficult to implement hierarchical screening in a computer-based format, the branching criteria used for the gating and decision tree methods would require additional programming and testing. Furthermore, there are differences in the response scales of the measures (e.g., “Not at all” to “Nearly every day”, “Yes” / “No”). Development of adaptive

measures that use a common stem and response scale is likely to simplify the screening process.

The promising results of the present study add to the evidence that emerging, tailored methods of screening may result in efficient and precise assessment of an individual's needs. In particular, service settings that have time and resource constraints on the assessment of mental health status, such as primary care and schools, may benefit from the efficiency gains of hierarchical screening. Despite strong evidence for the effectiveness of e-health programs (Griffiths et al., 2010), there is a pressing need to develop efficient and precise screening programs to direct individuals to appropriate online programs, particularly as virtual mental health services become further developed (Christensen and Hickie, 2010). Population-based research trials may also benefit from using a hierarchical screening approach to more efficiently identify individuals at risk of multiple mental disorders. Hierarchical screening generates this efficiency while maintaining the robust psychometric properties of the existing screeners for the specific disorders. Other approaches to screening for multiple disorders (e.g., Donker et al., 2009) that rely on few items without secondary screening are briefer and easier to implement, but tend to have lower accuracy.

Based on the current findings, screening for multiple disorders using serial presentation of multiple scales is an inefficient method for identifying individuals who are likely to meet criteria for one or more disorders. Health professionals, school professionals and researchers are likely to benefit from the administration of a two-phase hierarchical screener for multiple disorders. First-phase screening using gating items or a decision tree to identify whether an individual would benefit from more comprehensive screening is likely to markedly reduce overall response burden. Hierarchical screening is likely to be of benefit in a number of settings where established screening programs link to appropriate care, including primary care, schools and virtual clinics. The validation of hierarchical screening methods is

an important step towards the provision of efficient and accurate screening for multiple mental disorders.

References

- Batterham, P.J., in press. Recruitment of mental health survey participants using internet advertising: Content, characteristics and cost effectiveness. *International Journal of Methods in Psychiatric Research*.
- Batterham, P.J., Christensen, H., Mackinnon, A.J., 2009. Modifiable risk factors predicting major depressive disorder at four year follow-up: a decision tree approach. *BMC Psychiatry* 9, 75.
- Christensen, H., Hickie, I.B., 2010. Using e-health applications to deliver new mental health services. *Med J Aust* 192, S53-56.
- Clover, K., Carter, G.L., Mackinnon, A., Adams, C., 2009. Is my patient suffering clinically significant emotional distress? Demonstration of a probabilities approach to evaluating algorithms for screening for distress. *Support Care Cancer* 17, 1455-1462.
- Cuijpers, P., Smits, N., Donker, T., ten Have, M., de Graaf, R., 2009. Screening for mood and anxiety disorders with the five-item, the three-item, and the two-item Mental Health Inventory. *Psychiatry Res* 168, 250-255.
- Donker, T., van Straten, A., Marks, I., Cuijpers, P., 2009. A brief Web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 11, e19.
- Griffiths, K.M., Farrer, L., Christensen, H., 2010. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. *Med J Aust* 192, S4-11.
- Husky, M.M., Sheridan, M., McGuire, L., Olfson, M., 2011. Mental health screening and follow-up care in public high schools. *J Am Acad Child Adolesc Psychiatry* 50, 881-891.

Kass, G.V., 1980. An exploratory technique for investigating large quantities of categorical data. *Applied statistics* 29, 119-127.

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., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., Walters, E.E., 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 62, 593-602.

Kessler, R.C., Calabrese, J.R., Farley, P.A., Gruber, M.J., Jewell, M.A., Katon, W., Keck, P.E., Nierenberg, A.A., Sampson, N.A., Shear, M.K., Shillington, A.C., Stein, M.B., Thase, M.E., Wittchen, H.U., 2012. Composite International Diagnostic Interview screening scales for DSM-IV anxiety and mood disorders. *Psychol Med*, 1-13.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 16, 606-613.

Lucas, C.P., Zhang, H., Fisher, P.W., Shaffer, D., Regier, D.A., Narrow, W.E., Bourdon, K., Dulcan, M.K., Canino, G., Rubio-Stipec, M., Lahey, B.B., Friman, P., 2001. The DISC Predictive Scales (DPS): efficiently screening for diagnoses. *J Am Acad Child Adolesc Psychiatry* 40, 443-449.

Mattick, R.P., Clarke, J.C., 1998. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther* 36, 455-470.

Pilkonis, P.A., Choi, S.W., Reise, S.P., Stover, A.M., Riley, W.T., Cella, D., 2011. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. *Assessment* 18, 263-283.

Radloff, L.S., 1977. The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1, 385-401.

Slade, T., Johnston, A., Oakley Browne, M.A., Andrews, G., Whiteford, H., 2009. 2007 National Survey of Mental Health and Wellbeing: methods and key findings. *Aust N Z J Psychiatry* 43, 594-605.

Spitzer, R.L., Kroenke, K., Williams, J.B., 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. Jama* 282, 1737-1744.

Spitzer, R.L., Kroenke, K., Williams, J.B., Lowe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 166, 1092-1097.

Sunderland, M., Slade, T., Andrews, G., 2012. Developing a short-form structured diagnostic interview for common mental disorders using signal detection theory. *Int J Methods Psychiatr Res* 21, 247-257.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., Stewart-Brown, S., 2007. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes* 5, 63.

Weist, M.D., Rubin, M., Moore, E., Adelsheim, S., Wrobel, G., 2007. Mental health screening in schools. *J Sch Health* 77, 53-58.

Table 1: Characteristics of the two online samples

	Sample 1 (N = 1360)	Sample 2 (N = 630)	χ^2	p
Age group			126.2	<0.001
18-24	436 (32.1%)	63 (10.0%)		
25-49	394 (29.0%)	297 (47.1%)		
≥50	487 (35.8%)	249 (39.5%)		
Missing/refused	43 (3.2%)	21 (3.3%)		
Gender			126.8	<0.001
Male	553 (40.7%)	98 (15.6%)		
Female	759 (55.8%)	508 (80.6%)		
Other/missing/refused	48 (3.5%)	24 (3.8%)		
Location			6.0	0.201
Metropolitan	742 (54.6%)	323 (51.3%)		
Regional	359 (26.4%)	183 (29.0%)		
Rural/remote	215 (15.8%)	100 (15.9%)		
Refused/missing	44 (3.2%)	24 (3.8%)		
Marital status			60.1	<0.001
Married/de facto	489 (36.0%)	274 (43.5%)		
Single, never married	541 (39.8%)	148 (23.5%)		
Separated/divorced	217 (16.0%)	161 (25.6%)		
Widowed	51 (3.8%)	23 (3.7%)		
Missing/refused	62 (4.6%)	24 (3.8%)		
Educational attainment			23.8	<0.001
Secondary not completed	179 (13.2%)	136 (21.6%)		
Completed secondary school	414 (30.4%)	163 (25.9%)		
Post-secondary education	721 (53.0%)	308 (48.9%)		
Speak non-English at home	204 (15.0%)	42 (6.7%)	27.4	<0.001
Depression (PHQ-9)	340 (25.0%)	212 (33.7%)	16.1	<0.001
Generalised Anxiety Disorder (GAD-7)	274 (20.1%)	166 (26.3%)	9.6	0.002
Panic Disorder (PHQ-panic)	185 (13.6%)	140 (22.2%)	23.4	<0.001
Social Phobia (SOPHS)	213 (15.7%)	42 (6.7%)	31.2	<0.001
Number of disorder criteria met				
0	850 (62.5%)	330 (52.4%)	38.8	<0.001
1	224 (16.5%)	124 (19.7%)		
2	130 (9.6%)	116 (18.4%)		
3	96 (7.1%)	36 (5.7%)		
4	60 (4.4%)	24 (3.8%)		

Table 2: Sensitivity and mean number of items presented for each screening method compared to detection by separate multiple screening tests

Method	Sensitivity	Mean items presented	Projected items presented
1. No hierarchy	–	22.0	22.0
2. K6 hierarchy	95.7%	21.8	18.2
3. Psychological Distress hierarchy	99.7% ²	18.4	16.2
4. Disorder-based hierarchy	99.9% ¹	13.8	12.4
	100.0% ²	16.3	15.2
5. Gating hierarchy	100.0%	11.8	10.1

Notes: sensitivity assessed relative to presenting all items; ¹ based on the Survey 1 sample (n = 1360); ² based on the Survey 2 sample (n = 630)

Figure 1: The K6 hierarchy (n = 1990)

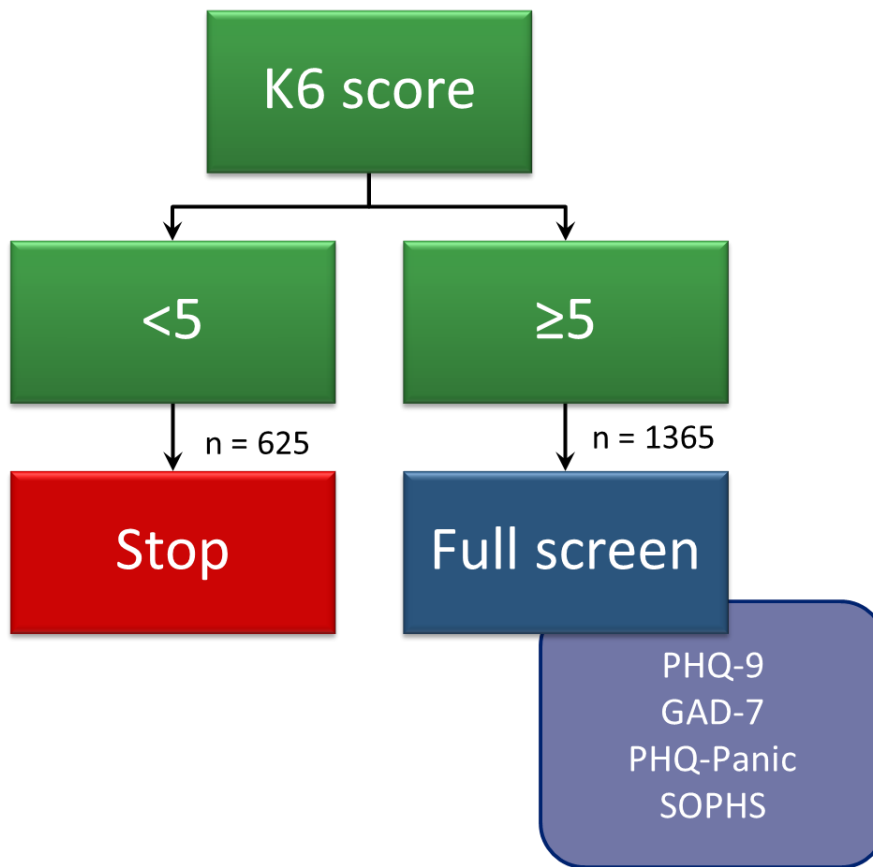
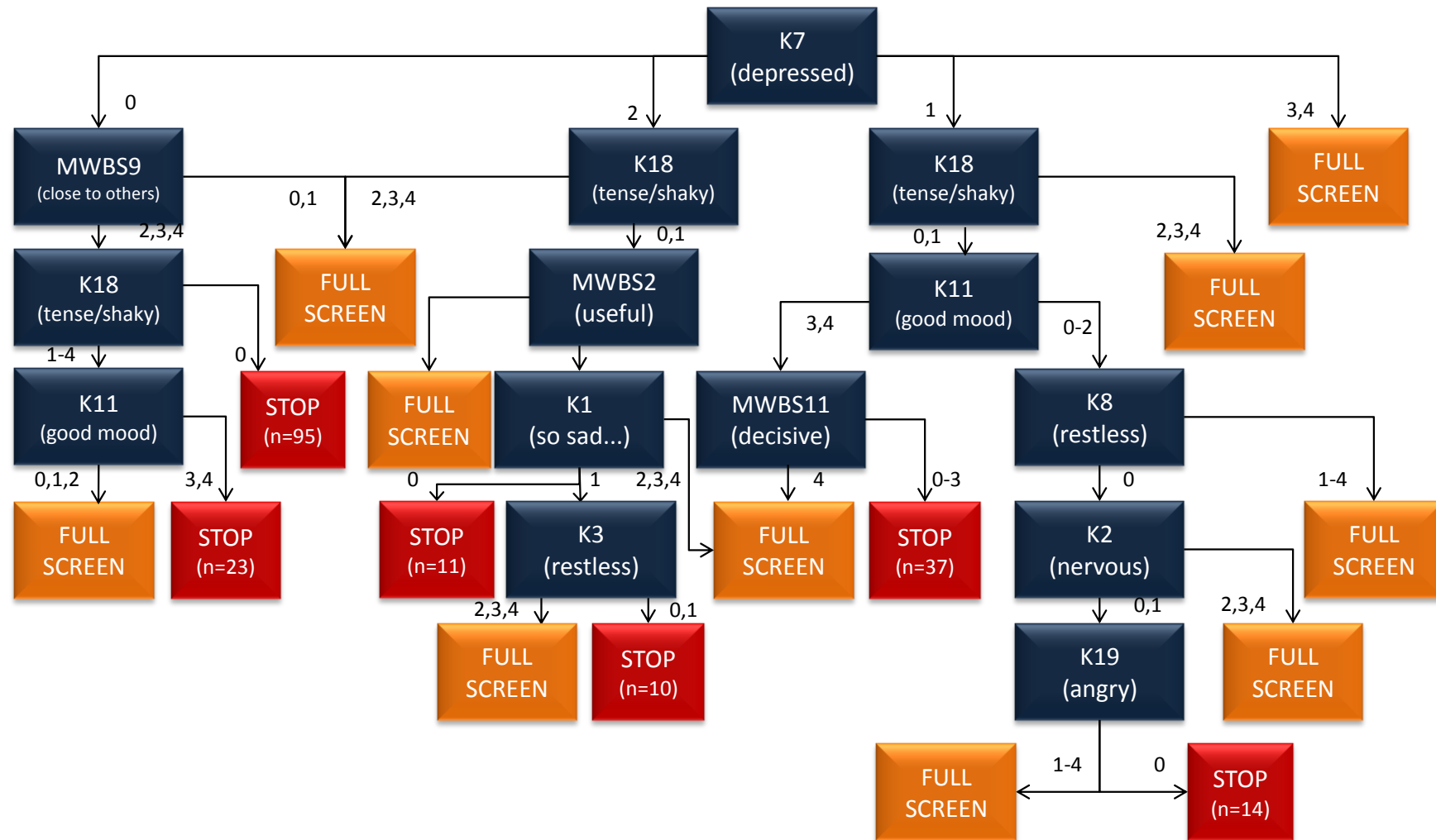
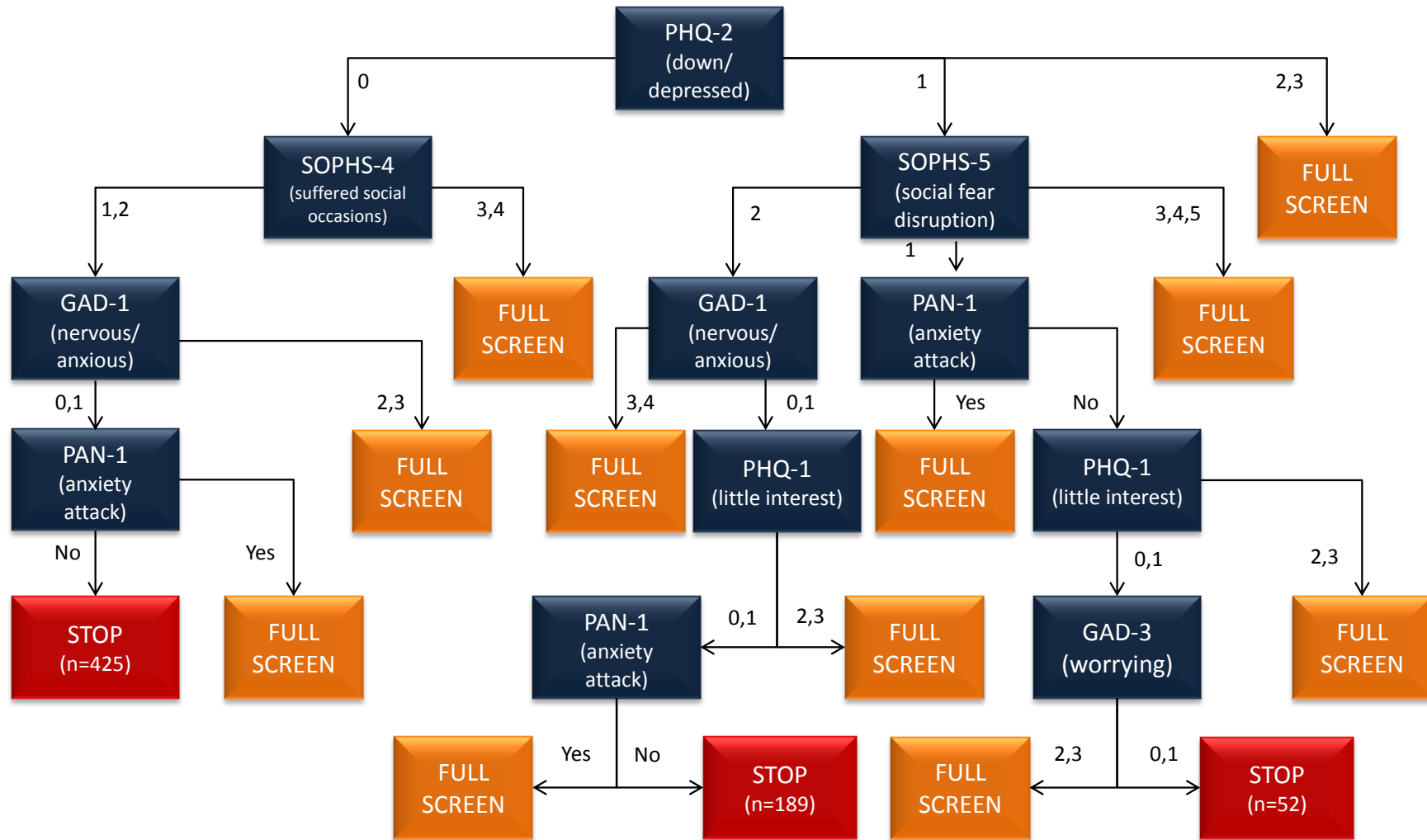


Figure 2: The decision tree used in the Psychological distress hierarchy, developed using Sample 2 (n = 630)



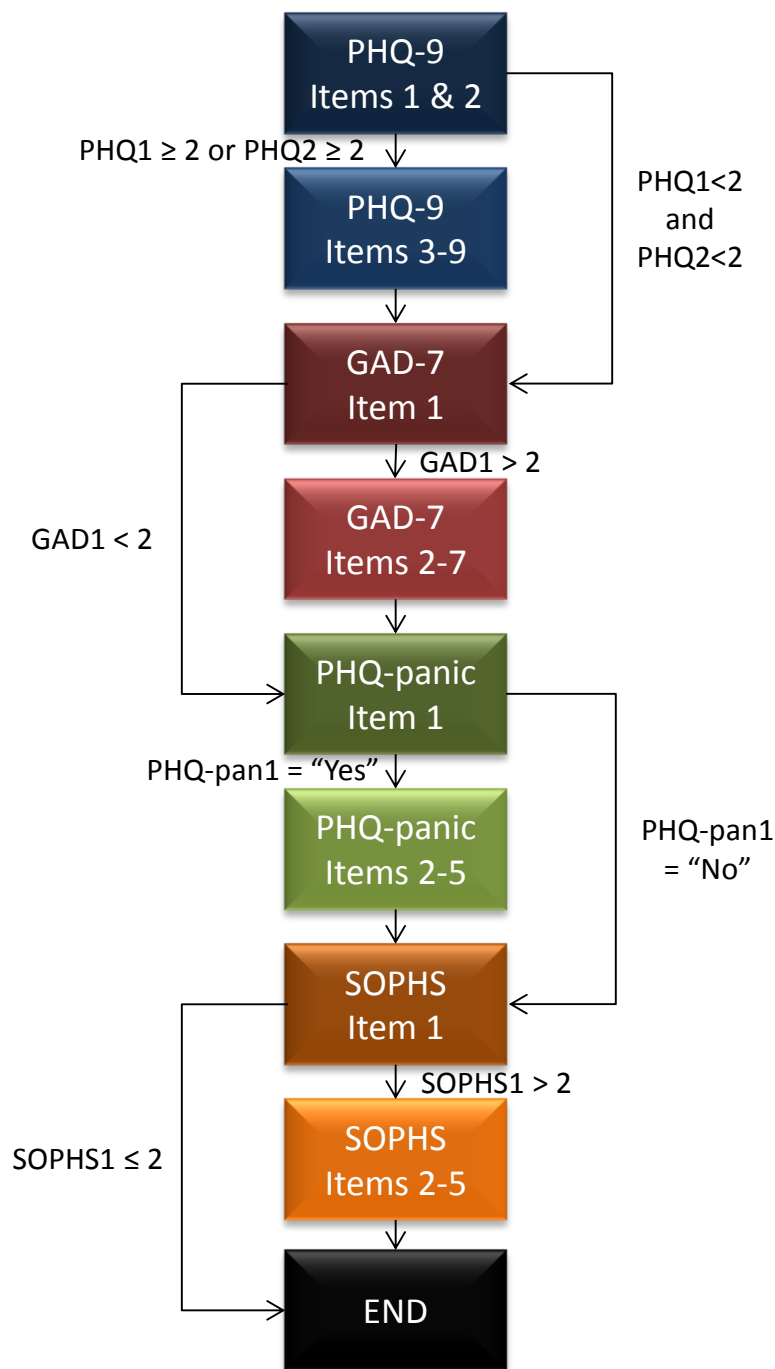
Notes: Items in dark blue boxes from Kessler-10 (K) and Warwick-Edinburgh Mental Well-being Scale (MWBS) with item numbers indicated; item responses for branching are indicated beside arrows (for K10: 0=“none of the time”, 1=“a little of the time”, 2=“some of the time”, 3=“most of the time”, 4=“all of the time”; for MWBS: 0=“none of the time”, 1=“rarely”, 2=“some of the time”, 3=“often”, 4=“all of the time”); “full screen” indicates response patterns that require full screening (PHQ-9, GAD-7, PHQ-panic, SOPHS); “stop” indicate response patterns for which full screening is not required

Figure 3: The decision tree used in the Disorder-based hierarchy, developed using Sample 1 (n = 1360)



Notes: Items in dark blue boxes from PHQ-9 (PHQ), GAD-7 (GAD7), SOPHS (SOPHS) and PHQ-panic (PAN) with item numbers indicated; item responses for branching are indicated beside arrows (for PHQ-9 / GAD-7: 0=“Not at all”, 1=“Several days”, 2=“More than half the days”, 3=“Nearly every day”; for SOPHS: 0=“Not at all”, 1=“A little”, 2=“Moderately”, 3=“Severely”, 4=“Extremely”); “full screen” indicates response patterns that require full screening (i.e., remaining items from PHQ-9, GAD-7, PHQ-panic, SOPHS); “stop” indicates response patterns for which full screening is not required

Figure 4: The gating hierarchy



Notes: item responses for branching are indicated beside arrows (for PHQ-9 / GAD-7: 0="Not at all", 1="Several days", 2="More than half the days", 3="Nearly every day"; for SOPHS: 0="Not at all", 1="A little", 2="Moderately", 3="Severely", 4="Extremely")