The Australian Psychological Society Ltd

Percentile Norms and Accompanying Interval Estimates from an Australian General Adult Population Sample for Self-Report Mood Scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS)

John Crawford,[1] Carol Cayley,[2] Peter F Lovibond,[3] Peter H Wilson,[4] and Caroline Hartley[1]

[1]School of Psychology, University of Aberdeen, Aberdeen, UK; [2]School of Psychology, Flinders University, Adelaide, South Australia, [3]School of Psychology, University of NSW, and [4]University of Kurdistan Hawler, Erbil, Iraq

Despite their widespread use, many self-report mood scales have very limited normative data. To rectify this, Crawford et al. have recently provided percentile norms for a series of self-report scales. The present study aimed to extend the work of Crawford et al. by providing percentile norms for additional mood scales based on samples drawn from the general Australian adult population. Participants completed a series of self-report mood scales. The resultant normative data were incorporated into a computer programme that provides point and interval estimates of the percentile ranks corresponding to raw scores for each of the scales. The programme can be used to obtain point and interval estimates of the percentile ranks of an individual's raw scores on the Beck Anxiety Inventory, the Beck Depression Inventory, the Carroll Rating Scale for Depression, the Centre for Epidemiological Studies Rating Scale for Depression, the Depression, Anxiety, and Stress Scales (DASS), the short-form version of the DASS (DASS-21), the Self-rating Scale for Anxiety, the Self-rating Scale for Depression, the State–Trait Anxiety Inventory (STAI), form X, and the STAI, form Y, based on normative sample sizes ranging from 497 to 769. The interval estimates can be obtained using either classical or Bayesian methods as preferred. The programme (which can be downloaded at http://www.abdn.ac.uk/~psy086/dept/MoodScore_Aus.htm) provides a convenient and reliable means of obtaining the percentile ranks of individuals' raw scores on self-report mood scales.

Key words: anxiety; Bayesian methods; computer scoring; depression; interval estimates; percentile norms; self-report scales.

What is already known on this topic

1. Self-report scales provide a quick and reliable means of assessing mood.
2. Interpretation of scores is currently primarily based on cut-off scores.
3. There is a need for normative data for the general adult Australian population.

What this article adds

1. Percentile norms for self-report mood scales.
2. A method of quantifying the uncertainty over the percentile norms.
3. Guidance on the use of an accompanying computer programme for referring a case's scores to the normative data.

Introduction

Self-report scales are widely used in the assessment of anxiety, depression, and related constructs. They can serve as a useful complement to the clinical interview and are generally quick to administer. They also generally have high internal consistency (i.e., they are reliable) and, unlike clinician rating scales, directly assess an individual's psychological state (Crawford et al., 2009a).

Crawford et al. (2009a) have recently provided UK percentile norms for a range of self-report mood scales. The main aim of the present study was to extend this work by providing percentile norms obtained from samples of the general Australian adult population for further self-report scales. The scales selected were: the Beck Anxiety Inventory (BAI; Beck & Steer, 1993); the Beck Depression Inventory (BDI; Beck & Steer, 1987); the Carroll Rating Scale for Depression (CRSD; Carroll, Feinberg, Smouse, Rawson, & Greden, 1981); the Centre for Epidemiological Studies Rating Scale for Depression (CES-D; Radloff, 1977); the Depression, Anxiety, and Stress scales (DASS; Lovibond & Lovibond, 1995); the short-form version of the DASS, DASS-21(Lovibond & Lovibond, 1995); the Self-Rating Scale for Anxiety (SRAS; Zung, 1971); the Self-Rating Scale for Depression (SRDS; Zung, 1965); the State–Trait Anxiety Inventory (STAI), form X (STAI-X; Spielberger, Gorsuch, & Lushene, 1970); and the STAI, form Y (STAI-Y; Spielberger, Goruch, Lushene, Vagg, & Jacobs, 1983).

The present study was mainly prompted by the scarcity of existing Antipodean normative data for self-report scales. Such data as are available are mainly from special populations such as

Correspondence: John Crawford, School of Psychology, University of Aberdeen, William Guild Building, Aberdeen, AB24 2UB, UK. Fax: +44 (0)1224 273426; email: j.crawford@abdn.ac.uk

Accepted for publication 18 April 2010

doi:10.1111/j.1742-9544.2010.00003.x

students (Lovibond & Lovibond, 1995; Tully, Zajac, & Venning, 2009, the elderly (Collie, Shafiq-Antonacci, Maruff, Tyler, & Currie, 1999), or clinical populations such as oncology patients (Osborne, Elsworth, & Hopper, 2003). Normative data are available for the STAI, SRDS, and BDI obtained from general adult population samples in New Zealand (Knight, 1984; Knight, Waal-Manning, & Spears, 1983) but the sample sizes were relatively modest.

Percentile norms allow clinicians or researchers to quantify the abnormality or rarity of an individual's scores and provide a useful supplement to previously existing cut-off scores. Expressing an individual's score as a percentile rank (rather than simply as being above or below a given cut-off) is in keeping with the commonly held view that anxiety and depression should be treated as dimensional rather than categorical constructs (Crawford, Henry, Crombie, & Taylor, 2001).

In keeping with Crawford et al. (2009a) we decided that, from the point of view of the end-user, the best approach was to develop a computer programme to express the raw scores on the various mood scales as percentile ranks. This approach has the advantage of it being quicker and less error-prone than referring to multiple, voluminous sets of conversion tables. With regard to this last point, research shows that clinicians make more simple clerical errors than we like to imagine when scoring or converting test scores (e.g., Faust, 1998; Sherrets, Gard, & Langner, 1979; Sullivan, 2000).

Also following Crawford et al. (2009a), a second aim of the present study was to provide interval estimates of the percentile ranks corresponding to raw scores on the various scales. When psychologists refer an individual's score to percentile norms, their interest is in the standing (percentile rank) of the individual's score in the normative population, rather than its standing in the particular group of participants who happen to make up the normative sample. Although, in the present case, the normative samples used to provide the basis of conversion from raw scores to percentile ranks were fairly large, it is still the case that there is uncertainty about these quantities. Thus the percentile rank for a raw score obtained from a normative sample must be viewed as a point estimate of the percentile rank of the score in the population and should be accompanied by an interval estimate (Crawford, Garthwaite, & Slick, 2009b). Interval estimates serve the useful general purpose of reminding us that normative data are fallible and serve the specific purpose of quantifying this fallibility (Crawford & Garthwaite, 2002; Gardner & Altman, 1989).

Method

The normative samples were recruited to be broadly representative of the general Australian adult population in terms of the distributions of age, education, and gender. Recruitment took place in Adelaide, South Australia, between 1995 and 2000, and used as wide a variety of sources as was practical, including local and national businesses, public service organisations, community centres, and recreational groups. The majority of participants were recruited from urban/suburban locations, although rural/semirural dwellers were also represented.

Participants were asked to complete the relevant questionnaires and place them in a sealed envelope. The questionnaires were filled in anonymously, that is, the participants were asked not to write their name on any part of the questionnaire or envelope. The questionnaires were either collected at a later date by the investigators or returned by the participants by mail (or, more rarely, by hand). The combined rate of refusals, non-returns, and incomplete returns ranged from 24.6% (for the CES-D scale) to 28.5% (for the Beck scales). Therefore a very substantial majority of those approached took part. Ethical approval was obtained from the Ethics Committee of the Flinders University of South Australia.

Summary statistics (sample size, mean age, and years of education, and so forth) for the normative samples used to generate the percentile norms are presented in Table 1. Table 2 presents the breakdown of the overall sample by age band and gender. The overall sample is defined here as those participants who had data for at least one scale; the *N* for this sample was 785 (396 females, 389 males). In Table 2 the percentages in each cell are compared with the expected percentages for the Australian adult population from census statistics (Australian Bureau of Statistics, 1994). A chi-square goodness-of-fit test comparing the distribution of gender in the sample against the expected values (52% female versus 48% males) revealed that the sample did not differ significantly from the census figures:

Table 1 Summary Statistics for the Samples of the General Australian Adult Population Used to Generate Percentile Norms for the Various Mood Scales

Scale	*n*	(F, M)	Age			Years of education		
			Mean	*SD*	Range	Mean	*SD*	Range
Beck (BDI and BAI)	729	(368, 361)	40.95	16.72	18–90	11.14	1.56	0–14
CRSD	765	(385, 380)	41.26	16.93	18–90	11.12	1.57	0–14
CES-D	769	(387, 382)	41.10	16.83	18–90	11.12	1.56	0–14
DASS	497	(221, 276)	42.14	17.93	18–86	11.07	1.65	0–14
DASS-21	497	(221, 276)	42.14	17.93	18–86	11.07	1.65	0–14
STAI-X	760	(378, 382)	41.19	16.85	18–90	11.13	1.55	0–14
STAI-Y	760	(378, 382)	41.19	16.85	18–90	11.13	1.55	0–14
Zung (SRDS and SRAS)	759	(381, 378)	41.40	16.92	18–90	11.10	1.58	0–14

BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; CRSD = Carroll Rating Scale for Depression; CES-D = Centre for Epidemiological Studies Rating Scale for Depression; DASS = Depression, Anxiety, and Stress Scales; SRSA = Self-Rating Scale for Anxiety; SRDS = Self-Rating Scale for Depression, STAI-X = State–Trait Anxiety Inventory, form X; STAI-Y = State–Trait Anxiety Inventory, form Y.

Table 2 Percentage of Distribution of Age by Gender in the Sample (Obtained) and in the Australian Adult Population (Expected)

Age group (in years)	Males		Females	
	Obtained	Expected	Obtained	Expected
18–24	19.5	15.9	18.2	14.9
25–34	20.6	22.3	21.7	21.6
35–44	23.7	21.0	19.4	20.1
45–54	18.5	15.9	13.1	14.7
55–69	11.3	16.7	17.9	16.5
Over 69	6.4	8.4	9.6	12.0

Table 3 Percentage of Distribution of Educational Level by Gender in the Sample (Obtained) and in the Adult Australian Population (Expected)

Education level	Males		Females	
	Obtained	Expected	Obtained	Expected
<Year 12	26.5	35.0	44.2	45.5
Year 12 or equivalent	29.3	13.2	18.9	13.6
Post-secondary (no degree)	29.6	35.9	22.5	28.2
University degree	14.7	15.9	14.4	12.7

$\chi^2 = 0.760$, $df = 1$, $p = 0.383$. A comparison of the number of females in each age band (see Table 2) against the expected numbers for females from the census data using a goodness-of-fit test was also not statistically significant: $\chi^2 = 5.820$, $df = 5$, $p = 0.324$. The same analysis for males revealed a significant effect ($\chi^2 = 15.239$, $df = 5$, $p = 0.009$). However, the effect size for this comparison was very small (Cramer's V = 0.01), and, as can be seen from Table 2, the fit is reasonably good (with very large samples, even small effects will be significant).

Table 3 presents the breakdown of the sample by years of education and gender, and again the percentages are accompanied by expected percentages for the Australian adult population. It can be seen that the correspondence between the sample and the census figures is reasonable. However, too few males with less than Year 12 education were recruited and too many males with a Year 12 level of education were recruited. Chi-square goodness-of-fit tests revealed that the distribution of education for males differed significantly from the census figures ($\chi^2 = 89.25$, $df = 3$, $p = <0.001$); a significant difference was also observed for females ($\chi^2 = 102.92$, $df = 3$, $p = <0.001$). However, the effect sizes in both cases were small (Cramer's V = 0.28 and 0.29, respectively). Moreover, if the sample is dichotomised into those with up to Year 12 education versus those with more than Year 12 education, the correspondence is good (59% of the sample had up to Year 12 education versus an expected percentage of 54%; 41% had more than Year 12 versus an expected percentage of 46%). In summary, although the fit is not perfect, the distribution of demographic variables in the sample is a reasonable approximation to the distribution in the adult Australian population.

Note that collection of data on the DASS and DASS-21 started later than collection for the other scales. Hence the sample sizes for these two scales are smaller, approximately 500 versus approximately 750 for the other scales (see Table 1). Brief details of the self-report mood scales are presented in the next sections.

The BAI

Each of the 21 items of the BAI describes a common symptom of anxiety. The respondent is asked to rate how much he or she has been bothered by each symptom over the past week on a 4-point scale ranging from 0 (*Not at all*) to 3 (*Severely—I could barely stand it*). Items are summed to obtain a total score which ranges from 0 to 63 (Beck, Epstein, Brown, & Steer, 1988).

The BDI

The BDI is a self-report measure designed to assess depression. It consists of 84 self-evaluative statements grouped into 21 categories. These assess the affective, cognitive, motivational, and physiological symptoms of depression. Items are rated in terms of increasing severity from 0 to 3, with 0 indicating the absence of a particular symptom (Beck & Steer, 1987). Scores for each item are summed, giving a range from 0 to 63.

The CRSD

The CRSD (Carroll et al., 1981) was developed as a direct self-rating adaptation of the 17-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960). Thus, in common with the HRSD, it was designed to be an index of severity of depressive symptomatology, rather than a diagnostic measure. It was developed to address the lack of congruence between clinician and self-ratings of depression, as these were assessed at the time. The items on the CRSD assess behavioural and somatic aspects of depression. Items in the HRSD that are scored 0–4 are represented by four statements, denoting progressively increasing severity of illness. Similarly, those scored 0–2 are represented by two statements. Items are initially scored as 0 for "yes" and 1 for "no," with scores reversed for positively worded items. The response indicative of depression is "yes" for 40 statements and "no" for 12 statements. As for the HRSD, the maximum score is 52. Each statement is scored as one point towards the total score, with a score of 10 or more taken as an indication of clinical depression (Feinberg & Carroll, 1986).

The CES-D

This scale was primarily designed for use in studies of the epidemiology of depressive symptomatology in the general (i.e., non-psychiatric) population. Developed for survey research by Radloff (1977), its purpose differs from those scales designed for diagnosis at clinical intake and/or evaluation of the severity of illness during a course of treatment. Items were generated on the basis of those used in previously validated scales such as the BDI and the SRDS. They were chosen to represent six major components of depression identified in clinical writings and factor analyses: depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance (Radloff, 1977). The CES-D consists of 20 items tapping four content subdomains: depressed affect, positive affect, somatic vegetative signs, and interpersonal distress. Four of the items are worded positively, the remaining 16 negatively. Each item is rated in terms of the frequency with which it has been experienced during the previous week on a scale from 0 to 3: rarely or none of the time (less than 1 day), some or a little of the time (1–2 days), occasionally or a moderate amount of time (3–4 days), and most or all of the time (5–7 days). Higher scores are indicative of greater levels of depression. The range for the CES-D total score is from 0 to 60.

The DASS

The DASS (Lovibond & Lovibond, 1995) is a 42-item self-report measure yielding three scales of 14 items each: depression, anxiety, and stress. Individual items are scored on a 4-point scale (0 to 3). The DASS was designed to maximize discrimination between the three constructs it measures, while at the same time maintaining their breadth. The stress scale is unique among self-report measures, and refers to a syndrome involving tension, irritability, and difficulty relaxing. The DASS has a high reliability, has a factor structure that is consistent with the allocation of the items to subscales, and exhibits high convergent validity with other measures of anxiety and depression in both clinical and community samples (Brown, Chorpita, Korotitsch, & Barlow, 1997; Crawford & Henry, 2003). Total scores on each of the three DASS scales range from 0 to 42.

The DASS-21

The DASS-21 (Lovibond & Lovibond, 1995) is a short-form of the DASS in which each of the three subscales contain seven (rather than 14) items. The DASS-21 has high reliability, has a factor structure that is consistent with the allocation of the items to subscales, and exhibits high convergent validity with other measures of anxiety and depression (Henry & Crawford, 2005). It has a number of advantages over the full-length version. First, and most obviously, it takes less time to complete (and thus is more acceptable to both patients with limited concentration and busy clinicians). Second, the items retained from the full-length version are generally superior to those omitted and, as a result, it has a cleaner factor structure. The downside is that, although the DASS-21 has high reliability, its reliability is a little lower than that of the full-length DASS (Henry & Crawford, 2005). Note that the DASS manual (Lovibond & Lovibond, 1995) recommends that DASS-21 raw scores be doubled to render them comparable to full-length DASS scores. In the present case this was not applied (that is, when using the present norms, it is the simple raw scores on each DASS-21 subscale that should be entered; thus the maximum obtainable scores are 21 on each subscale).

The STAI-X

While the term "anxiety" is most often used to describe an emotional state characterised by subjective feelings of tension, apprehension, nervousness and worry, and by activation or arousal of the autonomic nervous system (Spielberger, 1972; Spielberger et al., 1983), it is also used to describe relatively stable individual differences in anxiety as a personality trait (Spielberger, 1972). The STAI was developed to measure these different constructs.

Form X of the STAI (Spielberger, Gorsuch, & Lushene, 1970) contains 20 state anxiety items and 20 trait anxiety items. The state anxiety items are each rated on a 4-point intensity scale, from 1 for "not at all" to 4 for "very much so." The trait anxiety items are rated on a 4-point frequency scale (from "almost never" to "almost always"). Respondents are asked to indicate how they generally feel. Scoring is reversed for anxiety-absent items (e.g., "I feel calm"). The range of scores for each of the two scales is 20–80.

The STAI-Y

The STAI-Y (Spielberger et al., 1983) is a modification of the original STAI-X in which six items each in the state and trait scales were replaced to (1) balance the number of anxiety-present versus anxiety-absent items and (2) remove items deemed to be non-optimal (e.g., Form X contains the item "I feel blue," which would be more at home in a depression scale). In all other respects, the STAI-Y is the same as the original (e.g., the range of scores is 20–80 for both the state and trait scales).

The Zung SRAS

This self-report scale was developed using what Zung (1971) considered to be the most common characteristics of an anxiety disorder. It consists of 20 items: five affective symptoms and 15 somatic complaints. Subjects are asked to rate each of the 20 items as to how it applied to them within the past week in on a 4-point scale, from 1 for "none or a little of the time" to 4 for "most or all of the time." To reduce the effect of response bias, five items are worded symptom-negatively and the scoring is reversed. Raw scores can be converted to an index score by dividing the total by the maximum possible score (80) and multiplying by 100. Index scores are not used by all researchers, and it is not always clear whether raw or indexed scores are referred to in individual studies. In the present study, the percentile norms provided are for raw scores, not index scores.

The Zung SRDS

This is a 20-item self-report scale designed to provide an assessment of the symptoms of patients whose primary diagnoses were that of a depressive disorder. The items were identified in previous factor analytic studies of depression (Zung, 1965). According to Zung (1986) the content of the SRDS overlaps almost entirely with the HRSD and with Diagnostic and Statistical Manual of Mental Disorders criteria for depression. Ten of the items are worded in a symptomatically positive direction, and 10 in a symptomatically negative direction. In contrast to the BDI, which assesses symptom severity, the SRDS asks respondents to rate the frequency of depressive symptoms on a 4-point scale from 1 for "none or a little of the time," to 4 for "most or all of the time." Scoring for positively worded items is reversed (Zung, 1965). As with the SRAD, raw scores can be converted to an index score. In the present study the percentile norms provided are for raw scores, not index scores.

Indices of GPD

Some of the scales featured in the present study have subscales (i.e., the DASS, DASS-21, STAI-X, and STAI-Y), or partner scales (i.e., the BDI is commonly paired with the BAI while the SRDS is commonly paired with the SRAS). In some circumstances, it would be useful to have a single measure of GPD for each scale (Crawford et al., 2001; Spinhoven et al., 1997); see Cole (1987) and Tanaka and Huba (1984) for further discussion of the rationale for the construct of GPD. Because of this, percentile norms for general distress were derived for each of these mood scales by summing the raw scores from each of their subscales or partner scales.

In the case of the Spielberger STAI, the nature of this general index (whether formed from form X or form Y) is ostensibly different from the other GPDs as it is solely based on measures of anxiety, whereas the others combine anxiety and depression scales. However, even for this scale, a good case can be made that it yields a reasonable index of GPD as, like most other anxiety scales, its correlation with measures of depression approaches (and in some studies even exceeds) its correlation with alternative measures of anxiety (Beiling, Antony, & Swinson, 1998).

Point Estimates of Percentile Ranks

The standard method of obtaining percentile ranks was used (Crawford et al., 2009b; Ley, 1972). That is,

$$\text{Percentile rank} = \left(\frac{m + 0.5k}{N}\right)100, \qquad (1)$$

where m is the number of members of the normative sample obtaining a score lower than the score of interest, k is the number obtaining the score of interest, and N is the overall normative sample size. The percentile ranks thus obtained were then rounded to an integer for scores up to the 99th percentile and to one decimal place for those at or above the 99th percentile. This approach preserves the simplicity of using integer values for percentile ranks that are not very extreme while allowing a distinction to be made between an individual with a score estimated to be exceeded by only one in a hundred members of the general adult population versus one in a thousand.

Interval Estimates of Percentile Ranks

As noted, a further aim of the present study was to accompany the point estimates of the percentile ranks corresponding to raw score with interval estimates of these quantities. A percentile rank is simply a proportion multiplied by 100; thus, methods of obtaining an interval estimate of a proportion (such as classical methods based on the binomial distribution) can be used to obtain interval estimates of a percentile rank. However, for the present problem, there is a complication. Although test scores are discrete (i.e., integer-valued), the underlying mood dimensions they index are generally taken to be continuous, real-valued quantities. Thus, a raw score of, say, 7 is regarded as a point estimate of a real-valued score which could lie anywhere in the interval from 6.5 to 7.4999 (plus an infinite number of additional 9s after the fourth decimal place). Put in another way, in principle, we could distinguish among individuals obtaining the same raw score were we to introduce tie-breaking items. This assumption of a continuous underlying score is ubiquitous in psychological measurement and motivates the standard definition of a percentile rank (equation 1).

Normative data for mood scales will always contain a sizeable number of tied scores; that is, a large number of people in the normative sample will obtain the same raw test scores. Indeed, if the normative sample is very large and the data are heavily skewed (as is usually the case), then there could literally be hundreds of such ties for a given raw score. The present problem therefore differs from those dealt with by standard binomial sampling in which there can be no possibility of multiple ties.

Crawford et al. (2009b) have recently developed classical and Bayesian methods that incorporate the additional uncertainty arising from tied scores. Crawford et al. (2009a) used this method to provide interval estimates for mood scales, and we also apply it to the present data. To illustrate, suppose that in a normative sample of 100 people, 89 obtained lower scores than a case and two obtained the same score as the case. Then the point estimate of the percentile rank for the case's score (using equation 1) is 90, and applying Crawford et al.'s (2009b) classical method, the interval estimate is from 82.15 to 95.27. Suppose, however, that 85 obtained lower scores and 10 obtained the same score. The point estimate of the percentile rank is the same as in the foregoing example (90), but the interval estimate is from 79.79 to 97.10; the latter interval is wider because of the increased uncertainty introduced by the larger number of ties (2 vs 10).

The classical method developed by Crawford et al. (2009b) is an extension of the standard Clopper–Pearson method (Clopper & Pearson, 1934) of obtaining interval estimates on a proportion to the situation in which there are ties (i.e., some members of the normative sample obtain the same raw score as a case). In practice, a mid-p variant of the Clopper–Pearson method is often used in the standard situation where ties are not an issue, because the Clopper–Pearson interval is quite commonly regarded as being too conservative (Brown, Cai, & DasGupta, 2001). For this reason, Crawford et al. also offered a mid-p variant of their classical method. For the second numerical example used earlier (in which out of 100 members of a normative sample, 85 scored below a case's score and 10 obtained the same score), the mid-p interval estimate of the percentile rank is 80.30 to 96.80. It can be seen that this interval is narrower, that is, less conservative than that obtained earlier (79.79 to 97.10).

Crawford et al. (2009b) also developed a Bayesian method of obtaining an interval estimate on a percentile rank in the presence of tied scores; the methods are based on a mixture of beta distributions; see Crawford et al. for the technical details. To illustrate the Bayesian intervals using the previous examples, the Bayesian interval estimate of the percentile rank for the first example (89 out of 100 scoring below and two at the score) is from 82.99 to 94.74. For the second example (85 below and 10 at the score) the interval estimate is 80.36 to 96.74. These results demonstrate three features of the intervals. First, as was the case for the classical methods, the Bayesian method captures the greater uncertainty in the second example and hence produces a wider interval. Second, it can be seen that the Bayesian and classical intervals show a reasonable degree of convergence (the convergence between the Bayesian and classical approaches being particularly close for the classical mid-*p* variant). Indeed the degree of convergence is very close indeed with samples as large as those employed to provide the normative data for the present study. This convergence is reassuring regardless of whether one is classical, Bayesian, or eclectic in orientation. Third, it can be seen that, when the sample size providing normative data is modest (as in the previous example where $n = 100$), there is considerable uncertainty over the percentile rank of a case's raw score.

One-sided Versus Two-sided Intervals

In practice, there will be occasions in which a one-sided interval may be preferred over a two-sided interval. For example, a clinician may be interested in whether a case's score is less extreme than is indicated by the point estimate but not particularly interested in whether the score is even more extreme (or vice versa). Both the classical and Bayesian methods developed by Crawford et al. (2009b) are easily adapted to provide a one-sided limit. However, without prior knowledge of which limit is of interest (the situation here, as the aim is to provide intervals for use by others) it is more convenient to generate $100(1 - [\alpha/2])$ two-sided intervals that then provide $100(1 - \alpha)$ one-sided lower and upper limits. For example, if a 95% lower limit on the percentile rank is required, then a 90% two-sided interval is generated: The user then simply disregards the upper limit of the two-sided interval and treats the lower limit as the desired one-sided 95% limit.

Computer Programme for Obtaining Point and Interval Estimates of Percentile Ranks for Raw Scores on the Various Mood Scales

The methods for obtaining interval estimates for a percentile rank developed by Crawford et al. (2009b) are complex and time-consuming to calculate. Moreover, for the problem at hand, these limits need to be provided for the percentile ranks corresponding to all possible raw scores for each of the mood scales. It therefore makes sense to implement the methods into a computer programme so that the limits can be obtained quickly and accurately. As noted previously, the use of a computer programme also does away with the need to consult voluminous sets of tables

Results

Summary Statistics and Reliabilities of the Mood Scales

The summary statistics (mean, median, standard deviation (*SD*), and range) for the mood scales are presented in Table 4. The equivalent data for the total scores on these scales (used as indices of GPD) are presented in Table 5. Tables 4 and 5 also present reliability coefficients (Cronbach's alpha) for the scales calculated from the overall samples; 95% confidence limits on these alphas were computed using Feldt's (1965) formula. It can be seen that the reliabilities of the mood scales are generally acceptable and in many cases are very high. The reliabilities ranged from a low of 0.83, for the SRAS, to a high of 0.95 for the depression scale of the DASS. It can also be seen from the narrowness of the confidence limits on these reliability coefficients that they provide accurate estimates of the true reliability of the scales.

Effects of Demographic Variables on Mood Scores

The relationships between the mood scales and demographic variables were examined by computing the Pearson product–moment correlation between scores on each of the scales and age, years of education, and gender (males coded as 0, females as 1); the latter set of coefficients are termed point biserial

Table 4 Summary Statistics for the Mood Sales

Scale	Subscale	Mean	Median	*SD*	Range	α (95% CLs)
Beck	Depression	6.25	4	6.94	0–42	0.89 (0.88, 0.90)
	Anxiety	6.16	4	7.16	0–51	0.90 (0.89, 0.91)
CRSD	Depression	7.94	6	7.33	0–45	0.90 (0.89, 0.91)
CES-D	Depression	10.24	8	9.67	0–54	0.90 (0.89, 0.91)
DASS	Depression	5.02	2	7.54	0–39	0.95 (0.94, 0.96)
	Anxiety	3.36	1	5.07	0–31	0.88 (0.86, 0.90
	Stress	8.10	5	8.40	0–42	0.94 (0.93, 0.95)
DASS-21	Depression	2.57	1	3.86	0–20	0.90 (0.89, 0.91)
	Anxiety	1.74	1	2.78	0–17	0.79 (0.76, 0.82)
	Stress	3.99	3	4.24	0–21	0.89 (0.88, 0.90)
STAI-X	State Anxiety	33.91	31	11.75	20–77	0.94 (0.93, 0.95)
	Trait Anxiety	36.44	34	10.93	20–73	0.91 (0.90, 0.92)
STAI-Y	State Anxiety	33.16	30	11.69	20–77	0.94 (0.93, 0.95)
	Trait Anxiety	36.35	34	11.39	20–78	0.94 (0.93, 0.95)
Zung	Depression	34.16	33	9.11	20–71	0.86 (0.85, 0.87)
	Anxiety	31.52	30	7.72	20–69	0.83 (0.81, 0.85)

α = Cronbach's alpha; CLs = confidence limits; CRSD = Carroll Rating Scale for Depression; CES-D = Centre for Epidemiological Studies Rating Scale for Depression; DASS = Depression, Anxiety, and Stress Scales; STAI-X = State–Trait Anxiety Inventory, form X; STAI-Y = State–Trait Anxiety Inventory, form Y.

Table 5 Summary Statistics for Total Scores (General Psychological Distress (GPD)) on the Mood Scales

Scale	Mean	Median	SD	Range	α (95% CLs)
Beck GPD	12.41	8	12.97	0–78	0.94 (0.93, 0.95)
DASS GPD	16.48	10	19.24	0–107	0.97 (0.97, 0.97)
DASS-21 GPD	8.30	5	9.83	0–54	0.94 (0.93, 0.95)
STAI-X GPD	70.35	65	21.55	40–146	0.95 (0.95, 0.96)
STAI-Y GPD	69.51	64	21.88	40–154	0.96 (0.96, 0.96)
Zung GPD	65.68	62	15.65	41–136	0.91 (0.87, 0.90)

Notes. α = Cronbach's alpha
CLs = confidence limits; DASS = Depression, Anxiety, and Stress Scales; STAI-X = State–Trait Anxiety Inventory, form X; STAI-Y = State–Trait Anxiety Inventory, form Y.

Table 6 Correlations Among the Scales and Demographic Variables. The final column reports the correlations for age when 18–24-year olds were excluded (these latter correlations are only reported for scales exhibiting an age effect in the overall sample).

Scale	Subscale	Education	Gender	Age	Age (25–90)
Beck	Depression	–0.02	0.10	–0.12	—
	Anxiety	0.01	0.15	–0.13	—
CRSD	Depression	–0.05	0.11	–0.11	—
CES-D	Depression	–0.01	0.12	–0.16	–0.06
DASS	Depression	0.06	0.07	–0.20	–0.13
	Anxiety	–0.05	0.10	–0.14	–0.01
	Stress	0.07	0.06	–0.18	–0.16
DASS-21	Depression	0.06	0.06	–0.19	–0.12
	Anxiety	–0.05	0.11	–0.11	0.03
	Stress	0.07	0.07	–0.17	–0.16
STAI-X	State anxiety	0.04	0.08	–0.19	–0.15
	Trait anxiety	0.03	0.12	–0.19	–0.12
STAI-Y	State anxiety	0.04	0.10	–0.18	–0.13
	Trait anxiety	0.05	0.10	–0.21	–0.14
Zung	Depression	–0.09	0.16	–0.07	—
	Anxiety	–0.07	0.21	–0.07	—

CRSD = Carroll Rating Scale for Depression; CES-D = Centre for Epidemiological Studies Rating Scale for Depression; DASS = Depression, Anxiety, and Stress scales; STAI-X = State–Trait Anxiety Inventory, form X; STAI-Y = State–Trait Anxiety Inventory, form Y.

correlation coefficients but are computed in the same way as the Pearson coefficient. The correlations are presented in Table 6.

It can be seen from Table 6 that, for all scales, the correlations with years of education were very modest, ranging from a minimum of 0.01 (for the BAI scale) to a maximum of –0.09 (for the SRAS). These results indicate that there was little or nothing to be gained by calculating percentile ranks separately for different levels of education (and much to be lost in terms of reducing the sample sizes on which the norms would be based).

For most scales the (point biserial) correlations with gender were also modest. Excluding the Zung scales (see later), the correlations ranged from a minimum of 0.06 (for the stress scale of the DASS and depression scale of the DASS-21) to a maximum of 0.15 (for the BAI). Therefore it was decided that there was little to be gained by presenting percentile norms separately for males and females for these scales. In the case of the Zung scales, particularly the anxiety scale, it can be seen that the correlation with gender, although still relatively modest, is of concern ($r = 0.21$ for the anxiety scale; females scored higher). As a result, percentile norms were generated separately for males and females on these latter scales

Turning to the relationship between the scales and age, it can be seen that although the correlations were all relatively modest, for some scales (the CES-D, DASS, DASS-21, STAI-X, and STAI-Y) there were appreciable age effects; the correlations range from a minimum of –0.07 (for both Zung scales) to a maximum of –0.21 (for the STAI-X trait scale). Examination of scatter plots and box plots revealed that much of this age effect was attributable to younger participants (those aged between 18 and 24 years) scoring higher on the scales. The last column of Table 6 records the correlations between the scales and age when this age group was excluded from the analysis; it can be seen that the correlations are reduced in magnitude. On the basis of these findings it was decided to generate separate percentile norms for two age groups (18–24 years and 25–90 years) for the aforementioned scales.

Table 7 presents summary statistics for the scales for the two age groups; data are presented only for those scales in which norms were generated separately for these age groups. Table 8 presents summary statistics for the Zung scales (SRAS and SRDS) for males and females (as noted, the Zung scales were the only scales exhibiting appreciable gender effects).

Obtaining Point and Interval Estimates of the Percentile Ranks for Raw Scores

As previously noted, a computer programme for personal computers, MoodScore_PRs_Aus.exe was written (using the Delphi programming language) to express an individual's raw scores on the various scales as percentile ranks. The programme is free and can be downloaded (either as an uncompressed executable or as a zip file) from the first author's web pages at http://www.abdn.ac.uk/~psy086/dept/MoodScore_PRs_Aus.htm. (After downloading the programme it can be run by clicking on the programme in the web browser, or, if a shortcut to the programme has been created on the user's desktop, by clicking on the shortcut icon).

The programme prompts the user to select the scale they wish to score. The user is then prompted to enter the individual's raw scores for the selected mood scale. When using the Zung scales (SRAS and SRDS), the user also needs to identify the individual's gender through the use of radio buttons as the normative data are organised separately for females and males. In the case of scales that exhibited age effects, the user selects the individual's age group using a further set of radio buttons. Finally, there is the option of entering identifying information for the individual (in the form of user's notes) for future reference.

The output from the programme consists of a brief listing (resembling that in Tables 4 and 5) of the summary statistics for the scales, that is, the mean, median, *SD*, range, and reliability (the default is to suppress this listing; if it is required it can be selected with the appropriate radio button). The summary data

Table 7 Summary Statistics for Those Scales in which Normative Data are Provided Separately for Two Age Groups

Scale	Subscale	18–24 years of age					25–90 years of age				
		n	*M*	*Mn*	*SD*	Range	*n*	*M*	*Mn*	*SD*	Range
CES-D	Depression	145	14.08	12	10.91	0–42	624	9.35	7	9.13	0–54
DASS	Depression	102	7.75	5	8.87	0–39	395	4.31	1	7.00	0–38
	Anxiety	102	5.34	4	6.16	0–29	395	2.85	1	4.63	0–31
	Stress	102	10.04	7	9.37	0–42	395	7.61	5	8.07	0–41
	GPD	102	23.14	16.5	22.65	0–107	395	14.76	8	17.90	0–101
DASS-21	Depression	102	3.96	3	4.52	0–20	395	2.21	1	3.60	0–19
	Anxiety	102	2.76	2	3.25	0–15	395	1.48	0	2.60	0–17
	Stress	102	4.78	3	4.71	0–21	395	3.79	3	4.10	0–21
	GPD	102	11.51	8.5	11.51	0–54	395	7.48	4	9.18	0–52
STAI-X	State	142	37.56	37	13.10	20–74	618	33.07	30	11.27	20–77
	Trait	142	40.11	38	11.56	20–73	618	35.60	34	10.62	20–73
	GPD	142	77.66	72.5	23.67	40–144	618	68.67	64	20.69	40–146
STAI-Y	State	142	36.85	34.5	13.56	20–76	618	32.31	29	11.06	20–77
	Trait	142	40.39	38	12.32	20–78	618	35.42	33	10.96	20–74
	GPD	142	77.23	73	24.87	40–154	618	67.74	63	20.75	40–146

CES-D = Centre for Epidemiological Studies Rating Scale for Depression; DASS = Depression, Anxiety, and Stress Scales; GPD = general psychological distress; STAI-X = State–Trait Anxiety Inventory, form X; STAI-Y = State–Trait Anxiety Inventory, form Y.

Table 8 Summary Statistics for Females and Males on the Zung Scales (SRDS and SRAS)

Scale	Subscale	Females (n = 381)				Males (n = 378)			
		M	*Mn*	*SD*	Range	*M*	*Mn*	*SD*	Range
Zung	Depression	35.62	34	9.66	20–71	32.69	31	8.27	20–60
	Anxiety	33.12	32	8.15	20.69	29.90	29	6.90	20–65
	GPD	68.74	66	16.57	41–136	62.60	60	14.02	41–123

GPD = general psychological distress.

are followed by the User's Notes, if these have been entered, and the point and interval estimates of the percentile ranks for the individual's raw scores. These results can be viewed on screen, saved to a file, and/or printed.

As noted, the provision of a computer programme through which clinicians or researchers can obtain point and interval estimates of the percentile ranks for an individual's scores has advantages over the alternative of consulting voluminous sets of tables. However, to illustrate the mapping of raw scores to percentile ranks and the accompanying interval estimates, Table 9 provides these data for the BDI (interval estimates having been calculated using the Bayesian method).

Discussion

Despite the widespread use of self-report mood scales, percentile norms from large samples of the general adult population have not been available for these scales until recently (Crawford et al., 2009a). We suggest that the normative data and accompanying computer programme described in the present article provide an additional useful and convenient resource for clinical research and practice.

A Worked Example

The computer programme accompanying this article was designed to be intuitive to use but a brief example may be helpful. Suppose an individual has completed the BDI and BAI and obtains raw scores of 8 and 27, respectively. A screen capture of the programme set up to process these scores is presented in Figure 1. It can be seen that (1) the BDI–BAI pairing has been selected using the appropriate radio button, (2) the radio buttons for recording the individual's gender and age are disabled (because gender and age effects were minimal for these scales), (3) Bayesian (rather than classical) interval estimates of the percentile ranks have been requested (this being the default option), (4) the interval width has not been changed from its default value of 95%, (5) the user has requested a listing of the summary data (the default is to suppress this listing) and (6) the user has entered the individual's raw scores into their respective data fields. Note that if an individual had only been administered one of these scales, the data field for the other can simply be left blank.

From this setup, the results would then be obtained by clicking on the compute button. The results screen (not shown) provides the point estimates of the percentile ranks for the

Table 9 Conversion of Raw Scores on the Beck Depression Inventory to Percentile Ranks (PR) with Accompanying 95% Two-Sided Interval Estimate (95% CI). The interval estimates were obtained using Crawford et al.'s (2009b) Bayesian method.

Raw score	PR	95% CI	Raw score	PR	95% CI
0	7	0 to 15	32	99	98 to 99
1	20	14 to 26	33	99	98 to 99
2	30	24 to36	34	99	98 to 99
3	39	33 to 44	35	99	98 to 100
4	47	41 to 54	36	99.2	98.3 to 99.7
5	57	51 to 63	37	99.4	98.6 to 99.8
6	64	59 to 68	38	99.5	98.7 to 99.8
7	69	64 to 73	39	99.7	98.9 to 100
8	74	70 to 78	40	99.9	99.4 to 100
9	78	74 to 81	41	99.9	99.4 to 100
10	80	77 to 84	42	99.9	99.4 to 100
11	83	80 to 86	43	>99.9	99.7 to 100
12	85	82 to 88	44	>99.9	99.7 to 100
13	87	84 to 89	45	>99.9	99.7 to 100
14	88	86 to 91	46	>99.9	99.7 to 100
15	90	88 to 92	47	>99.9	99.7 to 100
16	92	89 to 94	48	>99.9	99.7 to 100
17	93	91 to 94	49	>99.9	99.7 to 100
18	93	91 to 95	50	>99.9	99.7 to 100
19	94	92 to 95	51	>99.9	99.7 to 100
20	94	92 to 96	52	>99.9	99.7 to 100
21	95	93 to 96	53	>99.9	99.7 to 100
22	96	94 to 97	54	>99.9	99.7 to 100
23	96	95 to 98	55	>99.9	99.7 to 100
24	97	95 to 98	56	>99.9	99.7 to 100
25	97	96 to 98	57	>99.9	99.7 to 100
26	98	96 to 99	58	>99.9	99.7 to 100
27	98	97 to 99	59	>99.9	99.7 to 100
28	98	97 to 99	60	>99.9	99.7 to 100
29	98	97 to 99	61	>99.9	99.7 to 100
30	98	97 to 99	62	>99.9	99.7 to 100
31	98	97 to 99	63	>99.9	99.7 to 100

individual's scores with their accompanying interval estimates: The percentile rank of the BDI score is 74 (95% confidence interval (CI) = 70 to 78), and the percentile rank of the BAI score is 98 (95% CI = 97 to 99).

In this example the (BDI) depression score is elevated relative to the average score for the general adult population (i.e., the score is above the 50th percentile) but is not that unusual. That is, a sizeable percentage of the general adult population (26%) would be expected to obtain higher scores. In contrast, the (BAI) anxiety score is extreme: 98% of the population is expected to have scores lower than the individual (and thus only 2% would be expected to score higher). Finally, the percentile rank of the individual's total (GPD) score (i.e., the sum of raw scores on the BDI and BAI scales) is 94 (95% CI = 92 to 95). Note that, in this specific and hypothetical case, the total score is of questionable utility given the discrepancies between the percentile ranks for the two scales.

The interval estimates of the percentile ranks in this example are fairly narrow, thereby indicating that the point estimates of percentile ranks of the raw scores obtained using the normative sample provide a fairly accurate estimate of the true percentile ranks of these raw scores in the population. This is generally the case for all of the mood scales included here. However, there will be considerably more uncertainty attached to percentile ranks for young individuals (18–24 years) on those scales that show age effects because the normative samples for this age group are smaller than for the older age group (25–90 years). Indeed it would be worth attempting to recruit larger samples for this younger age group in future work.

A further exception to the general rule that the interval estimates of the percentile rank are narrow for these normative data occurs for scores that are very low (regardless of which scale is used). For example, in the case of the BDI (see Table 9) the interval estimate of the percentile rank for a raw score of 0 is wide (from 0 to 15, point estimate = 7). This occurs because a large number of the normative sample obtained a score of 0 (that is, there was a large number of ties) and thus a high degree of uncertainty over an individual's percentile rank. It will be appreciated, however, that this is not of much practical concern as there is little need for a precise estimate for very low scores.

In interpreting the percentile ranks obtained from the present normative data, it is important to stress that a percentile rank does not have to be very extreme to be a potential cause for concern. For example, Shepherd, Cooper, Brown, and Kalton (1966) reported that between 30% and 40% of the general adult population in the UK suffer from anxiety to an extent that would benefit from clinical intervention.

Although point prevalence estimates for anxiety and depression in the general population vary from study to study, it is also clear that a sizeable percentage exhibit symptoms severe enough to warrant a clinical diagnosis (see Crawford et al., 2001 for a brief review). In Meltzer, Gill, Petticrew, and Hinds' (1995) survey of 10,000 UK households, the (1-week) prevalence of anxiety disorders was 13.9% (this percentage is based on the inclusion of cases diagnosed as mixed anxiety/depression). In Australia, the most recent (2007) National Survey of Mental Health and Wellbeing reported a 12-month prevalence rate for anxiety disorders of 14.4% (Australian Bureau of Statistics, 2007).

Point prevalence rates for depression tend to be lower but have been commonly reported to be around 3% to 4% (Horwath & Weissman, 1995). In the Australian survey previously referred to, the 12-month prevalence rates for affective disorder (defined as depressive episode, dysthymia, or bipolar affective disorder) was 6.2% (Australian Bureau of Statistics, 2007).

Bayesian Versus Classical Interpretations of the Interval Estimate on a Score's Percentile Rank

As Antelman (1997) notes, the frequentist (classical) conception of a CI is that, "It is one interval generated by a procedure that will give correct intervals 95% of the time. Whether or not the one (and only) interval you happened to get is correct or not is unknown" (p. 375). Thus, in the present context, the classical interpretation of the interval estimate on the percentile rank for a raw score on, say, the BAI is as follows, "if we could compute a CI for each of a large number of normative samples collected

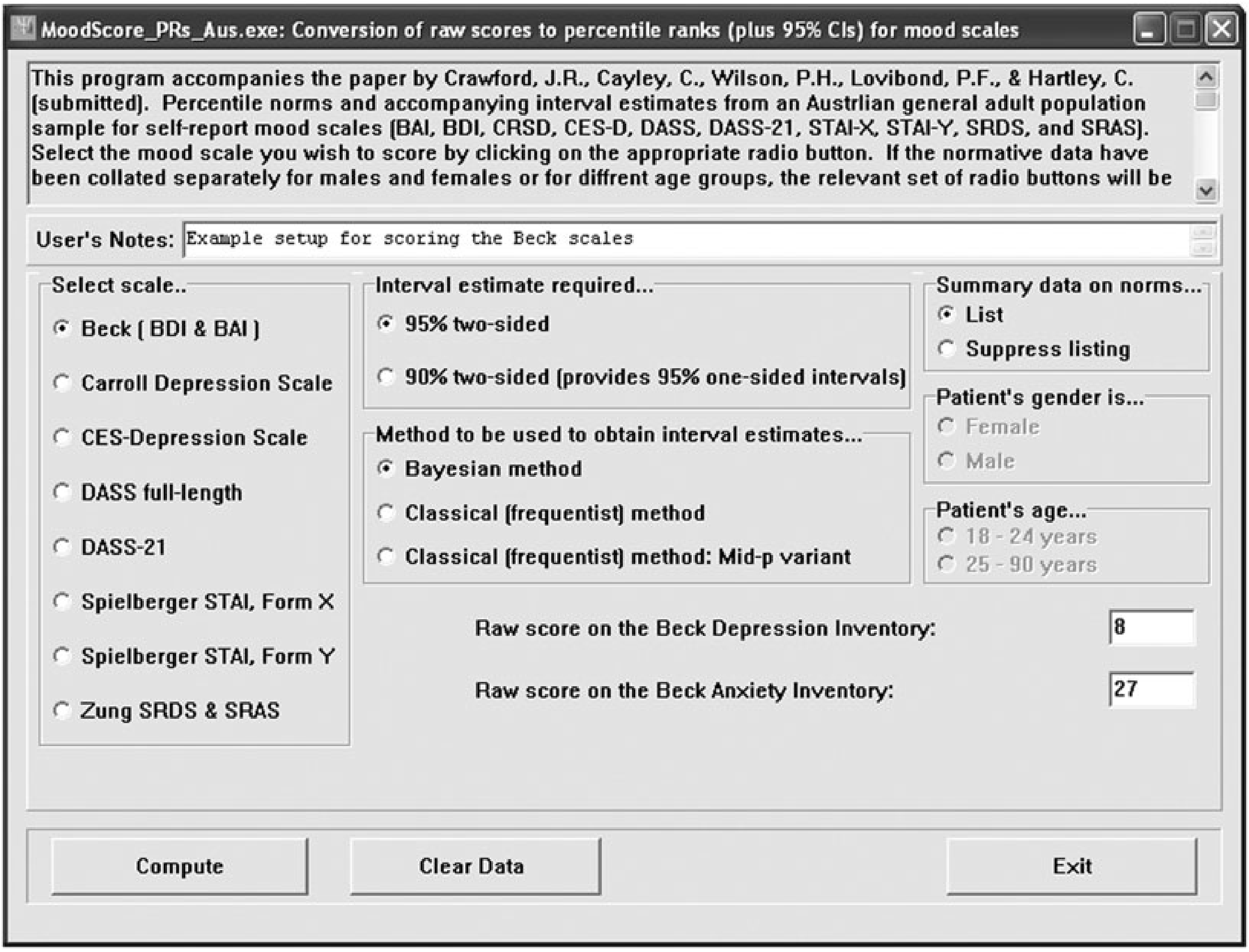

Figure 1 Screen Capture of the Computer Programme (MoodScore_PRs_Aus.exe) Used to Express Raw Scores on Various Commonly Used Self-report Mood Scales as Percentile Ranks (With Accompanying Interval Estimates). In this example, the program is set up to score the Beck scales (Beck Anxiety Inventory and Beck Depression Inventory), and a 95% (two-sided) Bayesian interval estimate has been selected.

in the same way as the present BAI normative sample, about 95% of these intervals would contain the true percentile rank of the individual's score."

The Bayesian interpretation of such an interval is "there is a 95% probability that the true percentile rank of the individual's score lies within the stated interval." This statement is not only less convoluted, but it also captures what a clinician would wish to conclude from an interval estimate (Crawford & Garthwaite, 2007). Indeed most psychologists who use frequentist confidence limits probably construe these in what are essentially Bayesian terms (Howell, 2002).

For the present problem, the frequentist and Bayesian approaches exhibit a high degree of convergence (Crawford et al., 2009b). This can readily be verified by the reader by comparing the two sets of interval estimates for scores on the mood scales featured in the present article. For instance, the intervals for the BDI/BAI example reported earlier (in which an individual obtained scores of 8 and 27 on the BDI and BAI, respectively) were calculated using the Bayesian method. However, the classical mid-*p* intervals are identical. The upshot is that psychologists can place a Bayesian interpretation on the interval estimate of a percentile rank, regardless of whether it was obtained using the Bayesian or classical methods (the classical methods were made available to cater for those who are strongly wedded to the classical approach to inference).

CIs Capturing Sampling Error Versus Measurement Error

The CIs on the percentile ranks provided here (whether obtained using classical or Bayesian methods) should not be confused with confidence limits derived from classical test theory that attempt to capture the effects of measurement error on an individual's score (Crawford & Garthwaite, 2009).

When the latter intervals are used, the clinician is posing the question "assuming scores are normally distributed, and assuming no error in estimating the population mean, *SD* and reliability coefficient of the test, how much uncertainty is there over an individual's score as a function of measurement error in the scale?" (Crawford & Garthwaite, 2008). In contrast, when using the intervals presented in the present article, the concern is solely with the score *in hand*. The more concrete question posed is "how much uncertainty is there over the standing (i.e., percentile rank) of the score the individual obtained as a function of error in using a normative sample to estimate its standing in the normative population?" That is, they do not address the issue of what score an individual might obtain on another occasion, or on a set of alternative, parallel items, but simply provide interval estimates for the percentage of the normative population who would score below the score obtained by the individual (Crawford et al., 2009b).

Provision of Comparison Standards for Group Studies

The emphasis in the current article has been on the use of these normative data with individual cases. However, they have other potential uses. For example, many studies of anxiety and depression in clinical populations (or, in particular, occupational groups and so forth) do not collect control data from a sample drawn from the general adult population (Crawford et al., 2009a). The summary statistics provided here (Tables 4 and 5) could serve as useful comparison standards against which to compare the means or medians of clinical samples. Alternatively, the median score obtained in a clinical sample could be expressed in terms of its percentile rank in the general population sample using the computer programme described earlier (the interval estimates provided should be ignored in such an application as they quantify the uncertainty over the standing of an individual's score rather than the average score of a second sample).

Future Developments

The normative data featured in the present article have all been gathered from samples of the general Australian adult population. However, it would be useful for clinicians if point and interval estimates for the percentile ranks of mood scores were also available for clinical populations. These would provide further context when evaluating an individual's score. Percentile norms could be gathered for clinical populations encountered in general medicine (e.g., in cardiology, oncology, and diabetic medicine) and for neurological populations (traumatic brain injury (TBI), stroke, etc.) as well as in mental health settings (Crawford et al., 2009a). For example, in the case of the BDI/BAI scores of an individual who has suffered a TBI, it would be useful to be able to obtain an estimate of how unusual or otherwise these scores are in the population of patients who have suffered a TBI. We would welcome collaborators in such a development: by pooling data it would be possible to provide clinicians with a single, reliable, and convenient source of normative data for a wide range of scales and populations.

In gathering such data, it would be important to supplement basic demographic summary information (such as that provided here for the general population samples) with summary data on the clinical characteristics of the samples. The nature of such additional background information will vary with the particular clinical condition. For example, in the case of TBI, the provision of summary information (means, medians, *SD*s) for measures of injury severity and time since injury, and so on would be useful. Relatedly, data gathering should not be limited to obtaining a single sample (however large) of patients with a particular clinical condition. Rather as many samples should be obtained as possible so that a psychologist could select the sample that most closely resembles the clinical characteristics of her or his particular client.

Some researchers may prefer to make normative data available separately, rather than pooling the data as suggested above. Given the importance placed on the use of interval estimates in contemporary biometry, we suggest that any such normative exercise should provide both point and interval estimates for the resultant percentile ranks. It would be particularly important to provide interval estimates if the normative samples were modest in size.

Conclusion

The present normative data and accompanying computer programme provide a quick and reliable means of obtaining percentile norms for a range of widely used self-report mood scales. The percentile norms allow clinicians to quantify the rarity or otherwise of an individual's score and are therefore a useful supplement to the traditional cut-off scores available for some (but not all) of these scales. Expressing an individual's score in terms of its percentile rank, rather than simply as below or above a cut-off, is also in keeping with a conception of anxiety and depression as dimensional rather than categorical constructs (Crawford et al., 2009a). Finally, the provision of interval estimates for the percentile rank of a score serves the general purpose of reminding clinicians that all normative data are fallible. It also serves the specific and practical purpose of quantifying the uncertainty over the standing of an individual's score when referred to such data.

Acknowledgements

This research was partially funded by a research grant awarded to the first author by the Research Board of the Flinders University of South Australia.

References

Antelman, G. (1997). *Elementary Bayesian statistics*. Cheltenam: Elgar.

Australian Bureau of Statistics (1994). *Estimated resident population by sex and age: States and territories of Australia. Preliminary June 1992 and June 1993. (Catalogue No. 3201.0)*. Canberra: ABS.

Australian Bureau of Statistics (2007). *National survey of health and wellbeing: Summary of results. (Catalogue No. 4326.0)*. Canberra: ABS.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*, 893–897.

Beck, A. T., & Steer, R. A. (1987). *Beck depression inventory manual*. San Antonio TX: The Psychological Corporation, Harcourt Brace Jovanovich.

Beck, A. T., & Steer, R. A. (1993). *Beck anxiety inventory manual*. San Antonio, TX: The Psychological Corporation.

Beiling, P. J., Antony, M. M., & Swinson, R. P. (1998). The state–trait anxiety inventory, trait version: Structure and content re-examined. *Behaviour Research and Therapy*, *36*, 777–788.

Brown, L. D., Cai, T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, *16*, 101–117.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the depression anxiety stress scales (DASS) in clinical samples. *Behaviour Research and Therapy*, *35*, 79–89.

Carroll, B. J., Feinberg, M., Smouse, P. E., Rawson, S. G., & Greden, J. F. (1981). The Carroll rating scale for depression I: Development, reliability and validation. *British Journal of Psychiatry*, *138*, 194–200.

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*, 404–413.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, *55*, 584–594.

Collie, A., Shafiq-Antonacci, R., Maruff, P., Tyler, P., & Currie, J. (1999). Norms and the effects of demographic variables on a neuropsychological battery for use in healthy ageing Australian populations. *Australian and New Zealand Journal of Psychiatry*, *33*, 568–575.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40*, 1196–1208.

Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology*, *24*, 343–372.

Crawford, J. R., & Garthwaite, P. H. (2008). On the “optimal” size for normative samples in neuropsychology: Capturing the uncertainty associated with the use of normative data to q-uantify the standing of a neuropsychological test score. *Child Neuropsychology*, *14*, 99–117.

Crawford, J. R., & Garthwaite, P. H. (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks. *The Clinical Neuropsychologist*, *23*, 193–204.

Crawford, J. R., Garthwaite, P. H., Lawrie, C. J., Henry, J. D., MacDonald, M. A., Sutherland, J. & Sinha, P. (2009a). A convenient method of obtaining percentile norms and accompanying interval estimates for self-report mood scales (DASS, DASS-21, HADS, PANAS, and SAD). *British Journal of Clinical Psychology*, *48*, 163–180.

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009b). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist*, *23*, 1173–1195.

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales: Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology*, *42*, 111–131.

Crawford, J. R., Henry, J. D., Crombie, C., & Taylor, E. P. (2001). Normative data for the HADS from a large non-clinical sample. *British Journal of Clinical Psychology*, *40*, 429–434.

Faust, D. (1998). Forensic assessment. In A. S. Bellack & M. Hersen (Ed.), *Comprehensive clinical psychology volume 4: Assessment* (pp. 563–599). Amsterdam: Elsevier.

Feinberg, M., & Carroll, B. J. (1986). The Carroll rating scale for depression. In N. Sartouris & T. Ban (Eds.), *Assessment of depression* (pp. 188–200). Berlin: Springer Verlag.

Feldt, L. S. (1965). The approximate sampling distribution of kuder-richardson reliability coefficient twenty. *Psychometrika*, *30*, 357–370.

Gardner, M. J., & Altman, D. G. (1989). *Statistics with confidence-confidence intervals and statistical guidelines*. London: British Medical Journal.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, *23*, 56–62.

Henry, J. D., & Crawford, J. R. (2005). The 21 Item version of the Depression Anxiety Stress Scales (DASS-21): Normative data and psychometric evaluation in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*, 227–239.

Horwath, E., & Weissman, M. M. (1995). Epidemiology of depression and anxiety disorders. In M. T. Tsuang, M. Tohen, & G. E. Zahner (Eds.), *Textbook in psychiatric epidemiology* (pp. 317–343). New York: Wiley.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Belmont, CA: Duxbury Press.

Knight, R. G. (1984). Some general population norms for the short form Beck Depression Inventory. *Journal of Clinical Psychology*, *40*, 751–753.

Knight, R. G., Waal-Manning, H. T., & Spears, G. F. (1983). Some norms and reliability data for the State-Trait Inventory and the Zung Self-Rating Depression Scale. *British Journal of Clinical Psychology*, *22*, 245–249.

Ley, P. (1972). *Quantitative aspects of psychological assessment*. London: Duckworth.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales*. Sydney: Psychology Foundation Monograph.

Meltzer, H., Gill, B., Petticrew, M., & Hinds, K. (1995). *The prevalence of psychiatric morbidity among adults living in private households*. London: HMSO.

Osborne, R. H., Elsworth, G. R., & Hopper, J. L. (2003). Age-specific norms and determinants of anxiety and depression in 731 women with breast cancer recruited through a population-based cancer registry. *European Journal of Cancer*, *39*, 755–762.

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.

Shepherd, M., Cooper, B., Brown, A. C., & Kalton, C. W. (1966). *Psychiatric illness in general practice*. London: Oxford University Press.

Sherrets, F., Gard, G., & Langner, H. (1979). Frequency of clerical errors on WISC protocols. *Psychology in the Schools*, *16*, 495–496.

Spielberger, C. D., Gorsuch, R., & Lushene, P. R. (1970). *Manual for the state-trait anxiety inventory (form X)*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D. (1972). Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1, pp. 23–49). New York: Academic Press.

Spielberger, C. D., Goruch, R. L., Lushene, P. R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory (form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Spinhoven, P., Ormel, J., Sloekers, P. P. A., Kempen, G., Speckens, A. E. M., & VanHemert, A. M. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine*, *27*, 363–370.

Sullivan, K. (2000). Examiners' errors on the Wechsler Memory Scale-Revised. *Psychological Reports*, *87*, 234–240.

Tanaka, J., & Huba, G. J. (1984). Confirmatory hierarchical factor analysis of psychological distress measures. *Journal of Personality and Social Psychology*, 622–625.

Tully, P. J., Zajac, I. T., & Venning, A. J. (2009). The structure of anxiety and depression in a normative sample of younger and older Australian adolescents. *Journal of Abnormal Child Psychology*, *37*, 717–726.

Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, *12*, 63–70.

Zung, W. W. K. (1971). A rating instrument for anxiety disorders. *Psychosomatics*, *12*, 371–379.

Zung, W. W. K. (1986). Zung self-rating depression scale and depression status inventory. In N. Sartorius & T. A. Ban (Eds.), *Assessment of depression* (pp. 221–231). Berlin: Springer Verlag.