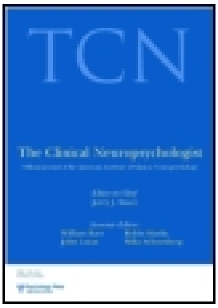

The Clinical Neuropsychologist
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ntcn20

Point and Interval Estimates of Percentile Ranks for Scores on the Texas Functional Living Scale

John R. Crawford [a], C. Munro Cullum [b], Paul H. Garthwaite [c], Emma Lycett [d] & Kate J. Allsopp [d]

[a] School of Psychology, University of Aberdeen, Aberdeen, UK

[b] Departments of Psychiatry and Neurology, The University of Texas Southwestern Medical Center at Dallas, Dallas, TX, USA

[c] Department of Mathematics and Statistics, The Open University, Milton Keynes, UK

[d] Pearson Assessment, London, UK

Version of record first published: 17 Sep 2012.

To cite this article: John R. Crawford, C. Munro Cullum, Paul H. Garthwaite, Emma Lycett & Kate J. Allsopp (2012): Point and Interval Estimates of Percentile Ranks for Scores on the Texas Functional Living Scale, The Clinical Neuropsychologist, DOI:10.1080/13854046.2012.720713

To link to this article: http://dx.doi.org/10.1080/13854046.2012.720713

The Clinical Neuropsychologist, 2012, 1–12, iFirst
http://www.psypress.com/tcn
ISSN: 1385-4046 print/1744-4144 online
http://dx.doi.org/10.1080/13854046.2012.720713

Point and Interval Estimates of Percentile Ranks for Scores on the Texas Functional Living Scale

John R. Crawford[1], C. Munro Cullum[2], Paul H. Garthwaite[3], Emma Lycett[4], and Kate J. Allsopp[4]
[1]School of Psychology, University of Aberdeen, Aberdeen, UK
[2]Departments of Psychiatry and Neurology, The University of Texas Southwestern Medical Center at Dallas, Dallas, TX, USA
[3]Department of Mathematics and Statistics, The Open University, Milton Keynes, UK
[4]Pearson Assessment, London, UK

Point and interval estimates of percentile ranks are useful tools in assisting with the interpretation of neurocognitive test results. We provide percentile ranks for raw subscale scores on the Texas Functional Living Scale (TFLS; Cullum, Weiner, & Saine, 2009) using the TFLS standardization sample data ($N = 800$). Percentile ranks with interval estimates are also provided for the overall TFLS T score. Conversion tables are provided along with the option of obtaining the point and interval estimates using a computer program written to accompany this paper (TFLS_PRs.exe). The percentile ranks for the subscales offer an alternative to using the cumulative percentage tables in the test manual and provide a useful and quick way for neuropsychologists to assimilate information on the case's profile of scores on the TFLS subscales. The provision of interval estimates for the percentile ranks is in keeping with the contemporary emphasis on the use of confidence intervals in psychological statistics

Keywords: IADLs; Percentile norms; Interval estimates; Bayesian methods; Computer scoring.

INTRODUCTION

The Texas Functional Living Scale (TFLS; Cullum, Weiner, & Saine, 2009) is a performance-based measure of functional competence and instrumental activities of daily living (IADLs). It consists of four subscales assessing basic functional living skills (Time, Money & Calculation, Communication, and Memory). The original version of the instrument was developed because of the limitations of other approaches to the measurement of functional ability in persons with dementia (Cullum et al., 2001). Most assessments of the ability to carry out IADLs are based on reports from family members. Unfortunately these reports may be biased by a variety of factors, including caregiver burden or insufficient patent–caregiver contact. Performance-based functional measures have been used to circumvent these limitations. However, many of the other available measures are lengthy or cumbersome to administer, assess a limited range of functions, and may not be suitable for patients with more pronounced deficits. For all of these reasons the

Address correspondence to: John R. Crawford, School of Psychology, College of Life Sciences and Medicine, King's College, University of Aberdeen, Aberdeen AB24 2UB, UK. E-mail: j.crawford@abdn.ac.uk
Accepted for publication: August 7, 2012. First published online: September 14, 2012.

current version of the TFLS was created to provide a brief and easily administered performance-based measure of IADLs for individuals with known or suspected dementia or other cognitive disorders across a wide age range (Cullum et al., 2009).

Quantifying performance on TFLS subscales

The TFLS provides an overall T score to quantify a patient's global level of functional capacity. At present the only way of quantifying performance on the four TFLS subscales is by referring to a table (Table A.1 of the test manual) of banded cumulative percentages. For example, using this table it can be ascertained that between 17% and 25% of the normative sample obtained a raw score of 7 on the TFLS Time subscale and a raw score of 6 on the Money & Calculation subscale. This table of banded cumulative percentages allows clinicians to gain a sense of how unusual or otherwise a given raw subscale score is. However, a useful alternative (and one that is perhaps more easily assimilated by users) is to provide the percentile rank for each raw score on each of the subscales. This is the principal aim of the present study.

In neuropsychological assessment, using percentile ranks as a means of expressing test scores has much to recommend it (Crawford & Garthwaite, 2009). Percentile ranks provide what is the most relevant piece of information for a clinician as they directly quantify the rarity or abnormality of a given score. In contrast, although standardized scores (such as IQ or T scores) are clearly useful, this is primarily because they allow clinicians to see how many standard deviation units a case's score is below (or above) the mean; with this information the clinician can then know (either from an internalized lookup table, or tables of the area under the normal curve) how unusual the score is. But of course this is the very information provided directly by the score's percentile rank.

In addition to providing tables for conversion of raw scores to percentile ranks (together with interval estimates; see later) a computer program was also developed to automate this process. This latter approach has the advantage that it is quicker and less error prone than referring to conversion tables. Research shows that clinicians frequently make many simple clerical errors when scoring or converting test scores (e.g., Faust, 1998; Sherrets, Gard, & Langner, 1979; Sullivan, 2000).

A second aim of the present study was to provide *interval* estimates of the percentile ranks corresponding to raw scores on the TFLS subscales. When neuropsychologists refer a case's score to percentile norms, their interest is in the standing (percentile rank) of the case's score in the normative *population*, rather than its standing in the particular group of participants who happen to make up the normative *sample* (in the case of the TFLS, the population in question is the general adult US population).

Despite the fact that, for the TFLS, the normative sample used to provide the basis of conversion from raw scores to percentile ranks was large, it is still the case that, as with any normative data, 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, Lawrie, et al., 2009). Interval estimates serve the useful general purpose of reminding us that all normative data are fallible, and serve the specific purpose of quantifying this fallibility (Crawford & Garthwaite, 2002; Gardner & Altman, 1989).

Finally, the primary score provided by the TFLS is the composite T score. Although composite scores have positive features (Crawford, 2012), it is an important neuropsychological tenet to examine the components (subscales in the present context) that contribute to a composite when deciding how to interpret that score and how much weight to place on it in clinical decision making. As with traditional IQ scores or Index scores, the TFLS T scores of two cases could be identical but this score could have arisen from very different combinations of subscale scores, reflecting different underlying patterns of cognitive dysfunction. The provision of percentile norms for the subscales allows for a *profile* of subscores to be derived, thereby assisting clinicians in assessing how much weight to give to the T score when reaching a formulation.

METHOD

Sample data

The data used in this paper are those obtained from the TFLS standardization sample (Cullum et al., 2009). The sample consisted of 800 persons (age range 16 to 90), recruited to be representative of the general adult US population in terms of key demographic variables (e.g., age, gender, educational background, geographic region etc).

Point estimates of percentile ranks

The standard method of obtaining percentile ranks was used (Crawford, Garthwaite, & Slick, 2009; Ley, 1972). That is,

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

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 one decimal place.

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 cognitive test scores are discrete (i.e., integer–valued), the underlying cognitive 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 6.5 to 7.4999 (plus an infinite number of additional 9 s after the fourth decimal place). Put 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 (formula 1).

Normative data for scales such as the TFLS 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 score. Indeed, if a normative sample is large and the data are skewed (as will be the case for scales that measure abilities that are largely within the competence of the majority of the general population), 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. (2009) have recently developed Bayesian and classical methods that incorporate the additional uncertainty arising from tied scores. Crawford, Garthwaite, Lawrie, et al. (2009) and Crawford, Cayley, Wilson, Lovibond, and Hartley (2011) have used these methods to provide interval estimates for self-report mood scales. The methods have also been used to provide interval estimates for a variety of neuropsychological test scores (Quintana et al. 2011; van der Werf & Vos, 2011). In the present study we apply the Bayesian method to the TFLS subscale data.

To illustrate the issue the methods address: suppose that in a normative sample of 100 people, 89 obtained lower scores than a case and 2 obtained the same score as the case. Then the point estimate of the percentile rank for the case's score (using formula 1) is 90 and applying Crawford et al. (2009) Bayesian method, the interval estimate is from 82.2 to 95.3. 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 79.8 to 97.1; the latter interval is wider because of the increased uncertainty introduced by the larger number of ties (10 versus 2).

The method uses a Jeffreys' prior and a mixture of beta distributions; the prior distribution is combined with the sample data to obtain the posterior distribution. Technical details of these methods are not set out here: see Crawford et al. (2009) for their derivation, and Garthwaite and Crawford (2011) for an additional mathematical treatment and evaluation.

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 neuropsychologist 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. The methods developed by Crawford et al. (2009) 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 which then

Table 1. Conversion of raw scores on the Time, Money & Calculation, and Memory subscales of the Texas Functional Living Scales to percentile ranks (with accompanying 95% credible intervals)

Raw score	Time	Money & Calculation	Memory	Raw score
0	0.0 (0.0 to 0.3)	0.0 (0.0 to 0.3)	0.3 (0.0 to 1.0)	**0**
1	0.1 (0.0 to 0.5)	0.1 (0.0 to 0.5)	1.3 (0.3 to 2.6)	**1**
2	0.2 (0.0 to 0.7)	0.2 (0.0 to 0.7)	3.9 (1.7 to 6.5)	**2**
3	0.5 (0.1 to 1.3)	0.8 (0.1 to 1.8)	7.8 (5.1 to 10.9)	**3**
4	0.9 (0.4 to 1.7)	2.3 (1.0 to 4.1)	17.2 (9.7 to 25.1)	**4**
5	2.8 (0.8 to 5.2)	5.1 (2.9 to 7.8)	62.3 (26.2 to 98.2)	**5**
6	7.3 (4.1 to 11.0)	13.9 (6.8 to 21.6)	–	**6**
7	16.4 (9.8 to 23.5)	30.1 (20.7 to 39.7)	–	**7**
8	38.2 (23.0 to 53.5)	69.6 (40.4 to 98.5)	–	**8**
9	76.8 (54.3 to 98.9)	–	–	**9**

provide100($1-\alpha$) 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.

Point and interval estimates of percentile ranks for TFLS T scores

The main aim of the present study is to provide percentile norms and accompanying interval estimates for TFLS *subscales*. However, it would be useful to also express TFLS T scores as percentile ranks and to quantify the uncertainty attached to them. Because there is a much wider range of raw scores available for conversion to T scores than there is for individual subscales, and because T scores are normalized, the uncertainty over the percentile rank of a case's T score can be quantified using parametric methods developed by Crawford and Garthwaite (2002). These methods use non-central *t* distributions. The technical details are not set out here: for a derivation of the methods see Appendix B of Crawford and Garthwaite (2002) and, for empirical confirmation (using Monte Carlo simulation) of their veracity, see Crawford, Garthwaite, and Porter (2010).

RESULTS

Obtaining point and interval estimates of the percentile ranks for raw scores

Raw scores on the Time, Money & Calculation, and Memory subscales can be converted to percentile ranks using Table 1. The point estimate of the percentile ranks are accompanied by their corresponding 95% credible intervals. So, for example, it can be seen that the percentile rank for a raw score of 5 on the Time subscale is 2.8 (95% CI = 0.8 to 5.2). Conversion of raw scores on the Communication subscale can be achieved using Table 2 (a separate conversion

Table 2. Conversion of raw scores on the Communication subscale of the Texas Functional Living Scales to percentile ranks, with accompanying 95% credible intervals

Raw score	Percentile Rank (PR)	95% Credible Interval on PR
0	0.0	0.0 to 0.3
1	0.0	0.0 to 0.3
2	0.0	0.0 to 0.3
3	0.0	0.0 to 0.3
4	0.0	0.0 to 0.3
5	0.0	0.0 to 0.3
6	0.0	0.0 to 0.3
7	0.1	0.0 to 0.5
8	0.2	0.0 to 0.7
9	0.3	0.1 to 0.8
10	0.3	0.1 to 0.9
11	0.4	0.1 to 1.1
12	0.6	0.2 to 1.3
13	0.6	0.2 to 1.4
14	0.6	0.2 to 1.4
15	0.8	0.3 to 1.7
16	1.1	0.5 to 2.0
17	1.4	0.7 to 2.6
18	1.9	1.1 to 3.0
19	2.4	1.4 to 3.8
20	3.6	2.2 to 5.3
21	5.9	3.7 to 8.5
22	8.7	6.5 to 11.2
23	12.4	9.0 to 16.1
24	18.1	13.9 to 22.6
25	26.9	20.5 to 33.5
26	39.8	31.7 to 48.0
27	57.3	46.6 to 67.8
28	83.7	67.6 to 99.2

table was used for the Communication subscale because it has many more items than the other scales; i.e., the range is from 0 to 28).

It can be seen from these two tables that, for low scores, the interval estimates of the percentile ranks are fairly narrow, thereby indicating that the point estimates of percentile ranks of the raw scores obtained using the TFLS normative sample provide an accurate estimate of the true percentile ranks of these raw scores in the normative population.

However, there is considerably more uncertainty attached to percentile ranks for high scores. This is an inevitable consequence of the fact that a large number of the normative sample (as expected given the nature of the test) obtained near perfect performance on the subscales (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 practical concern as there is little need

Table 3. Percentile ranks, with accompanying 95% confidence intervals, for T scores on the Texas Functional Living Scales

T score	Percentile Rank (PR)	95% Credible Interval on the PR
20	0.1	0.1 to 0.2
21	0.2	0.1 to 0.3
22	0.3	0.2 to 0.4
23	0.4	0.2 to 0.5
24	0.5	0.3 to 0.7
25	0.6	0.4 to 0.9
26	0.8	0.6 to 1.2
27	1.0	0.8 to 1.5
28	1.4	1.0 to 1.9
29	1.8	1.3 to 2.4
30	2.3	1.7 to 3.0
31	2.9	2.2 to 3.7
32	3.6	2.8 to 4.6
33	4.5	3.5 to 5.6
34	5.5	4.4 to 6.7
35	6.7	5.5 to 8.1
36	8.1	6.7 to 9.6
37	9.7	8.2 to 11.4
38	11.5	9.8 to 13.4
40	15.9	13.9 to 18.0
42	21.2	19.0 to 23.6
44	27.5	25.0 to 30.0
47	38. 2	35.5 to 40.9
51	54.0	51.2 to 56.7
55	69.1	66.5 to 71.7
60	84.1	82.0 to 86.1
66	94.5	93.3 to 95.6

for a precise estimate for high scores. That is, the TFLS is intentionally designed to measure functional living abilities that will be within the competence of most members of the general adult population. Clinical interest will be firmly focused on individuals who fail to achieve the expected level of performance.

Finally, Table 3 can be used to obtain point and interval estimates of the percentile ranks for TFLS T scores. As noted, a parametric method was used to compute the intervals. This, when combined with the large normative sample size, results in intervals that are narrow across the range of T scores. Thus it is the case that there is little uncertainty over the true percentile ranks of TFLS T scores. To illustrate the use of Table 3: it can be seen that the point estimate of the percentile rank for a TFLS T score of 37 is 9.7 and the accompanying 95% CI on this percentile rank is from 8.2 to 11.4.

Computer program for obtaining point and interval estimates of percentile ranks for raw scores on the TFLS subscales

As previously noted, a computer program for PCs, TFLS_PRs.exe, was written (using the Delphi programming language) to express a patient's raw scores on the TFLS subscales as percentile ranks. The program can be downloaded (either as an uncompressed executable or as a zip file) from the first author's web pages at www.abdn.ac.uk/~psy086/dept/TFLS_PRs.htm. After downloading the program it can be run by clicking on the program in Windows Explorer, or, if a shortcut to the program has been created on the user's desktop, by clicking on the shortcut icon).

The user need only enter the raw scores on the four TFLS subscales and click on the "Compute" button. There is the option of entering identifying information for the patient (in the form of User's Notes) for future reference. If a case has not been administered all subscales then the relevant data fields(s) can simply be left blank. (In these circumstances it is obviously not possible to convert the sum of the patient's raw scores to a TFLS T score so the T score and its percentile rank are not reported.)

The output from the program consists of the case's raw scores on the TFLS subscales (as entered by the user), and the point and interval estimates of the percentile ranks for these scores. The sum of the case's raw scores is converted to a TFLS T score and the point and interval estimate of the percentile rank for the T score is also reported. The results are arranged in a table so that users can quickly assimilate the case's profile of scores. The output can be viewed on screen, saved to a file, and/or printed.

DISCUSSION

An example

Suppose a case obtained raw subscale scores of 6 on the Time subscale, 5 on the Money & Calculation subscale, 23 on Communication, and 1 on Memory. The sum of these raw scores is 35 and hence (using either the TFLS manual or computer program) the case's TFLS T score is 32. Table 4 presents the point and interval estimates of the percentile ranks for the case's scores: the layout adopted is the same as the output provided by the computer program that accompanies this paper. It can be seen from Table 4 that the case's performance is particularly poor on Memory (PR = 1.3; thus only 1.3% of the normative population are expected to obtain a lower score).

Although still poor (i.e., a matter for concern), performance on Communication is a *relative* strength (PR = 12.4). The percentile rank for the case's overall score (T score) is 3.6, thus making it clear that there is an overall

Table 4. Percentile ranks (PR), with accompanying 95% credible intervals (CI), for a case's scores on the Texas Functional Living Scales subscales

Subscale	Raw Score	Percentile Rank (PR)	(95% CI on PR)
Time	6	7.3	(4.1 to 11.0)
Money & Calculation	5	5.1	(2.9 to 7.8)
Communication	23	12.4	(9.0 to 16.1)
Memory	1	1.3	(0.3 to 2.6)
T Score	32	3.6	(2.8 to 4.6)

Also shown are the point and interval estimates for the case's T score.

impairment of functional living abilities. It can be seen that there is relatively little uncertainty over the standing of the case's scores, particularly for the overall T score. For example, for the Memory score we can be (95%) confident that the percentile rank of the patient's score in the normative population (point estimate = 1.3) is not lower than 0.3 and not higher than 2.6%. For the T score (point estimate of percentile rank = 3.6) we can be 95% confident that the true percentile rank is not lower than 2.8 and not higher than 4.6)

Bayesian versus classical interpretations of the interval estimate on a score's percentile rank

The method used in the present paper to provide interval estimates of a patient's percentile ranks is a Bayesian method. As Antelman (1997) notes, the frequentist (classical) conception of a 95% confidence interval 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 Time subscale is as follows: "if we could compute a confidence interval for each of a large number of normative samples collected in the same way as the present TFLS normative sample, about 95% of these intervals would contain the true percentile rank of the patient's score".

The Bayesian interpretation of such an interval is "there is a 95% probability that the true percentile rank of the patient'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 classical/frequentist confidence limits probably construe these in what are essentially Bayesian terms (Howell, 2002).

Finally, for the present problem the Bayesian approach used here exhibits a very high degree of convergence with its classical equivalents (Crawford et al., 2009). For example, applying Crawford et al. (2009) classical Clopper-Pearson mid-*p* method to the examples featured earlier yielded intervals that were identical to the Bayesian interval to two decimal places.

The method for setting confidence intervals on the TFLS T scores is a classical method. However, a Bayesian approach to this problem (i.e., a Bayesian method that also assumes normality) gives the same results as the classical method (Crawford & Garthwaite, 2007). This convergence between different methods is reassuring regardless of whether a neuropsychologist is Bayesian, frequentist or eclectic in orientation. Importantly, it also means that it is legitimate to apply a Bayesian interpretation to these intervals as well as to the intervals for the TFLS subscales.

Confidence intervals capturing sampling error versus measurement error

The confidence intervals on the percentile ranks provided here 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 neuropsychologist is posing the question "assuming scores are normally distributed, and assuming no error in estimating the population mean, standard deviation 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, 2012; Crawford & Garthwaite, 2008). In contrast, when using the intervals presented in the present paper, 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 patient's TFLS subscale score (or T score) 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., 2009).

Future developments

The normative data used in the present paper came from the TFLS standardization sample, i.e., the participants were drawn from the (healthy) general adult population. However, it would be useful for clinicians if point and interval estimates for the percentile ranks of TFLS scores were also available for clinical populations. These would provide further context when evaluating an individual case's score. Percentile norms could be gathered for different clinical populations and different settings. For example, in the case of a community-dwelling patient with a diagnosis of probable AD, it would be useful to be able to obtain an estimate of how typical or unusual an individual's TFLS scores are within the population of other community-dwelling patients with the same diagnosis.

The methods applied here allow clinicians to examine a case's TFLS performance at the level of functional subdomains (i.e., at the subscale level). This also means that clinicians can examine a case's profile of TFLS scores (i.e., evaluating the case's relative strengths and weaknesses). Analysis of TFLS-derived profiles may aid in clinical interpretation and monitoring changes over time across various populations, although research is needed to explore the utility of profile analysis of the TFLS in various populations. However, although inferential methods are applied to the individual subscales (i.e., the uncertainty over the case's standing on each subscale is quantified via credible intervals), the present paper does not apply inferential methods to the differences among a case's subscale scores.

When tests yield data that can be treated as continuous and normally distributed there are a number of methods that can be applied to examine score differences (e.g., see Crawford & Garthwaite, 2007; McIntosh & Brooks, 2011). Unfortunately, analogous methods that could be applied to data that do not have these properties (i.e., where the data are of a form such that, as here, scores need to be expressed as percentile ranks and will contain "ties") have yet to be developed. Although there are significant technical problems to be overcome, we are hopeful that suitable methods can be developed in future.

Conclusion

The tabled values provided here, and particularly the accompanying computer program, provide a quick and reliable means of obtaining percentile norms for TFLS subscales. The percentile norms allow neuropsychologists to quantify the rarity or otherwise of the patient's subscale scores and T score. 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.

REFERENCES

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

Crawford, J., Cayley, C., Wilson, P. H., Lovibond, P. F., & Hartley, C. (2011). Percentile norms and accompanying interval estimates for self report mood scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). *Australian Psychologist*, *46*, 3–14.

Crawford, J. R. (2012). Quantitative aspects of neuropsychological assessment. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians* (2nd ed. pp. 129–155). Chichester, UK: Wiley.

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 quantify 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. (2009). 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., & Porter, S. (2010). Point and interval estimates of effect sizes in the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*, 245–260.

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). 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.

Cullum, C. M., Saine, K., Chan, L. D., Martin-Cook, K., Gray, K., & Weiner, M. F. (2001). A performance-based instrument to assess functional capacity in dementia: The Texas functional living scale. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *14*, 103–108.

Cullum, C. M., Weiner, M. F., & Saine, K. C. (2009). *Texas Functional Living Scale examiner's manual*. San Antonio, TX: Pearson.

Faust, D. (1998). Forensic assessment. *In Comprehensive clinical psychology volume 4: Assessment* (pp. 563–599). Amsterdam: Elsevier.

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

Garthwaite, P. H., & Crawford, J. R. (2011). Confidence intervals for a binomial proportion in the presence of ties. *Journal of Applied Statistics*, *38*, 1915–1934.

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

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

McIntosh, R. D., & Brooks, J. L. (2011). Current tests and trends in single-case neuropsychology. *Cortex*, *47*, 1151–1159.

Quintana, M., Pena-Casanova, J., Sanchez-Benavides, G., Langohr, K., Manero, M. R., Aguilar, M., … Blesa, R. (2011). Spanish multicenter normative studies (Neuronorma Project): Norms for the Abbreviated Barcelona Test. *Archives of Clinical Neuropsychology*, *26*, 144–157.

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

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

van der Werf, S. P., & Vos, S. H. (2011). Memory worries and self-reported daily forgetfulness: A psychometric evaluation of the Dutch translation of the Multifactorial Memory Questionnaire. *The Clinical Neuropsychologist*, *25*, 244–268.