Primary Care Mental Health 2006;4:7–20

International research

When clients do not provide discharge data in primary care services: using the CORE system to address attrition in CORE-OM outcome data

Tracy Mullin BA MRes
Research Assistant

Michael Barkham BEd MA MSc PhD
Professor of Clinical and Counselling Psychology

Psychological Therapies Research Centre, University of Leeds, UK

ABSTRACT

Rationale Missing data in routine practice settings in primary care present a major issue in determining the effectiveness of clients' outcomes in primary care. This is particularly true for client self-report measures, one of which is the CORE-OM which is widely used in primary care settings. In order to build a robust evidence base in practice settings, user-friendly procedures need to be disseminated to practitioners.

Aim To provide a statistically sound yet simple method for calculating a predicted CORE-OM end-of-therapy score for those clients who did not complete an outcome measure at discharge.

Results The models created were statistically able to predict outcome scores, using clients' pre-therapy CORE-OM score and therapist-identified problems. When testing the models on two independent datasets, there was some variability with predictive ability. Where the significance level for the difference between predicted and actual score was problematic, this was associated with small effect sizes.

Conclusion It is possible to construct equations for men and women which, with an inevitable degree of error, can be used to predict reasonably well CORE-OM scores for those patients in primary care who did not complete an outcome measure at discharge.

Keywords: attrition, outcomes, CORE-OM, CORE system

Introduction

Assessing outcomes of psychological interventions is a cornerstone of evidence-based practice and NHS healthcare policy.[1] The traditional model of evidence-based practice builds upon randomised controlled trials (RCTs). This technology is invariably resourced via additional external funding which supports and ensures delivery of both time 1 (T_1) pre-intervention assessment data, and time 2 (T_2) post-intervention outcome data. This simple model of outcome evidence has been extended within research settings, such that repeated administration of measures can be collected and greater specificity of the temporal shape of change can be obtained.[2,3] Pursuit of repeated measurement data has occurred because of the simplistic nature of deducing change (i.e. improvement or deterioration) from just two data points. That is, only a hypothetical straight line can be drawn between T_1 and T_2 data points, and few practitioners would consider change to be linear under all conditions.

However, although the T_1–T_2 model is deemed basic within traditional research settings, the

collection of T_2 data can prove extremely demanding for services in routine practice settings. This particularly applies to the collection of T_2 client self-report data. Using routine data – for comparative purposes or otherwise – carries a range of intrinsic pitfalls.[4] Indeed, attrition rates bedevil evaluation within routine practice, since this phenomenon arises from both client dropout rates – the mean psychotherapy dropout rate is about 47% – and the non-completion of T_2 measures for clients who remain in therapy.[5] This issue runs the danger of undermining the collection of data from the very people who are the focus of a 'patient-led' NHS, and replacing it with practitioner-completed measures. There is a strong argument that the development of a robust evidence base relating to client outcomes requires data from both client and practitioner rather than from any one party alone. Set against this context, our aim was to devise a pragmatic but robust procedure for deriving proxy T_2 client outcome scores that would be of practical use to routine services. To do so, we utilised a large anonymised and aggregated dataset using data drawn from a range of healthcare settings.

Missing data

Missing outcome data tend not to occur in a random pattern but rather are derived directly from those clients who drop out. This arises for a range of reasons and reflects that this group of service users can differ in important ways from those service users who yield complete data. Therefore, loss of data is important not merely because it results in a loss of sample size and statistical power, but because it can bias the remainder of the dataset.

Types of missing data

The most appropriate way to handle missing or incomplete data will depend upon how data points became missing. Three unique types of missing data mechanisms have been identified: missing completely at random (MCAR), missing at random (MAR) and non-ignorable (NI).[6] In instances where MCAR applies, cases with complete data are indistinguishable from cases with incomplete data. An example of MCAR missing data would be data sheets lost due to someone arbitrarily discarding some of the data sheets. By contrast, datasets with incomplete data where MAR is appropriate will differ from cases with complete data, but the pattern of the missing data is predictable from other variables in the database. For example, if a certain age group was less likely to fill out an outcome measure, but the score on the outcome measure was unrelated to age, then the data would be MAR. Missing data which fits the NI category is non-random and it is not predictable from other variables in the database. This is because the missingness pattern is explainable only by the very variable(s) on which the data are missing. For example, if a certain age group is less likely to provide their age when asked, this is NI data. In practice it is usually difficult to meet the MCAR assumption, whereas MAR is an assumption that is more often, but not always, tenable. The more relevant and related predictors one can include in statistical models, the more likely it is that the MAR assumption will be met.

Within the psychotherapy literature, a wide range of variables have been identified that are related to dropout, for example racial status and socio-economic status.[5] Hence it is likely that the pattern of missing data would be predictable from these variables. Therefore, the best assumption is that the data to be analysed is MAR, and thus the missingness is ignorable. Accordingly, we would not need to include in the model the variables on which the missingness is based. This is in line with our aim which is to predict the missing value rather than to predict whether data are missing or likely to be missing. It is unlikely that such demographic variables are likely to be able to predict post-therapy CORE-OM score.[7] Thus, the variables investigated were those that it was hypothesised would be related to post-therapy outcome score rather than missingness.

The method used in this paper, suitable for dealing with MAR data, is the traditional regression-based imputation. For this, a regression equation based on the data cases present was computed. Thus for cases missing, the independent variables identified to be part of the regression equation are input, and the predicted value is used as a substitute for the missing value. There are a number of strengths of using this method for dealing with missing data which include the following: (1) it saves the deletion of data; (2) the estimated data preserve deviations from the mean and the shape of the distribution; and (3) it is a more accurate method than 'simple techniques' such as mean substitution or casewise deletion.[8,9] Thus it has been concluded that the regression method may be best when data are missing in non-random patterns and >20% data are missing.[10] However, this method has limitations in that it requires all predicted values to fall within the expected range, and carries the proviso that the method may underestimate the variance statistic.

Although newer methods of missing data estimation (e.g. maximum likelihood estimation or bootstrapping) may in some circumstances be preferable due to their being superior in terms of statistics, one of the primary aims of the current paper was to provide a model from which primary care practitioners could predict T_2 data.[11,12] A key consideration in this

regard was to utilise procedures to produce a model that could be used by practitioners and was not computer intensive. With regard to this it was concluded that regression analysis would be most suitable.

Outcome measure

Our focus was on one outcome measure, namely the CORE-OM.[13–15] The CORE-OM has been widely adopted in NHS psychological therapy services as a routine assessment and outcome measure in large-scale studies of levels of distress and outcomes of treatment in primary and secondary care settings.[13,14,16,17] It has also been used among university students and older adults, and to monitor significant risk (including self-harm, suicide and risk to others) for patients with borderline personality disorder.[18–20] The CORE-OM is part of a larger CORE system which comprises information on the identified problems from the perspective of the practitioner, and is completed at intake and discharge.[21,22]

Although the CORE-OM is widely used in practice, there are serious issues relating to the T_2 attrition level for data. Using large accumulated CORE-OM datasets across multiple primary care sites, overall T_2 completion rates have been reported as 38%, while those for clients with planned endings have been reported as 73% and 88%.[16,23,24] However, for clients with unplanned endings, T_2 completion rates of 8% have been indicated.[23] A study in primary care settings found over three-quarters of clients were, in principle, willing to complete outcome measures, but in fact only one in four of these actually did complete the CORE-OM at T_1 and T_2.[17] In sum, T_2 completion rates vary enormously as a function of a range of factors – worthy of investigation in itself – but the most notable being whether they are drawn from client samples with planned or unplanned endings.

Aims and assumptions

Against this background, our aim was to realise a way of predicting T_2 CORE-OM scores when the actual T_2 data were not available. Two hypotheses underpinned our model. First, that the identified problems indicated by the therapist in their post-therapy assessment form would be able to predict, to a degree, T_2 CORE-OM scores because it is assumed that the therapist knows enough about the client to be able to summarise the client's current state. If an identified problem is highlighted, regardless of whether it has been addressed in therapy, it is assumed that this will be reflected in a client's CORE-OM score. Pre-therapy-identified problems were not included in the model because it was found that these were less accurate predictors of T_2 CORE-OM scores than post-therapy-identified problems, which is supported by literature which suggests that there is greater concordance between client self-evaluation and therapist assessment at discharge than at baseline.[25] Furthermore, excluding the pre-therapy-identified problems would help to ensure the model was as simple as possible to use, and having fewer variables in the model reduces the risk of modelling noise instead of outcome. It was also hypothesised that the T_1 CORE-OM score would be a predictor because there is likely to be a relationship between T_1 and T_2 CORE-OM scores, and research has found that the best individual predictor of outcome is intake score.[26] Client demographics were not included in the model because they are not adequate predictors of outcome.[7] Thus, we used T_1 CORE-OM scores and the T_2-identified problems to predict the T_2 CORE-OM score.

It was concluded that to create one all-encompassing equation would not provide the model with enough accuracy to reliably predict T_2 scores. Thus it was decided that because the majority of services that use CORE-OM, and from which the datasets derive, were primary care services, then the equation should be tailored to these specifically rather than to allow possible contamination and a loss of accuracy by using data from different types of services. Furthermore, in previous work on the CORE system it has been noted that there are differences between scores for males and females. This has led to there being different clinical cut-off scores on the CORE-OM for males and females, with the male score being higher than that for females. Thus it was hypothesised that to take into account these score differences, a different equation should be derived for each sex. Therefore three equations will be presented. Firstly, a brief overview of a general equation for primary care clients will be provided, that is intended to serve as an introduction. Secondly, two specific equations, one for each sex, will be presented that will provide enough detail to allow a thorough evaluation of them and enable their use in services to predict CORE-OM T_2 scores.

Methods

Datasets

In total three anonymised datasets were used. Aside from the key master dataset on which the analysis will be conducted, there are two further data sets on

which the output of the analysis will be tested, which are totally independent of the first set and of each other. They are derived from two different primary care population samples for clients who entered therapy, and can therefore be used to reliably validate the outcome of the original analysis. Dataset 1, the master dataset, comprises 33 services which contribute between 5 and 234 cases each.[16] Dataset 2 is an updated version from one primary care service.[16] Dataset 3 is drawn from a larger dataset, from one primary care trust, with services each contributing between 57 and 140 cases.[17] The descriptives for the datasets are presented in Table 1 and Figure 1.

Measures

CORE-OM

The CORE-OM is a 34-item self-report measure primarily completed at pre- and post-therapy.[13–15] Items cover domains of subjective wellbeing (4 items), symptoms/problems (12 items), and life/social functioning (12 items). In addition, it contains four items on risk to self, and two items on risk to others. The symptoms/problems and functioning dimensions of the CORE-OM can be divided further into specific problem areas. These item 'clusters' for the symptoms/problems dimension are: depression (4 items); anxiety (4 items); physical problems (2 items); trauma (2 items); and for the functioning dimension: close relationships (4 items); social relationships (4 items); general functioning (4 items). The items are rated on a scale of '0' (not at all) to '4' (most or all the time) with a time scale of 'over the last week'. Internal reliability, sensitivity to change, test–retest stability, convergent validity in relation to other measures, and discrimination between clinical and non-clinical populations have been established and reported, as well as comparisons between presentation at primary and secondary levels of NHS care.[13–15]

CORE-A (CORE-assessment)

The CORE-A comprises two practitioner-completed forms: the therapist assessment form (TAF) and the end of therapy form (EOT).[14,21,23,27] These forms were developed as an audit and evaluation tool that could be used alone or as a complement to the CORE-OM. Data from the CORE-A EOT used in this paper can be categorised into two groups as follows:

1 *sex* – male/female; *age* – in years
2 *identified problems/concerns*:
   – depression: including mood, emotional problems, hopelessness, dysthymia, cyclothymia
   – anxiety/stress: including generalised, stress, adjustment, irritability, phobia, panic, obsessive–compulsive disorder, dissociation

Table 1 Sample descriptive statistics

		Dataset 1	Dataset 2	Dataset 3
Overall	*n*	1706	444	508
	Age range (years)	12–86	12–80	13–87
	Mean age (SD)	39.7 (12.5)	37.3 (12.6)	35.1 (12.3)
Male	*n* (%)	460 (27)	113 (25)	145 (29)
	Age range (years)	11–89	14–80	15–83
	Mean age (SD)	37.5 (12.9)	38.1 (12.9)	35.0 (11.9)
Female	*n* (%)	1246 (73)	326 (73)	363 (71)
	Age range (years)	12–88	12–80	13–87
	Mean age (SD)	36.8 (13.3)	37.0 (12.4)	35.1 (12.5)
Frequency of therapy[a]	1	Person-centred (*n* = 710)	Person-centred (*n* = 331)	Cognitive–behavioural (*n* = 280)
	2	Integrative (*n* = 647)	Integrative (*n* = 190)	Supportive (*n* = 245)
	3	Psychodynamic (*n* = 356)	Cognitive–behavioural (*n* = 161)	Person-centred (*n* = 158)
	4	Cognitive–behavioural (*n* = 353)	Psychodynamic (*n* = 99)	Structured/brief (*n* = 127)

[a] Not mutually exclusive, may be received in combination

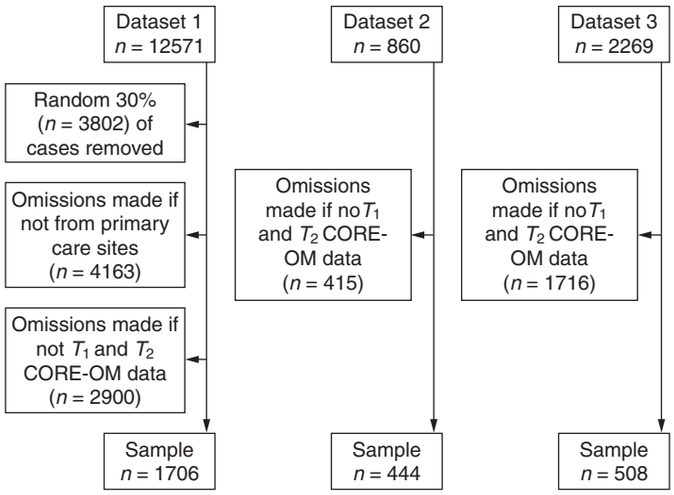

Figure 1 Flow diagram of data cleaning

- psychosis: including schizophrenia, schizotypal, psychopathic
- personality problems: including personality disorder, sociopathic, psychopathic
- cognitive learning: including dementia, memory loss, brain injury, learning difficulties, intellectual impairment, dyslexia
- eating disorder: including anorexia, bulimia
- physical problems: including disability, sleep problems, sexual problems, psychosomatic
- addictions: including alcohol, drug, gambling
- trauma/abuse: childhood and adulthood abuse, post-traumatic stress disorder, trauma
- bereavement/loss: death of significant other in last 6 months
- self-esteem: including loss of confidence in any area
- interpersonal/relationship: including specific relationship problems, e.g. with partner, and non-specific, e.g. shyness, inability to form relationships
- living/welfare: including financial problems, employment difficulties, housing problems, living conditions, lack of self-care
- work/academic: including motivation, concentration, performance
- other.

These identified problems are rated on a severity scale of 1 to 4 (i.e. causing minimal/mild/moderate/severe difficulty), with problems not occurring left blank.

Analyses

The initial analysis used the master dataset (dataset 1). Stepwise multiple regression analyses (suitable for exploratory purposes) were used to determine the combination of variables that best predicted CORE-OM post-therapy score, according to amount of variance explained (r^2) and with each variable significantly contributing to the model. In order to test both the accuracy and universality of the model, the equation was then tested on datasets 2 and 3. For all possible cases, the relevant data from the datasets were inputted into the regression equations to derive predicted CORE-OM post-therapy scores. These predicted scores were then compared to the actual scores reported in the dataset. We used both parametric and non-parametric tests to allow the results to be contrasted. It was found that the parametric *t* test gave more conservative results, and thus it was decided that these test results only would be reported. Due to the large sample sizes, the tests have large degrees of power and therefore are likely to detect even very small differences, and thus the likelihood of obtaining a statistically significant result is high. Therefore, the important statistics to note are the effect sizes (*r*). Subsequent analyses repeated these procedures using only male and female clients.

The mean of the CORE-OM items is often used in research studies but, consistent with more recent reporting procedures, we present the CORE-OM clinical score here, which is the mean of the completed items multiplied by 10, giving a more convenient 0–40 range.[13] Using a 0–40 range enhances the likelihood of practitioners assigning clinical meaning to scores, because it is easier to work with, and meaning can be assigned to whole numbers.[28] Importantly, however, this format does not alter the psychometric properties of the CORE-OM, and the published norms can easily be multiplied by 10 for direct comparison.[14,15] However, although we present

CORE-OM scores in their clinical form, the regression analyses and equations are based on the mean CORE-OM scores. This is intended to allow the equations to be utilized to produce both CORE-OM clinical and mean scores. To derive a mean CORE-OM score, use the equations as they are. To derive a CORE-OM clinical score, simply multiply the outcome of the equation by $\times 10$ (*see* Box 1).

Results

Model 1: a model for primary care sites

The identified problems that were found to be the strongest predictors of the outcome variable (see Table 2), and to explain the most variance when entered into a model together, were depression, anxiety, and interpersonal/relationship problems. The R^2 values indicate how much variability in the outcome is accounted for by each predictor. β (Beta) is the change in the outcome variable (OM score) associated with a unit change in the predictor variable. Because these β values are standardized, they are directly comparable even though the variables have been measured in differing units. The descriptive statistics for the variables included in the model are presented in Table 3 (n = 1706). Table 3 shows that CORE-OM scores decrease from pre-therapy to post-therapy. The small standard errors indicate that the mean values for the sample are very similar to those that would be found in a larger population. There are significant, but not large, individual correlations between each predictor variable and the outcome variable. These relationships suggest that the variables should be good predictors.

There is a medium-sized relationship between the predictor variables and the outcome variable (r = 0.59). The predictors account for 35% of the variation in post-therapy CORE-OM scores. The model was significantly better at predicting the outcome variable than using the sample T_2 CORE-OM mean, F (4,1701) = 225.17, $P < 0.001$, effect size (ES) r = 0.34. The standard error of the estimate (SEE), which is a measure of the size of a 'typical' error, is 4.83. Fewer than 5% of cases have a standardised residual outside ± 2, which is accurate for a normally distributed sample of the differences between the predicted and observed values. The model parameters are presented in Table 4. Thus the regression equation for the model, stated to three significant figures, is:

$$OM = 0.0479 + (0.287pre_i) + (0.162dep_i) + (0.0991anx_i) + (0.0491int_i) + e_i$$

Pre = client pre-therapy CORE-OM mean score; Dep = therapist post-therapy depression rating; Anx = therapist post-therapy anxiety rating; Int = therapist post-therapy interpersonal problem rating; e = error term (an unknown value). The subscript $_i$ represents a client and their score on that variable.

Box 1 Formal and practice formulae

To simplify the equation so it can be easily used it is recommended that the equation be rounded to one significant figure.

Equation for men

The *formal equation* for men presented in this paper is:

$$OM = 0.0590 + (0.309pre_i) + (0.134dep_i) + (0.111anx_i) + (0.0451int_i) + e_i$$

Thus, the *practice equation* for men is:

Male OM = 0.06 + (0.3pre) + (0.1dep) + (0.1anx) + (0.05int)

Equation for women

Similarly, the other two equations are also used to one significant figure only. Thus, the *practice equation* for women is:

Female OM = 0.04 + (0.3pre) + (0.2dep) + (0.09anx) + (0.05int)

Equation for men and women combined

And the general *practice equation* (though it is recommended that this one is only used where the gender of clients is unknown) is:

General OM = 0.05 + (0.3pre) + (0.2dep) + (0.1anx) + (0.05int)

Table 2 Strength of predictor variables

	General		Male		Female	
Variable	R^2	Standardised β[a]	R^2	Standardised β[a]	R^2	Standardised β[a]
Pre-therapy CORE score	0.172	0.414	0.200	0.447	0.170	0.412
Depression	0.290	0.539	0.302	0.549	0.299	0.547
Anxiety	0.251	0.501	0.244	0.494	0.265	0.515
Self-esteem	0.230	0.480	0.180	0.425	0.254	0.504
Interpersonal/ relationship	0.176	0.420	0.164	0.405	0.212	0.460
Work/ academic	0.157	0.396	0.218	0.467	0.113	0.336
Trauma	0.144	0.380	0.149	0.387	0.174	0.417
Living/welfare	0.140	0.374	0.225	0.475	0.107	0.327
Bereavement	0.134	0.366	0.212	0.460	0.107	0.327
Personality problems	0.120	0.347	0.330	0.575	0.072	0.268
Cognitive/ learning	0.080	0.283	0.296	0.544	0.025	0.158
Addictions	0.078	0.279	0.197	0.444	0.017	0.129
Eating disorder	0.077	0.278	0.017	0.131	0.108	0.329
Physical	0.069	0.263	0.151	0.389	0.092	0.304
Other	0.054	0.232	0.233	0.482	0.014	0.117
Psychosis	0.031	0.176	0.141	0.375	0.005	0.074

[a]All β values are significant ($P < 0.001$)

Testing the model

For dataset 2 (n = 444) there was a medium-sized significant correlation (r = 0.59, $P < 0.001$) between the predicted and actual results. The difference in mean (M) scores was small (actual M = 7.62 ± 0.52, standard deviation (SD) = 5.61; predicted M = 8.33 ± 0.32, SD = 3.48; ES_{diff} = 0.15) although there was a statistically significant difference between them: $t(443) = -3.31$, $P < 0.001$, ES r = 0.16. For dataset 3 (n = 508) there was a medium-sized significant correlation (r = 0.63, $P < 0.001$) between the predicted and the actual scores. The mean scores are very similar for the two sets of scores: actual scores, M = 8.11 ± 0.52, SD = 5.94; predicted scores, M = 7.97 ± 0.28, SD = 3.26; ES_{diff} = 0.02. Furthermore, there was no statistically significant difference between them: $t(507) = 0.69$, $P = 0.49$, ES r = 0.031.

Model 2: a model for males

The identified problems that were the strongest predictors of post-therapy CORE-OM score were different from those identified in the previous analysis (see Table 2), however when input into just one model the variables that together explained the most variance were the same as those identified in the previous analysis. The descriptive statistics are

Table 3 Model descriptives

Model		Mean	SD	SE	*r* with DV
1: site	Post-CORE-OM score (DV)	8.22	5.96	0.14	–
	Pre-CORE-OM score	17.76	6.42	0.16	0.41[a]
	Depression	0.77	0.97	0.02	0.48[a]
	Anxiety	1.00	0.99	0.02	0.40[a]
	Interpersonal	0.81	1.06	0.03	0.31[a]
2: male	Post-CORE-OM score (DV)	8.54	6.20	0.29	–
	Pre-CORE-OM score	17.33	6.83	0.32	0.44[a]
	Depression	0.81	0.98	0.05	0.45[a]
	Anxiety	1.03	1.05	0.05	0.38[a]
	Interpersonal	0.81	1.10	0.05	0.32[a]
3: female	Post-CORE-OM score (DV)	8.10	5.87	0.17	–
	Pre-CORE-OM score	17.93	6.26	0.18	0.40[a]
	Depression	0.76	0.97	0.03	0.49[a]
	Anxiety	0.99	0.96	0.03	0.41[a]
	Interpersonal	0.80	1.05	0.03	0.31[a]

DV, dependent variable.
[a] $P < 0.001$

Table 4 Model parameters

Model		Unstandardised		Standardised β	*t*	95% CI for β	
		β	SE			Lower	Upper
1: site	Constant	0.05	0.04		1.34	–0.02	0.12
	Pre-CORE-OM score	0.29	0.02	0.31	15.21[a]	0.25	0.32
	Depression	0.16	0.02	0.26	10.15[a]	0.13	0.19
	Anxiety	0.10	0.01	0.16	6.57[a]	0.07	0.13
	Interpersonal	0.05	0.01	0.09	3.88[a]	0.02	0.07
2: male	Constant	0.06	0.07		0.88	–0.07	0.19
	Pre-CORE-OM score	0.31	0.04	0.34	8.52[a]	0.24	0.38
	Depression	0.13	0.03	0.21	4.32[a]	0.07	0.19
	Anxiety	0.11	0.03	0.19	3.97[a]	0.06	0.17
	Interpersonal	0.05	0.03	0.08	1.80[b]	0.00	0.09
3: female	Constant	0.04	0.04		1.00	–0.04	0.13
	Pre-CORE-OM score	0.28	0.02	0.30	12.68[a]	0.24	0.32
	Depression	0.17	0.02	0.28	9.22[a]	0.14	0.21
	Anxiety	0.09	0.02	0.15	5.17[a]	0.06	0.13
	Interpersonal	0.05	0.02	0.09	3.47[a]	0.02	0.08

CI, confidence interval
[a] $P < 0.001$, [b] $P < 0.05$

presented in Table 3 ($n = 460$) and show that CORE-OM scores decrease from T_1. There were significant, but not large, individual correlations between each predictor variable and the outcome variable. There was a medium-sized relationship between the predictor variables and the outcome variable ($r = 0.58$). The predictors accounted for 34% of the variation in T_2 CORE-OM scores. The model was significantly better at predicting the outcome variable than using the sample T_2 CORE-OM mean, F (4,455) = 58.47, $P < 0.001$, ES $r = 0.33$. The SEE value is 5.06. Fewer than 5% of cases have a standardised residual outside ±2.

The model parameters are presented in Table 4. Thus the regression equation for the model, to three significant figures, is:

$$OM = 0.0590 + (0.309pre_i) + (0.134dep_i) + (0.111anx_i) + (0.0451int_i) + e_i$$

Testing the model: dataset 2, n *= 113*

The left and middle plots in Figure 2 present the spread of scores with the specific percentile values in the graph for the actual results being 4.10, 7.60 and 12.20, and for the predicted results being 6.60, 8.60, and 10.70. The difference in mean scores was small: actual $M = 8.52 \pm 1.13$, SD = 6.14; predicted $M = 8.80 \pm 0.65$, SD = 3.52; $ES_{diff} = 0.06$. There was a statistically significant correlation between the predicted and actual scores, $r = 0.61$, $P < 0.001$. Furthermore, there was no significant difference between the two sets of scores, $t(112) = –0.61$, $P = 0.55$, ES $r < 0.001$.

Testing the model: dataset 3, n *= 145*

The left and middle plots in Figure 3 show the spread of scores, with the exact percentile values for the actual scores at 3.50, 6.20 and 11.05, and for the predicted scores at 6.09, 8.09 and 10.12. There was a small difference between the scores: actual $M = 8.14 \pm 1$, SD = 6.17; predicted $M = 8.49 \pm 0.55$, SD = 3.38; $ES_{diff} = 0.07$. There is a statistically significant correlation between the actual and predicted scores, $r = 0.67$, $P < 0.001$. There was no statistically significant difference between the two set of scores, $t(144) = –0.93$, $P = 0.36$, ES $r = 0.077$.

Model 3: a model for females

The strongest predictor-identified problems were very similar to those identified previously (see Table 2), and again when input into one model the combination that predicts the greatest amount of variance in the outcome variable (post-therapy CORE-OM score) is the same as that in the model for males. The descriptive statistics for the model are presented in Table 3 ($n = 1246$). There was a medium-sized relationship between the predictor variables and the outcome variable ($r = 0.59$). The predictors accounted for 35% of the variation in T_2 CORE-OM scores. The model was significantly better at predicting the outcome variable than using the sample T_2 CORE-OM mean, $F(4,1241) = 167.56$, $P < 0.001$, ES $r = 0.34$. The SEE value is 4.74. Fewer than 5% of cases have a standardised residual outside ±2. The model parameters are presented in Table 4. The regression equation for the model, to three significant figures, is:

$$OM = 0.0420 + (0.281pre_i) + (0.172dep_i) + (0.0932anx_i) + (0.0509int_i) + e_i$$

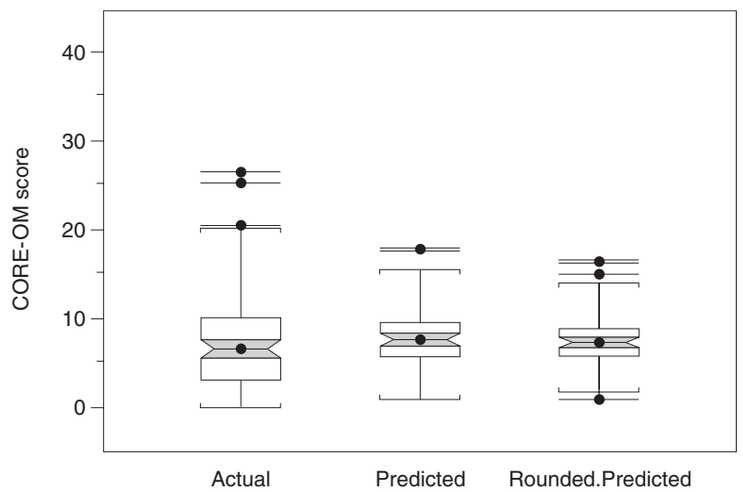

Figure 2 Testing the male model, dataset 2

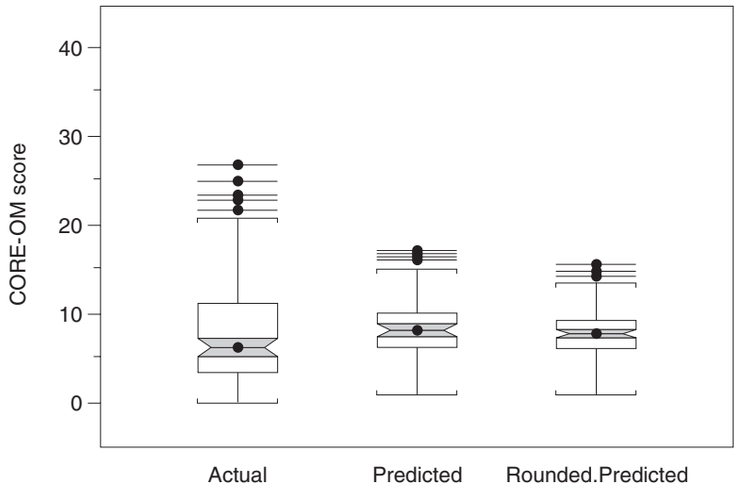

Figure 3 Testing the male model, dataset 3

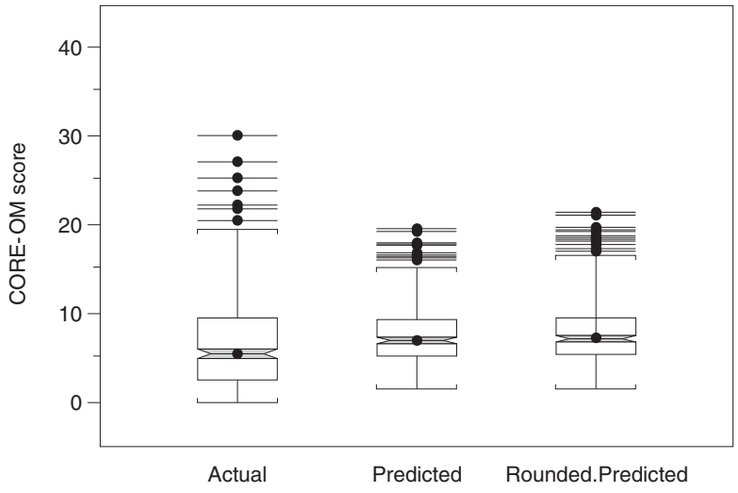

Figure 4 Testing the female model, dataset 2

Testing the model: dataset 2, n *= 326*

There was a difference between the two scores: predicted $M = 8.15 \pm 0.38$, SD = 3.50; actual $M = 7.32 \pm 0.59$, SD = 5.42; $ES_{diff} = 0.18$. The spread of scores is presented graphically in the left and middle plots of Figure 4 with the exact percentiles for the actual scores being 3.15, 6.15 and 10.00, and for predicted scores 5.58, 7.45, and 9.70. There was a statistically significant correlation between the actual and predicted scores, $r = 0.58$, $P < 0.001$. However, there was a statistically significant difference between the two sets of scores, albeit with a small effect size, $t(325) = -3.39$, $P < 0.001$, ES $r = 0.18$.

Testing the model: dataset 3, n *= 363*

There was a small difference between the two scores: predicted $M = 7.77 \pm 0.33$, SD = 3.22; actual $M = 8.09 \pm 0.60$, SD = 5.86; $ES_{diff} = 0.07$. The spread of the scores is presented graphically in the left and middle plots of Figure 5, for which the exact percentile values were 3.50, 7.10 and 10.90 for the actual scores and 5.46, 7.34 and 9.73 for the predicted scores. There was a statistically significant correlation between the actual and predicted scores, $r = 0.62$, $r < 0.001$, but no statistically significant difference between the two sets of scores, $t(362) = -1.34$, $P = 0.18$, ES $r = 0.07$.

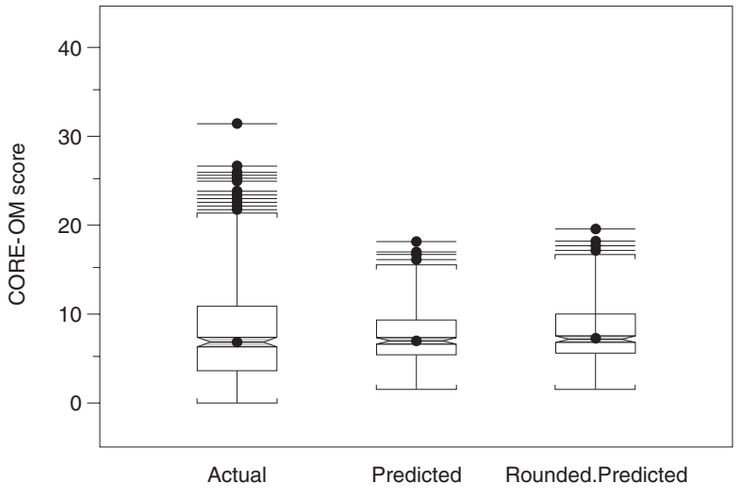

Figure 5 Testing the female model, dataset 3

Discussion

The aim of this study was to determine whether it was feasible to devise a method that would provide a 'good enough' proxy for clients' outcomes in the absence of their completing a measure at the end of therapy. In terms of the models presented, it is noteworthy that the best predictors are the combined factors of client pre-therapy score and the cluster of depression, anxiety, and interpersonal issues at the end of treatment, as perceived by the practitioner. The prominence of these components is consistent with the phenomenon that scores at T_1 are often the best predictors of the subsequent score.[26] In addition, the literature reporting the predominance of mixed depression and anxiety as well as the role of interpersonal relationships makes these four components eminently plausible from a clinical perspective.[29] Supporting this, the most common identified problems at post-therapy were depression (73% of clients), anxiety (66% of clients) and interpersonal issues (54% of clients).

The models provide different estimates of T_2 data depending on whether the focus is on the service as a whole or on males or females. Using dataset 1, all three models are statistically significant predictors of CORE-OM scores at T_2. Testing the models on dataset 2, all models had medium-sized correlations between actual and predicted scores. However, for the general and female models there is a statistically significant, albeit small, difference between predicted and actual scores. Testing the models on dataset 3, the correlations between actual and predicted results were all of a fair size and there were no significant differences between the results. It is recommended that the 'general' equation be used only if the individual sex equations cannot be used due to a lack of specificity within this model. The use of three different datasets illustrates that the model is generalisable at different levels of evaluation/audit.

The limitations of the models are most apparent when we consider the outliers within the actual datasets. Within a sample there will always be outliers due to the nature of distributions. However, a statistical model will never produce outliers because such data points do not follow the typical pattern within the data that the model is based upon. Outliers can have a disproportionately large effect on the model. However, the actual number of outliers is relatively small and such data points were investigated to establish that the extent to which they are untypical and that the effect that they are having on the model is not large.

A conceptual limitation of the procedures that have been carried out relate to a potential 'catch 22' phenomenon. The only way of testing whether the predicted models are plausible is to test them on data that have actual outcomes. However, by its very definition, it is likely that the clients – or their experience of psychological interventions – who complete T_2 data will differ in some way from clients who do not yield T_2 data. For example, it is suggested that those clients who drop out of therapy have the poorest outcomes, and there are certain groups of people who are more likely to drop out (e.g. minority, less-educated and lower income groups).[5,30] However, these patterns of predictors of therapy dropout are not strong enough that the type of missing data should be NI and not MAR. For example, there is evidence to suggest that clients may drop out of therapy simply because it was not needed any more – that is, for a positive and not a negative reason. Accordingly, their prospective CORE-OM

T_2 scores would not necessarily be worse than those of the completers (they may indeed be better), although they may be different.[31,32] Hence, by definition, the predicted scores should not be an exact match of the actual scores.

In practical terms, the equations can be used with a pre-completed EOT and T_1 CORE-OM, to predict an individual's score to assess the effectiveness of therapy for a particular client, or to predict more than one individual's score to enable group assessment of therapy outcome, for example for service/therapist/therapy evaluation. However, it should be noted that the equations have more accuracy for predicting T_2 data at a group rather than at an individual level, which is a function of the type of model used. The models are therefore best utilised when outcomes for a number of cases need to be evaluated, for example at a service level, in which there is likely to be 'routine' missing T_2 data, especially because the model is more accurate than simply deleting cases. This will also help to prevent the issue of possible conflicting interests, in which therapists are tempted to score a client who has not completed a T_2 CORE-OM as less severe, to improve their evaluation rates when utilising the models. Furthermore, although the models predict outcome significantly well, they do only explain 34% to 35% of the variance within outcome scores. Thus, 65% to 66% of the variance in outcome scores is unaccounted for by the models, and therefore other variables must influence CORE-OM T_2 scores. The equations are therefore valuable in that they provide a way of dealing with the high attrition rate for individuals in therapy which makes clinical governance problematic.

In order to make the models more accessible for practitioners, they have been simplified so they are more user friendly. Thus, the equations have been changed from three significant figures to one significant figure, which retains the values but with less specificity. This is important in terms of reliability. The most important components of the models are the variables involved (i.e. depression, anxiety etc). These are the factors that have been found to influence T_2 CORE-OM scores, and therefore it is assumed that if the methodology were to be repeated with a different primary care dataset then these variables would again be the predictors and would therefore remain constant. However, the numbers involved would change by an unknown small degree. Therefore it is concluded that in practice, rounding the numbers involved in the equations would have little impact on the predicted T_2 CORE-OM scores and not affect their reliability, and would make them easier for practitioners to use. Figures 2–5 (predicted and rounded predicted) show the difference between the specific models and the rounded models, and illustrate that there is little difference between the predicted values from these models. Box 1 provides an example of how to utilise the rounded male equation, and the same procedure can be used with the other two equations.

However, set against these efforts, it should be stated that such procedures are second best to organisational efforts that address the central issue as to why T_2 data completion rates are not higher. However, it is unlikely that any service could achieve a 100% completion rate. Hence, there is a need for both organisational and statistical supportive procedures, but it should clearly be the aim that the former achieves a majority completion rate perhaps something in the order of 6:4 or 7:3. As it is, these proportions are reversed, with the statistical procedures potentially being used to represent a majority of clients. These issues should be considered in conjunction with the guidance on the outcomes agenda disseminated in a report from the National Institute for Mental Health in England, which sets out differing levels of outcomes rollout.[33] However, for these to be achieved requires increased organisational resources to yield quality data from a majority of clients in routine service settings.

REFERENCES

1 Parry G. Evidence-based psychotherapy: an overview. In: Rowland N and Goss S (eds). *Evidence-based Counselling and Psychological Therapies: research and applications*. New York: Routledge, 2000, pp. 55–75.
2 Barkham M, Stiles WB and Shapiro DA. The shape of change in psychotherapy: a longitudinal assessment of personal problems. *Journal of Consulting and Clinical Psychology* 1993;61:667–77.
3 Lutz W, Leach C, Barkham M *et al*. Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology* 2005;73:904–13.
4 Powell AE, Davies HTO and Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Quality and Safety in Health Care* 2003; 12:122–8.
5 Wierzbicki M and Pekarik G. A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice* 1993;24:190–5.
6 Little RJA and Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, 1987.
7 Clarkin JF and Levy KN. The influence of client variables on psychotherapy. In: Lambert MJ (ed). *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change*. New York: John Wiley, 2004, pp. 194–226.
8 Little RJA. Missing data adjustments in large surveys. *Journal of Business and Economic Statistics* 1988;6:296–7.

9 Buck SF. A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society. Series B (Methodological)* 1960;22:302–6.
10 Raymond MR and Roberts DM. A comparison of methods for treating incomplete data in selection research. *Educational and Psychological Measurement* 1987;47:13–26.
11 Roth PL. Missing data: a conceptual review for applied psychologists. *Personnel Psychology* 1994; 47:537–60.
12 Wothke W. Longitudinal and multi-group modeling with missing data. In: Little TD, Schnabel KU and Baumert J (eds). *Modeling Longitudinal and Multiple Group Data: practical issues, applied approaches and specific examples*. Rahway, NJ: Lawrence Erlbaum Associates, 1998, pp. 219–40.
13 Barkham M, Margison F, Leach C *et al*. Service profiling and outcomes benchmarking using the CORE-OM: towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology* 2001;69:184–96.
14 Barkham M, Gilbert N, Connell J *et al*. Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry* 2005;186:239–46.
15 Evans C, Connell J, Barkham M *et al*. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry* 2002;180:51–60.
16 Evans C, Connell J, Barkham M *et al*. Practice-based evidence: benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology and Psychotherapy* 2003;10:374–88.
17 Gilbert N, Barkham M, Richards A *et al*. The effectiveness of a primary care mental health service delivering brief psychological interventions: a benchmarking study using the CORE System. *Primary Care Mental Health* 2005;3:241–51.
18 Connell J, Barkham M and Mellor-Clark J. Mental health norms using CORE-OM of students attending university counselling services benchmarked against an age-matched primary care sample. *British Journal of Guidance and Counselling* (in press).
19 Barkham M, Culverwell A, Spindler K *et al*. The CORE-OM in an older adult population: psychometric status, acceptability, and feasibility. *Aging and Mental Health* 2005;9:235–45.
20 Whewell P and Bonanno D. The care programme approach and risk assessment of borderline personality disorder: clinical validation of the core risk sub-scale. *Psychiatric Bulletin* 2000;24:381–4.
21 Mellor-Clark J and Barkham M. Quality evaluation: methods, measures and meaning. In: Feltham C and Horton I (eds). *Handbook of Counselling and Psychotherapy*. London: Sage Publications, 2000, pp. 255–70.
22 Mellor-Clark J and Barkham M. The CORE System: developing and delivering practice-based evidence through quality evaluation. In: Feltham C and Horton I (eds). *Handbook of Counselling and Psychotherapy* (2e). London: Sage Publications, 2006, pp. 207–24.
23 Barkham M, Connell J, Stiles WB *et al*. Dose–effect relations and responsive regulation of treatment duration: the good enough level. *Journal of Consulting and Clinical Psychology* 2006;74:160–7.
24 Mellor-Clark J, Connell J, Barkham M *et al*. Counselling outcomes in primary health care: a CORE system data profile. *European Journal of Psychotherapy Counselling and Health* 2001;4:65–86.
25 Dorz S, Borgherini G, Conforti D *et al*. Comparison of self-rated and clinician-rated measures of depressive symptoms: a naturalistic study. *Psychology and Psychotherapy: Theory, Research and Practice* 2004;77:353–61.
26 Hill CE and Lambert MJ. Methodological issues in studying psychotherapy processes and outcomes. In: Lambert MJ (ed). *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5e). New York: John Wiley & Sons, 2004, pp. 84–135.
27 Mellor-Clark J, Barkham M, Connell J *et al*. Practice-based evidence and standardised evaluation: informing the design of the CORE system. *European Journal of Psychotherapy Counselling and Health* 1999;2:357–74.
28 Leach C, Lucock M, Barkham M *et al*. Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *British Journal of Clinical Psychology* 45:153–66.
29 Sartorius N, Ustun TB, Lecrubier Y *et al*. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. *British Journal of Psychiatry* 1996;168(suppl 30):38–43.
30 Rainer J and Campbell LF. Premature drop-out in psychotherapy: identification and intervention. *Journal of Psychotherapy in Independent Practice* 2001;2:19–41.
31 Pekarik G. Improvement in clients who have given different reasons for dropping out of treatment. *Journal of Clinical Psychology* 1983;39:909–13.
32 April D and Nicholas LJ. Premature termination of counselling at a University counselling centre. *International Journal for the Advancement of Counselling* 1997;19:379–87.
33 National Institute for Mental Health in England (NIMHE). *Outcome Measures Implementation: best practice guidance*. London: NHS-NIMHE, 2005.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Michael Barkham, Psychological Therapies Research Centre, 17 Blenheim Terrace, University of Leeds, Leeds LS2 9JT, UK. Tel: +44 (0)113 343 5699; email: m.barkham@leeds.ac.uk

Received 3 October 2005
Accepted 19 April 2006