The contribution of therapist effects to patient dropout and deterioration in the psychological therapies

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Clinical Psychology &amp; Psychotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CPP-1124.R1</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Research Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Saxon, David; University of Sheffield, ScHARR Barkham, Michael; University of Sheffield, Psychology Foster, Alexis; University of Sheffield, ScHARR Parry, Glenys; University of Sheffield, ScHARR</td>
</tr>
<tr>
<td>Keywords:</td>
<td>therapist effects, outcomes, drop-out, deterioration, CORE</td>
</tr>
</tbody>
</table>
The contribution of therapist effects to patient dropout and deterioration in the psychological therapies

David Saxon (Centre for Psychological Services Research, University of Sheffield)*
Michael Barkham (Centre for Psychological Services Research, University of Sheffield)
Alexis Foster (School of Health & Related Research, University of Sheffield)
Glenys Parry (Centre for Psychological Services Research, University of Sheffield)

*Correspondence concerning this article should be addressed to David Saxon, School for Health & Related Research, University of Sheffield, 30 Regents Court, Sheffield S1 4DA, United Kingdom.

E-mail: d.saxon@sheffield.ac.uk
ABSTRACT

Background: In the psychological therapies, patient outcomes are not always positive. Some patients leave therapy prematurely (dropout) while others experience deterioration in their psychological wellbeing.

Methods: The sample for dropout comprised patients (N = 10,521) seen by 85 therapists and who attended at least the initial session of 1-to-1 therapy and completed a Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) at pre-treatment. The sub-sample for patient deterioration comprised patients (N = 6,405) seen by the same 85 therapists but who attended 2 or more sessions, completed therapy, and returned a CORE-OM at pre- and post-treatment. Multilevel modeling was used to estimate the extent of therapist effects for both outcomes after controlling for patient characteristics.

Results: Therapist effects accounted for 12.6% of dropout variance and 10.1% of deterioration variance. Dropout rates for therapists ranged from 1.2% - 73.2%, while rates of deterioration ranged from 0% - 15.4%. There was no significant correlation between therapist dropout rate and deterioration rate (Spearman’s rho = 0.07, p=0.52).

Conclusions: The methods provide a reliable means for identifying therapists who return consistently poorer rates of patient dropout and deterioration compared to their peers. The variability between therapists and the identification of patient risk factors as significant predictors have implications for the delivery of safe psychological therapy services.
Key practitioner message:

- Therapists play an important role in contributing to patient dropout and deterioration, irrespective of case mix.

- Therapist effects on patient dropout and deterioration appear to act independently.

- Being unemployed as a patient was the strongest predictor of both dropout and deterioration

- Patient risk to self or others was also an important predictor

Keywords: Deterioration, dropout, outcomes, variability, therapist effects, CORE-OM
Introduction

Background

Clinical practice and research have, understandably, focused on the improvement patients experience when engaging in a course of psychological therapy (Lambert, 2013). However, outcomes for patients are not always positive and patients may leave therapy prematurely without making meaningful improvement (Cahill et al., 2003). Moreover, others may experience deterioration in their psychological wellbeing during the course of therapy (Craze et al., 2014; Lambert, 2010). There has to date, however, been limited research into negative outcomes in routine services and few have considered therapist effects on those outcomes. In part this has been due to the absence of sufficiently large datasets to study therapist effects, but also to inconsistencies in the definitions of the range of negative outcomes. Linden (2013) classified negative outcomes, as different types of ‘unwanted events’, some of which are adverse reactions to the therapy, while others may or may not be therapy related. Two manifestations of the latter are unplanned endings, often termed dropout, and patient deterioration.

Patient dropout

Patient dropout from therapy has been of concern in the psychological therapies for over 50 years and continues to have implications for service delivery and patient outcomes (Barrett, Chua, Crits-Christoph, Gibbons, & Thompson, 2008; Garfield, 1994; Rogers, 1951). Dropout occurs where a patient unilaterally ends therapy by ceasing to attend sessions, prior to the endpoint planned with their therapist (Westmacott, Hunsley, Best, Rumstein-McKean, & Schindler, 2010). The reported rates of dropout have ranged between 20-60% depending on the patient population, service setting, how dropout has been defined, and the methodology adopted (for details, see Reneses, Munoz, & Lopez-Ibor, 2009). A meta-
analysis of 669 studies of psychological and psychosocial interventions reported a dropout rate of 17% for efficacy studies and 26% for effectiveness studies (Swift & Greenberg, 2012).

In the UK, successive national audits by the Royal College of Psychiatrists (RCP) of psychological therapy services have reported treatment dropout rates of 25% and 24% respectively (RCP, 2011, 2013), while a report on 32 UK services comprising the initial national rollout of the Improving Access to Psychological Therapies (IAPT) initiative yielded a rate of 21.6% (Glover, Webb, & Evison, 2010). However, these UK reports did not include patients who failed to engage with therapy. These patients attended only one appointment, which has been consistently found to be the modal number of psychotherapy sessions attended (e.g., Gibbons et al., 2010). The current study considers patient dropout at any point after the initial session.

Patient deterioration

Patient deterioration, a shorthand term for deterioration in a patient’s mental state after therapy, may be defined as any negative change between pre- and post-therapy outcome score. Because this definition would include small changes that may be due to the inherent unreliability of outcome measures (Jacobson & Truax, 1991), a more stringent criterion of ‘statistically reliable deterioration’ has been adopted by researchers (as a mirror opposite of reliable improvement) in which measurement error is taken into account. Using this procedure to determine rates of reliable deterioration based on selected completer samples has yielded an estimate for primary care of 1.5% (Cahill, Barkham, & Stiles, 2010) and upwards of 6% for secondary care (Barkham et al., 2001). Reports from the US have tended to yield higher rates; for example, an average figure of 8.2% across a range of different clinical settings (Hansen, Lambert, & Forman, 2002).
However, it is debatable whether the criterion for deterioration should be the same as for improvement. The natural propensity for patient recovery, the normative trajectory of patient change, and any statistical regression to the mean, make therapy more likely to lead to some level of improvement rather than deterioration. In the same way that Linden (2013) argues that if therapy does not produce the expected outcome (i.e., improvement), then the outcome is an ‘unwanted event’, then reliable deterioration should not be viewed as a mirror opposite of reliable improvement. Practitioners are likely to want to be flagged about possible deterioration in their patients at a less stringent threshold than improvement. Furthermore, services should be concerned if some of their practitioners have significantly more patients who deteriorate compared to their peers, when a less stringent threshold is used.

**Therapist effects**

The study of therapist effects focuses on the extent of variability between therapists and the impact the individual therapists have on patient outcomes. The recommended methods for estimating such effects, for example multilevel modeling (Goldstein & Spiegelhalter, 1996; Snijders & Bosker, 2012), require large samples of patients, and in particular therapists (Maas & Hox, 2005). Randomised controlled trials (RCTs) are usually underpowered to estimate therapist effects and very large datasets drawn from routine practice are best suited to provide the statistical power and external validity needed in this field (e.g., Castonguay, Barkham, Lutz, & McAleavy, 2013; Wampold & Brown, 2005). Most studies of therapist effects have considered positive outcomes such as clinical improvement or recovery rates and there is a relative paucity of research into therapist effects on negative outcomes (Baldwin & Imel, 2013). An exception is a recent study of patient dropout, using multilevel modeling (MLM), which found a significant therapist effect.
(6.21%), after controlling for initial impairment, although the sample size, particularly the number of patients per therapist, was a recognised limitation (Zimmerman, Rubel, Page & Lutz, submitted 2016). There have been no studies to date which have used MLM to estimate therapist effects for patient deterioration. Krause et al (2011) analysed the outcomes for 696 therapists in the context of naturalistic treatment and found some therapists demonstrated large, negative treatment effect sizes ($d = -0.91$ to $-1.49$).

However, case mix was not controlled for in the analysis.

**Case-mix**

In order to make valid comparisons between therapists’ outcomes it is necessary to control for patient characteristics that have a significant impact on outcome (i.e. case-mix). Some likely candidates for patient dropout are: younger age (e.g., Edlund et al., 2002); non-white ethnicity and socio-economic deprivation (e.g. Garfield, 1994) and greater intake severity (Kazdin, Mazurick, & Siegel, 1994, Zimmerman et al, submitted 2016).

Few studies have considered the patient characteristics associated with deterioration and one study failed to identify any statistically significant predictors of reliable deterioration in a sample of 1416 UK outpatients (Shepherd, Evans, Cobb & Ghossain, 2012).

In the development of models for both dropout and deterioration, the current study will test all available patient variables as possible case-mix variables.

**Study aims**

In the current study, we employed a large-scale practice-based dataset to estimate the extent of therapist effects, while also controlling for those patient variables that have a significant impact on outcome.

Accordingly, the study had three aims:

1) To estimate the therapist effect for patient dropout using MLM.
2) After applying varying indices of deterioration to the data, to estimate therapist effects on patient deterioration for treatment completers.

3) To combine the variability between therapists on both dropout and deterioration and consider whether those therapists with higher dropout rates are also those therapists with higher deterioration rates for their treatment completers.

**Method**

*Original dataset*

The original data set – the Clinical Outcomes in Routine Evaluation Practice-Based Evidence National Database-2008 – comprised information on 70,245 clients, routinely collected by 1,059 therapists in 35 UK counselling and clinical psychology services between 1999 and 2008. This data set was an updated version of earlier datasets used in studies by our research group (e.g., Stiles, Barkham, Connell, & Mellor-Clark, 2008). Ethics approval was covered by the UK National Health Service’s Central Office for Research Ethics Committee, application 05/Q1206/128.

*Study-specific dataset*

For the current study, in order to exclude practitioners who may have been selective in their submission of patient data, therapists were only included if they provided treatment ending information for over 90% of the patients they treated. The figure of 90% was chosen as it is a target for the UK Improving Access to Psychological Therapies initiative (Department of Health, 2012). Patients were included if they were 18 years old and over, were assessed and accepted for individual therapy, completed a specified pre-therapy outcome measure (see below), provided demographic data, and had the type of therapy ending recorded.
addition, in order to estimate therapist variability more reliably, only therapists with 30 or
more patients were included (Soldz, 2006).

These criteria yielded a study-specific sample of 85 therapists and 10,521 patients from
14 sites with a range of patients per therapist of 30 – 468. In this sample, the patient mean
(SD) age was 40.3 (13.00) years, 71.2% were female, 23.9% were unemployed, and 4.6%
were of non-white ethnicity. No formal diagnoses were made but therapists recorded
patients’ problems on a standardized form (CORE Assessment form; Barkham, Gilbert,
Connell, Marshall, & Twigg, 2005). This indicated that 76.8% of patients had some level of
depression (44.7% rated as ranging between moderate and severe) and 82.7% had some
level of anxiety (54.6% rated as ranging between moderate and severe).

Deterioration sub-sample

The deterioration dataset was a sub-sample of the study-specific dataset. It comprised
patients who completed therapy, had two or more sessions, and provided a pre- and post-
therapy CORE-OM score. This yielded 6,405 patients, with the same 85 therapists, who saw
between 13 –180 patients each. Therapists with less than 30 patients were not excluded, in
order to compare all 85 therapists on both outcomes. The mean (SD) age of this sub-sample
was 41.9 (13.02) years, 71.6% were female, while 21.0% of patients were unemployed and
3.8% were non-white. A flowchart describing how the samples of patients ($N_p$) and
therapists ($N_T$) were derived is presented in Figure 1.

Baseline and outcome variables

Baseline patient demographic and severity data were collected using the CORE Assessment
form (Barkham et al., 2005) and CORE-OM (Barkham et al., 2001; Evans et al., 2002). The
CORE-OM is a self-report measure of a patient’s condition over the past week and
comprises 34 items addressing the domains of subjective wellbeing, symptoms, functioning,
and risk. The risk domain captured both risk-to-self (4 items: e.g., I have made plans to end my life) and risk-to-others (2 items: e.g., I have been physically violent to others). Items are scored on a 0 to 4 scale and yield an overall CORE-OM score that can be separated into a CORE non-risk score and CORE risk score, each with a range from 0 to 40. The 34-item scale has a reported internal consistency of .94 (Barkham et al., 2001) and a one-month test-retest correlation of .88 (Barkham, Mullin, Leach, Stiles, & Lucock, 2007).

Patients completed the CORE-OM prior to therapy and at the end of their final treatment session. Therefore final outcome scores were not available for patients that dropped out of therapy. The two outcomes for the study were whether patients had completed or dropped out of therapy, as recorded by the therapist at case closure and whether those patients that completed therapy deteriorated or not as reflected in their CORE-OM score.

Reliable change in CORE-OM scores has been defined as a pre-post change in CORE-OM scores of five points or more (Connell et al., 2007). However, for the reasons stated above and due to the rarity of reliable deteriorations, pre-post deteriorations of fewer than five points were also considered.

**Analysis**

Subsequent to describing patient intake severity and patient outcomes, MLM was used to produce a multilevel model for each outcome. MLM is a recommended method where there is a hierarchical structure in the data (i.e., where patients at level 1 are ‘nested’ within therapists at level 2) and differences between the higher-level units (i.e., therapists) are of interest (Goldstein & Spiegelhalter, 1996; Snijders & Bosker, 2012). Explanatory variables were added to the models, with continuous variables grand mean centred (Hofmann & Gavin, 1998) and tested for significance by dividing the derived coefficients by their standard errors. Values greater than 1.96 were considered significant at the 5% level.
Multilevel modeling software, MLwiN v2.30 (Rasbash, Charlton, Browne, Healy, & Cameron, 2009) was used to estimate the parameters in each model, initially by marginal quasi-likelihood (MQL) methods, before applying these estimates as ‘priors’ for Markov chain Monte Carlo (MCMC) estimation procedures. This simulation approach produces a large number of estimates of the unknown parameters that can be summarised to both a mean estimate and a 50th percentile estimate. In addition, a 95% probability interval (PrI), analogous to 95% confidence intervals, can be taken as the 2.5 and 97.5 percentile values (Browne, 2012). During development, MCMC models were compared using the Deviance Information Criteria (DIC), which balances ‘fit’ and ‘complexity’, with reductions in DIC indicating improvements in the model fit (Spiegelhalter, Best, Carlin, & van der Linde, 2002).

Because the study samples used in these analyses are much reduced compared with the full dataset, sensitivity analysis was carried out. Logistic regression models were developed for larger data samples, where exclusion criteria were not applied, and the included predictor variables and their odds ratios (ORs) were compared with those derived from the smaller study samples.

The therapist effect on outcome is defined as the percentage of the total variance that is at level 2 (therapist level). In the current study, variance on the logistic scale derived from a linear threshold method was used (Rasbash et al., 2009; Snijders, & Bosker, 2012). Assumptions of normality in the data were tested by plotting the patient level and therapist level residuals produced by the model to normal distribution curves using quantile-quantile (q-q) plots.

The residual for each therapist represents the degree to which a therapist’s outcomes depart from those of the average therapist while controlling for patient characteristics (case-mix) and can be seen as the additional, unexplained impact of the therapist on

John Wiley & Sons
outcome (Goldstein & Spiegelhalter, 1996; Rasbash et al., 2009; Saxon & Barkham, 2012).

The therapist residuals from the dropout and deterioration models were considered separately by ranking and plotting with confidence intervals (CIs; Goldstein & Healy, 1995; Rasbash et al., 2009). Thus for each outcome, therapists could be described as average, where their CI crossed the average (residual = 0) in their impact on outcome, while those that did not cross the average were identified as significantly above or below average.

The therapist residuals from the two models were also plotted against each other as a scatterplot, placing each therapist in one of four quadrants: Quadrant 1 comprising those therapists better than average on both outcomes; Quadrant 2 those therapists worse than average on both outcomes; and in Quadrants 3 and 4, those therapists better on one and worse on the other outcome.

**Results**

The results are presented in three main sections, reflecting the three study aims. The two sections on dropout and deterioration begin with descriptives of the samples, followed by descriptions of the multilevel models and the reporting of therapist effects. The models and significant case-mix variables are presented in Appendix A, Appendix B and Table 1. The third section of the Results compares and combines the results found for dropout and deterioration.

**Patient dropout**

For the dropout sample (N = 10,521), the proportion of patients who dropped out of therapy was 33.8%, with over half of these (52.7%) dropping out before session 3. The mean (SD) number of sessions attended for dropouts was 2.8 (1.91) sessions, compared with 6.1 (2.68) for treatment completers. The mean (SD) patient dropout rate for therapists was 31.5% (13.8) with a range between 1.2% - 73.2% (IQR: 23.6% - 39.9%).

John Wiley & Sons
The mean (SD) patient CORE-OM score at intake was 18.1 (6.31) with 90.0% of patients scoring above the clinical cut-off score of 10. For patients who dropped out of therapy (N=3,554), the mean (SD) intake score was 18.9 (6.28) and 91.8% were above clinical cut-off. This compares to 17.8 (6.28) and 89.1% for patients who completed therapy (N=6,967).

**Dropout model development**

A single level logistic regression model containing significant predictors of outcome (dropout or not) was developed, prior to extending it to a multi-level model to allow for therapist variability. Following MCMC procedures, the difference between the DICs of the multilevel model compared to the single level model (688.7) indicated that the multilevel model was a better fit for the data. Tests of convergence showed a chain length of 57,000 iterations to be sufficient and q-q plots were fairly linear, indicating that Normality can be assumed. The dropout multilevel model is presented in Appendix A.

Table 1 shows the patient variables identified as predictors of dropout, with their odds ratios (ORs) and 95% probability intervals (PrIs) produced by the exponentials of the 2.5, 50 and 97.5 percentile values for the model coefficients. Patients who were younger, non-white, unemployed, or had higher CORE non-risk scores were more likely to drop out.

In addition, patients answering in the affirmative (either: *only occasionally*, *sometimes*, *often*, or *most of the time*) to the risk questions ‘I have hurt myself physically or taken dangerous risks with my health’ (N=850; OR=1.19) and ‘over the past week I have been physically violent to others’ (N=534; OR=1.39), were both predictive of dropout compared to patients indicating no risk on these items. There were no significant interactions between variables in the model. In relation to risk, this suggests that the two questions, ‘risk to self’ and ‘risk to others’, are identifying two separate types of risk. This is supported by the data showing that of those patients reporting risk on either item (N=2,316), only 19% scored on...
both items. No significant random slopes were found, indicating that each of the variables in the model impacted on outcomes similarly for all therapists.

Sensitivity analysis was carried out on a sample (N=38,354), representing all patients accepted for therapy (N=55,070) minus those with missing data (N=16,715). A single level logistic regression model produced by the larger data sample contained the same significant variables as above and minimal differences in ORs. The variable showing the greatest difference was 'Ethnicity' with an OR (95% PrI) of 1.12 (1.01, 1.23) in the larger sample compared with 1.29 (1.05, 1.59) in the study sample.

**Therapist effects for dropout**

Individual therapists had a varying impact on outcome after controlling for the significant patient predictors identified above, with a significant therapist effect (95% PrI) of 12.6% (9.1, 17.4). No therapist factors were available but number of patients per therapist was considered in the model and was found to have minimal effect, reducing the therapist effect to 12% but indicating a poorer model fit (larger DIC). Therefore the final treatment dropout model (Appendix A) included only patient variables.

Figure 2 plots the therapist intercept residuals (with 95% CIs) produced by the model for the 85 therapists ranked best to worst, from left to right. The plot shows that the majority of therapists (61.1%), shown in grey, had treatment ending outcomes that were not significantly different to the average therapist (indicated by the dashed horizontal line where the residual is zero), while 13 (15.3%) therapists, on the left of the chart, had significantly better than average outcomes and 20 (23.5%), on the right of the chart, had outcomes that were significantly poorer than average. In order to gauge the actual differences in dropout rates between these three groups of therapists, their aggregated means were calculated. The aggregated mean (SD) dropout rate for average therapists was
29.7% (6.4), while for above average therapists it was 12.0% (7.3) compared with 49.0% (10.4) for below average therapists.

**Patient Deterioration**

For the deterioration sample (N=6,405), where patients completed therapy, the mean (SD) CORE-OM score at intake was 17.8 (6.23) while the proportion scoring above clinical cut-off was 89.1%. Their mean (SD) outcome score of 8.9 (6.25) yielded a pre-post effect size of 1.43. Most patients (72.2%) improved by 5 points or more on the CORE-OM and could be considered reliably improved, while 26.8% made no reliable change, 6.2% deteriorated to some degree, and 1.0% reliably deteriorated. The mean (SD) reliable deterioration rate for therapists was 1.2% (1.67) with a range between 0% and 7.1% (IQR: 0 - 1.9%).

Table 2 shows the deterioration rates for six different levels of deterioration, ranging from any change on the CORE-OM to a change of ≥5 CORE-OM points (the degree of change considered as reliable deterioration) and the number of therapists that had no deteriorations for each level. There were significant positive correlations (one-tailed, all \( p \)-values<0.001) between the different rates and rankings for therapists. Correlation coefficients (Spearman’s \( \rho \)ho) ranged from .50 for the association between ‘any deterioration’ and ‘≥5’ point change, to .92 for the association between ‘any deterioration’ and ‘>1’ point change.

The large proportion of therapists with no deteriorations was problematic in multilevel model development and only where deterioration was defined as ‘any deterioration’ or ‘> 1’ did the models stabilise to produce reliable estimates of therapist effects. Therefore, a model with deterioration at the level of ‘more than 1 point’ was used as the patient outcome in the multilevel analysis. The correlation between therapists ranked using this level of deterioration and reliable deterioration (‘≥5’) was .56 (\( p < .001 \)).
Deterioration model development

As with the dropout model, a single level logistic regression model containing significant
predictors of outcome was extended to allow for therapist variability. Following MCMC
procedures, the change in DIC of 45.9 indicated the multilevel model to be a better fit for
the data than the single level model. Tests of convergence indicated that the chain length of
128,000 iterations was sufficient and the q-q plots were fairly linear, indicating that
Normality can be assumed. The deterioration model is presented in Appendix B.

Table 1 shows the patient variables identified as predictors of deterioration by more than 1
point. Patients who were older and less severe at intake were more likely to deteriorate.
However, the latter is likely to be a statistical factor with higher scores having less scope to
deteriorate. The risk item ‘I have thought of hurting myself’, was a significant predictor of
deterioration (N=1,829; OR= 1.55) and, consistent with the drop-out model, patients who
were unemployed were more likely to deteriorate than patients not unemployed. Again,
there were no interactions between variables and no significant random slopes on any of
the predictor variables indicating that they have a similar impact on outcome for all
therapists.

Sensitivity analysis, using the largest possible sample (N = 24,499) representing all those
patients who completed therapy and had a pre and post CORE OM score (N = 30,978) minus
those with missing variable data (N= 6,479), produced a logistic regression model containing
the same four predictor variables as in Table 1. The ORs for age and CORE non-risk score
were almost identical to those produced by the smaller sample. The ORs (95%PrI) for
unemployment and the risk question, of 2.04 (1.22, 2.33) and 1.41 (1.21, 1.66) respectively,
were reduced, although for both variables the PrIs overlap their corresponding PrIs derived from the smaller samples.

_Therapist effects for deterioration_

The therapist effect for deterioration of more than 1 point was 10.1% (95% PrI: 4.9, 17.8).

The number of patients per therapist was not significant in the model. As with the caterpillar plot for dropout (Figure 2), Figure 3 plots the therapist intercept residuals produced by the deterioration model (with 95% CIs) for the 85 therapists ranked best to worst, from left to right.

Indicative of the rarity of the event and the smaller numbers of patients per therapist, the 95% CIs are generally wider than in the dropout model, with only one therapist being significantly better than average, and four therapists significantly worse than average. The vast majority of therapists (94.1%) could be considered average with regard to patient deterioration, they had an overall mean (SD) deterioration rate of 4.6% (3.7). The better than average therapist had no patients who deteriorated, while for the four below average therapists, their rates of deterioration were, from left to right, 11.8%, 12.1%, 14.1% and 14.9%. The statistically reliable deterioration rates (deterioration by ≥5 points) for these four therapists, were 1.5%, 3.5%, 3.1% and 3.0% respectively, compared with a mean (SD) rate of 1.1% (1.7), for the average therapists.

_Combining therapist variability on dropout and deterioration_

In order to consider whether those therapists with more treatment dropouts also had more treatment completers that deteriorated, the therapist rankings and residuals from Figures 2 and 3 were compared. There was no significant correlation between the rankings (Spearman’s rho = 0.07, $p=0.52$), suggesting that, overall, therapists that were less able to retain patients in therapy did not generally have more patients that deteriorated after
completing treatment. To consider the relationship between the two outcomes further, the therapist residuals for each outcome were plotted against each other in a scatterplot (Figure 4).

In Figure 4, the $x$-axis measures the therapist residual for dropout, while the $y$-axis measures the therapist residual for deterioration. Zero on each axis represents the average therapist and each therapist is placed in a quadrant of the plot based on their residuals derived from each model. The 20 therapists significantly below average for dropout are represented by black circles while the four therapists identified as significantly below average for deterioration are represented by grey squares. The 95% CIs from Figures 2 and 3, which would be represented by a cross through every therapist point, are not shown, but in all instances at least one CI crossed zero. Therefore, no therapist was found to be significantly below average on both outcomes.

**Discussion**

In this practice-based study, our aim was to establish the degree to which therapists contribute to the variability in two negative patient outcomes, namely unplanned endings (i.e., dropouts) and deterioration. For both outcomes, we found significant therapist effects, of 12.6% and 10.1% respectively, that were larger than the range of effects (5%-8%) found in similar studies of patient improvement (e.g., Lutz, Leon, Martinovich, Lyons, & Stiles, 2007; Wampold & Brown, 2005). In a context where the overall effect of therapy, which includes all aspects of therapy including therapist factors, treatment adherence, and alliance is estimated at 20% (Baldwin & Imel, 2013), these therapist effects of over 10% are both statistically significant and clinically important.

Locating the focus for patient outcomes with the therapist supports findings from studies of addiction services (Brorson, Arnevik, Rand-Hendriksen, & Duckert, 2013) and adolescent
services (de Haan, Boon, de Jong, Hoeve, & Vermeiren, 2013). These studies concluded that
the simple study of patient variables in isolation was of limited value and the study of such
factors as the alliance and therapist variables would be more useful, in part because they
are variables that can be changed (de Haan et al., 2013).

Therapist variables that have been associated with negative outcome include lack of
empathy, negative countertransference, overuse of transference interpretations, and
disagreement with patients about therapy process (Mohr, 1995). Type and amount of
training, theoretical orientation, and gender were not predictive of patient outcome (Okiishi
et al., 2006), while studies of therapist competence, have yielded contradictory results
(Ginzburg et al., 2012; Webb, de Rubeis & Barber, 2010). Our finding that those therapists
worse than average for dropout were no worse than average for deterioration, suggests
that different therapist factors may be associated with different negative outcomes. Further
research is necessary to identify therapist factors and their interactions with patient
characteristics that may explain the degree of variability between therapists in their
negative outcomes.

Our finding that around a third of patients dropped out of therapy is within the range
of 20%-60% reported elsewhere (Reneses et al., 2009) and is 10% larger than reported
rates where session 1 was excluded (e.g., RCP, 2011; 2013). The mean therapist rate of
31.5% was similar to the 33.2% found by Zimmerman et al (2016), however, our therapist
effect for dropout was twice that found in their study. We can only speculate as to why
there was such a difference, but reasons may include differences in methodology
(Goldstein, Rasbash & Browne, 2002; Snijders & Bosker, 2012), sample size (Soldz, 2006),
service delivery models and available patient variables.
Our analysis identified 23.5% of therapists whose dropout rates were significantly higher than average. Aggregated dropout rates indicated that patients seen by these therapists were around four times more likely to dropout than patients seen by the 15.3% of therapists who had significantly lower than average dropout (49% compared with 12%).

The results for patient deterioration were less reliable, reflected in the wide Probability Interval for the therapist effect and the wider confidence intervals for therapist residual estimates. This unreliability was due to the rarity of the outcome, the smaller number of patients per therapist and the adoption of a measure of deterioration that was less than ‘statistically reliable’. That said, where patient safety and possible harm are paramount, it would seem appropriate to ‘flag’ therapists at the below average end, as soon as possible, regardless of the confidence intervals or number of patients they have treated.

We found significant outcome variability between therapists, with patients seen by therapists identified as below average being over twice as likely to deteriorate as patients seen by therapists identified as average. That those therapists identified as below average, using our less stringent criteria, also showed higher than average rates of reliable deterioration suggests that the model is correctly identifying therapists with higher rates of negative change.

**Case-mix variables**

A number of patient variables were significant predictors of outcomes and were controlled for in estimating the impact of the therapist. We found that these variables effected therapists similarly, i.e., there were no random slopes. For dropout, the patient variables identified were similar to those reported elsewhere: greater symptom severity at intake (Kazdin, Mazurick, & Siegel, 1994); younger age (e.g., Edlund et al., 2002), and non-
white ethnicity and unemployment, which may be proxy measures of socio-economic
deprivation (Garfield, 1994; Wierzbiki & Pekarik, 1993; Williams, Ketring, & Salts, 2005). In
addition, and possibly of greater concern, was the finding that patients at risk of harming
themselves or others were more likely to dropout than patients with no risk, a finding that
supports previous research from a single service study using CORE risk items (Saxon,
Ricketts & Heywood, 2010). We found that patients who had been ‘physically violent to
others’ were 39% more likely to dropout than those who had not.

For deterioration, we found that in addition to answering in the affirmative to the risk
question ‘over the past week I have thought of hurting myself’, patient age and employment
status were also predictive of outcome. Younger patients were more likely to drop out than
older patients, but if they completed therapy they were less likely to have deteriorated,
while unemployed patients were 44% more likely to drop out than patients who were not
unemployed, and if they stayed in therapy to a planned ending they were more than twice
as likely to have deteriorated.

Study Limitations

Crucial in any practice-based study is the issue of the representativeness of included data
(Brown, Lambert, Jones, & Minami, 2005). In order to reduce any bias due to the failure to
collect data from patients, only those therapists who provided data for over 90% of their
patients were included, therefore results may only be generalizable to therapists with high
return rates. Also, our sample contained counsellors and clinical psychologists in primary
care who had seen at least 30 patients for dropout or 13 for deterioration, therefore results
may not be generalizable to therapists who have seen fewer patients or deliver other types
of therapy in different settings. The small number of sites, and therapists per site, prevented any analysis of the impact treatment sites might have on both outcomes.

In addition to concerns about the reliability of the deterioration analysis outlined above, the deterioration rates reported here may underestimate actual rates as they are based on treatment completers only. No last CORE-OM was available for patients who dropped out, therefore it was not possible to measure their pre-post change, but research indicates that they are likely to have had poorer clinical outcomes (Delgadillo et al., 2014; Saxon, Firth & Barkham, submitted 2016). To address these limitations, it would be informative to replicate this analysis with a larger multi-site dataset that contains a wider range of patient and therapist variables and outcome measures for the last session attended.

**Clinical and service implications**

These results have important implications for quality improvement in psychological therapy services. Services may not be meeting the needs of some sections of the community and should take steps to better engage patients who are younger, of non-white ethnicity or unemployed. With regard to risk, heightened patient risk may be associated with greater severity and complexity of condition and possible borderline personality disorder.

Guidelines suggest that brief, psychological therapies in primary care are unsuitable for patients with borderline personality disorder, who have higher levels of self-harm, or anti-social personality disorder where higher levels of aggression are characteristic, therefore patients may need to be referred-on to more appropriate services (NICE, 2009a,b).

Our results show that patient characteristics alone cannot account for drop out and clinical deterioration and that therapists account for a large proportion of the variance in these negative outcomes. This is an important factor that is often neglected, for example when
considering ways of reducing early withdrawal from treatment (Barrett et al. 2008). The
implication is that therapists who are below average for negative outcomes should be made
aware of this so that remedial action, for example through greater support, supervision or
training, can be taken. However, caution is necessary because although the statistical
methods employed in this study can raise questions about therapist outliers, other
unmeasured factors may influence therapist performance. Therapists and service managers
need to use these methods only as a starting point for exploration.

Conclusion

In conclusion, using sophisticated and appropriate methods, we found large therapist
effects for both types of negative outcomes, indicating significant variability between
therapists in their ability to retain patients in therapy and to prevent patient deterioration.
This study illustrates that the reporting of simple aggregated outcomes for services and
practitioners, usually focused on improvement and recovery, is limited and may mask
important factors for safe and effective service delivery in the psychological therapies.
Acknowledgements

We thank CORE IMS for the collection of the data and in particular John Mellor-Clark and Alex Curtis Jenkins.

Conflict of interest

MB was a member of the research group that developed the CORE-OM.

Funding source

The research was funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0408-15144). They had no involvement in any aspect of this manuscript.
References


John Wiley & Sons


Table 1: Odds Ratios for the predictor variables in each model, with their 95% Probability Intervals (PrIs)

<table>
<thead>
<tr>
<th>Variable in model</th>
<th>Drop-out Model</th>
<th>Deterioration Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployed</td>
<td>1.44 (1.30, 1.60)</td>
<td>2.71 (2.05, 3.57)</td>
</tr>
<tr>
<td>Age - grand mean</td>
<td>0.97 (0.96, 0.97)</td>
<td>1.02 (1.01, 1.03)</td>
</tr>
<tr>
<td>CORE non-risk – grand mean</td>
<td>1.02 (1.01, 1.02)</td>
<td>0.90 (0.88, 0.92)</td>
</tr>
<tr>
<td>Ethnicity (not white)</td>
<td>1.29 (1.05, 1.59)</td>
<td>NS</td>
</tr>
<tr>
<td>‘I have been physically violent to others’</td>
<td>1.39 (1.21, 1.60)</td>
<td>NS</td>
</tr>
<tr>
<td>‘I have hurt myself physically or taken dangerous risks with my health’</td>
<td>1.19 (1.05, 1.34)</td>
<td>NS</td>
</tr>
<tr>
<td>‘I have thought of hurting myself’</td>
<td>NS</td>
<td>1.55 (1.12, 2.14)</td>
</tr>
</tbody>
</table>
Table 2: The number of patients who deteriorated for each level of deterioration, ranging from any deterioration to deterioration of 5 points or more, the mean (SD) deterioration rates for therapists (N=85) and the number of therapists with no deteriorations at each level.

<table>
<thead>
<tr>
<th>Deterioration</th>
<th>Patient rate</th>
<th>Therapist rate</th>
<th>Therapist rate</th>
<th>N (%) of therapists with no deteriorations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterioration</td>
<td>N (%)</td>
<td>Mean (SD)</td>
<td>Range %</td>
<td>N (%)</td>
</tr>
<tr>
<td>Any deterioration</td>
<td>399 (6.2)</td>
<td>6.8 (5.29)</td>
<td>0 – 28.6</td>
<td>11 (12.9)</td>
</tr>
<tr>
<td>&gt;1 point</td>
<td>287 (4.5)</td>
<td>5.0 (4.12)</td>
<td>0 – 15.4</td>
<td>16 (18.8)</td>
</tr>
<tr>
<td>&gt;2 points</td>
<td>191 (3.0)</td>
<td>3.2 (2.87)</td>
<td>0 – 10.3</td>
<td>24 (28.2)</td>
</tr>
<tr>
<td>&gt;3 points</td>
<td>134 (2.1)</td>
<td>2.2 (2.17)</td>
<td>0 – 7.7</td>
<td>31 (36.5)</td>
</tr>
<tr>
<td>&gt;4 points</td>
<td>93 (1.5)</td>
<td>1.2 (1.96)</td>
<td>0 – 7.7</td>
<td>36 (42.4)</td>
</tr>
<tr>
<td>5 or more points</td>
<td>67 (1.0)</td>
<td>1.2 (1.67)</td>
<td>0 – 7.1</td>
<td>44 (51.8)</td>
</tr>
</tbody>
</table>
Appendix A: Multilevel model for patient dropout. Key: Ending\(_{ij}\) : Outcome (drop-out or not) for patient \(i\) seen by therapist \(j\). Outcomes follow a Binomial distribution with parameters, \(\text{denom}_{ij}\) (=1 for binary outcomes) and \(\pi_{ij}\) (the probability that patient \(i\) seen by therapist \(j\) will drop-out.

\[
\text{logit}(\pi_{ij}) = \beta_0 + \beta_0\text{cons} + 0.030(0.002)\text{Age-gm} + 0.365(0.053)\text{EmployStatus}_1 + 0.259(0.106)\text{Ethnicity}_1 + 0.015(0.003)\text{CoreXRpre-gm} + 0.331(0.071)\text{coR06p}_1 + 0.170(0.063)\text{coR34p}_1
\]

\[
\beta_0 = -1.043(0.084) + u_{0j}
\]

\[
[u_{0j}] \sim N(0, \Omega) : \Omega_u = [0.483(0.094)]
\]

\[
\text{Var}(\text{Ending}_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}
\]

\[
\text{Deviance (MCMC)} = 12187.190(10521 of 10521 cases in use)
\]

The coefficients (SE) of included patient variables: \text{Age-gm} (age minus grand mean); \text{EmployStatus}_1 (unemployed); \text{Ethnicity}_1 (not white); \text{CoreXRpre-gm} (pre CORE score excluding risk, minus grand mean); \text{coR06p}_1 (affirmative response to CORE question 6, 'over the past week I have been physically violent to others'); \text{coR34p}_1 (affirmative response to CORE question 34, 'I have hurt myself physically or taken dangerous risks with my health').

\(\beta_0\) : Therapist intercept consisting of two terms, the fixed component (SE) = -1.043 (0.084) plus the therapist specific component, \(u_{0j}\).

The random effect \([u_{0j}]\) follows a Normal distribution (N) with a mean of 0 and a variance \((\Omega_u)\) of 0.483 (SE:0.094).

\[
\text{Var}(\text{Ending}_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}
\]

\(\text{Deviance (MCMC)}\): used to produce the Deviance Information Criteria, an indicator of model 'fit'.
Appendix B: Multilevel model for deterioration of more than one point on CORE-OM. Key: Deterioration$\geq 1_{ij}$: Outcome (deterioration by more than 1 point on CORE-OM or not) for patient $i$ seen by therapist $j$. Patient variables CoreXRpre-gm, EmployStatus_1 and Age-gm are described in Appendix 1, Cor09p_1 indicates an affirmative response to CORE question 9, ‘I have thought of hurting myself’. The model notation and other parameters are described in Appendix A.
Figure 1: Flowchart showing how the study samples were derived from the full data sample.

203x266mm (96 x 96 DPI)
Figure 2: Plot of therapist residuals (with 95% CI) for patient dropout.
226x128mm (132 x 132 DPI)
Figure 3: Plot of therapist residuals (with 90% CIs) for patient deterioration
270x172mm (96 x 96 DPI)
Figure 4: Scatterplot of therapist residuals for both dropout and deterioration
241x203mm (96 x 96 DPI)