
Article:
Ali, S., Rhodes, L., Moreea, O. et al. (6 more authors) (2017) How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. Behaviour Research and Therapy (94). pp. 1-8. ISSN 0005-7967

https://doi.org/10.1016/j.brat.2017.04.006

Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., Lucock, M., Lutz, W., Delgadillo, J. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy* **94**, 1–8.

doi: 10.1016/j.brat.2017.04.006

How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study

Shehzad Ali [a], Laura Rhodes [b], Omar Moreea [c], Dean McMillan [d], Simon Gilbody [d], Chris Leach [e], Mike Lucock [e], Wolfgang Lutz [f], and Jaime Delgadillo [g*]

a. Department of Health Sciences and Centre for Health Economics, University of York, York, UK
b. Leeds Community Healthcare NHS Trust, Leeds, UK
c. Centre for Clinical Practice, National Institute for Health and Care Excellence, Manchester, UK
d. Hull York Medical School and Department of Health Sciences, University of York, York, United Kingdom
e. South West Yorkshire Partnership NHS Foundation Trust and University of Huddersfield, Huddersfield, UK
f. Department of Psychology, University of Trier, Trier, Germany
g. Clinical Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK

* Correspondence: Dr Jaime Delgadillo, Clinical Psychology Unit, University of Sheffield, Sheffield, United Kingdom. Email: jaime.delgadillo@nhs.net

ABSTRACT

***Background*:** Depression and anxiety disorders are relapse-prone conditions, even after successful treatment with pharmacotherapy or psychotherapy. Cognitive behavioural therapy (CBT) is known to prevent relapse, but there is little evidence of the durability of remission after low intensity forms of CBT (LiCBT).

***Method*:** This study aimed to examine relapse rates 12 months after completing routinely-delivered LiCBT. A cohort of 439 LiCBT completers with remission of symptoms provided monthly depression (PHQ-9) and anxiety (GAD-7) measures during 12 months after treatment. Survival analysis was conducted to model time-to-relapse while controlling for patient characteristics.

***Results*:** Overall, 53% of cases relapsed within 1 year. Of these relapse events, the majority (79%) occurred within the first 6 months post-treatment. Cases reporting residual depression symptoms (PHQ-9 = 5 to 9) at the end of treatment had significantly higher risk of relapse (hazard ratio = 1.90, $p < 0.001$).

***Conclusions*:** The high rate of relapse after LiCBT highlights the need for relapse prevention, particularly for those with residual depression symptoms.

***Key words*:** depression; anxiety; relapse; cognitive behavioural therapy

1. Introduction

Depression is known to have a high recurrence rate, even after the successful treatment of acute-phase symptoms (Burcusa & Iacono, 2007; Yiend et al., 2009; Harter et al., 2007; Hardeveld et al., 2010; Gopinath et al., 2007). For example, after a first episode of depression, the probability of a further episode is approximately 50%; this rises to 70% following two episodes and 90% after a third episode (Burcusa & Iacono, 2007; Kessler et al., 1996). It also appears that with each further episode there is an increase in the severity of depressive symptoms and an increased probability that these symptoms will become resistant to treatment (Kendler et al., 2000). Similarly, research on anxiety disorders suggests high recurrence rates between 39% and 56% after treatment (Bruce et al., 2005; Eisen et al., 1999; Vervliet et al., 2013).

Literature in the field draws conceptual distinctions between *relapse* – a deterioration after initial response to treatment– and *recurrence* –a new episode of the disorder following a period of recovery– (Bockting et al., 2015). Meta-analyses of trials in this area show that accessing cognitive behavioural therapy (CBT) reduces the risk of depression relapse (Vittengl et al., 2007; Cuijpers et al., 2013) by comparison to acute-phase pharmacological treatment. The prophylactic effects of CBT appear to be as durable as to those of long-term maintenance on pharmacological treatment, but better at preventing relapse and recurrence compared to acute-phase pharmacological treatment without a maintenance phase (Hollon et al., 2005). Similarly, CBT is associated with sustained maintenance of improvements after the acute phase of treatment in various anxiety disorders (Otto, Smits, & Reese, 2005). It is, however, unclear if this

apparent durability of therapeutic effects primarily applies to conventional CBT delivered by qualified psychotherapists or psychologists for up to 20 sessions, as applied in efficacy trials.

Recent decades have seen the development of briefer and 'low intensity' versions of CBT (LiCBT) which can be delivered as guided self-help interventions supported by didactic materials (Bennett-Levy et al., 2010). LiCBT is becoming a common form of psychological care in many services, for example it is the predominant treatment option offered to thousands of patients each year in the *Improving Access to Psychological Therapies* (IAPT) programme in England (Clark, 2011). LiCBT involves brief (<8 sessions), highly structured (manual driven) psycho-educational interventions delivered in a variety of flexible formats (e.g., in person, via telephone, in groups, assisted by computerized learning modules). In the UK, for example, LiCBT is typically delivered by coaches (*psychological wellbeing practitioners*) who do not have formal psychotherapy or clinical psychology qualifications, but who are trained to a standardized curriculum and competency framework (e.g., see Richards & Whyte, 2009).

Although LiCBT can be effective at alleviating symptoms of depression and anxiety (Gellatly et al., 2007), it is as yet unclear if these effects are sustained after the acute-phase of treatment. For example, Coull et al. (2011) carried out a meta-analysis of 13 controlled trials comparing LiCBT versus waitlist or usual care controls, which estimated a statistically significant but small mean weighted effect size of d = 0.32 favouring LiCBT for 9 studies that reported follow-up data (up to 12 months), and this effect reduced to d = 0.19 after excluding a study with low quality rating. This finding is closely comparable to the statistically significant between-group effect size (d = 0.20) reported in a meta-analyses of computerized CBT for

depression that examined (up to 4 months) follow-up assessments (Richards & Richardson, 2012). However, this stands in contrast to a meta-analysis of internet-based CBT for depressive and anxiety disorders which concluded that clinical improvements were maintained at follow-up (median of 26 weeks) with no evidence of relapse (Andrews et al., 2010). The mixed evidence about its therapeutic durability raises important questions about the clinical and cost-effectiveness of LiCBT interventions, particularly when delivered in routine care conditions. Furthermore, evidence from meta-analyses is usually expressed in the form of effect sizes, masking information about the actual numbers (and proportions) of cases that may have relapsed in primary studies.

This study presents the findings of a naturalistic cohort study investigating remission and relapse rates following the completion of LiCBT interventions delivered in routine clinical care. The objectives of this study were to quantify post-treatment relapse rates at 12 months' follow-up and to explore predictors of time-to-relapse.

2. Method

2.1. Design and context

This was a prospective, longitudinal cohort study that included patients who completed LiCBT interventions in a primary care mental health service linked to the English IAPT programme. LiCBT interventions offered in the service followed national clinical guidelines (National Institute for Health and Clinical Excellence [NICE], 2011); they were highly standardised and delivered under regular (weekly or fortnightly) clinical supervision. These included individual and group guided self-help, as well as computerized CBT

interventions supported by psychological wellbeing practitioners. LiCBT interventions delivered in IAPT are based on principles of CBT, and aim to teach patients to apply coping skills including self-monitoring, goal setting, behavioural activation, graded exposure, problem solving, sleep hygiene, cognitive restructuring and relapse prevention (Bennett-Levy et al., 2010; Richards & Whyte, 2009).

Unlike controlled trials where the duration of treatment is often fixed *per protocol*, naturalistic studies typically show wide variability in the length of treatment, since patients attend a sufficient number of sessions to attain a *good enough level* (GEL) of improvement and then discontinue therapy (Barkham et al., 2006). This study aimed to assess the durability of LiCBT under routine care conditions, and therefore our definition of completion was consistent with the GEL model: completers attended at least 1 LiCBT therapy session after their initial assessment and had an end of treatment mutually agreed with their therapist (e.g. did not unilaterally drop out). Consistent with this definition, patients were eligible to take part in the study if they met 4 criteria: (1) they had case-level depression and/or anxiety symptoms at assessment; (2) they attended at least 1 LiCBT session; (3) they had a planned end of treatment; (4) they had below-threshold depression and anxiety symptoms, meeting IAPT criteria for recovery (Clark et al., 2009), as described in section 2.3 below. Those referred for ongoing psychological care (i.e., stepped-up to formal psychotherapeutic interventions) were excluded from the study. Eligible participants were identified from treatment discharge records and were invited to participate within 1 month of treatment completion. Consenting participants were contacted once per month using their preferred method (post, telephone, email) and were asked

to complete questionnaires (described below) to monitor symptoms of depression and anxiety.

Participants remained in the study until they met one of three end-point criteria: (1) their symptom scores were indicative of relapse as defined below; (2) they failed to respond to 2 consecutive monthly assessments, in which case they were considered lost to follow-up; or (3) their symptoms remained in remission until the 12-month follow-up assessment. Participants classified as relapsed were contacted by the research team; they received information about support options and were encouraged to re-engage with medical and/or psychological care. Service users were involved in the development of the study, and their views informed several aspects of the design (monthly frequency of assessments, end-point criteria and relapse support strategies) to ensure the study was feasible and acceptable. The study was approved by the NHS Health Research Authority and reviewed by an independent ethics committee (Yorkshire and Humber REC; Ref: 12/YH/0095).

2.2. Measures and data sources

Two validated patient reported outcome measures are routinely used in IAPT services to monitor depression and anxiety symptoms. The PHQ-9 is a nine-item screening tool for major depression (Kroenke et al., 2001). Each item is rated on a 0 to 3 scale, yielding a total depression severity score between 0-27. A cut-off ≥ 10 is used to detect clinically significant depression symptoms (Kroenke et al., 2001) and a reliable change index of ≥ 5 points has been recommended to monitor improvement or deterioration over time (McMillan et al., 2010). The GAD-7 is a seven-item measure developed to screen for anxiety disorders (Spitzer et al., 2006). Each item is

also rated on a 0 to 3 scale, yielding a total anxiety severity score between 0-21. A cut-off score ≥8 is recommended to identify the likely presence of a diagnosable anxiety disorder (Kroenke et al., 2007), with a reliable change index of ≥5 points (Richards & Borglin, 2011). Standard cut-off scores have been recommended for these measures to identify mild (≥5; residual or sub-threshold symptoms), moderate (≥10), moderately-severe (≥15) and severe (≥20) symptoms (Kroenke et al., 2001, 2007).

De-identified clinical assessment records were also collected for participants, including demographic (age, gender, ethnicity, employment, socioeconomic deprivation) and clinical information from the acute-phase of treatment (primary diagnosis, session-to-session outcome measures, family history of mental health problems, number of prior treatment episodes as a proxy for a relapsing condition). Furthermore, we created two variables using PHQ-9 and GAD-7 scores from the treatment phase: (1) Early change scores were estimated by subtracting symptom scores at session 3 from baseline (session 1) scores. (2) We used the last treatment session scores to derive a binary variable denoting the presence of residual symptoms (PHQ-9 = 5 to 9; GAD-7 = 5 to 7). Functional impairment was assessed using the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002). The WSAS measures the extent to which mental health problems impair daily functioning across 5 domains (work, home chores, social leisure, private leisure, and relationships). Responses are captured using 5 items rated on a nine-point scale ranging from "not at all" to "severely impaired". Socioeconomic deprivation was assessed by matching participants' home postcodes to the *English Index of Multiple Deprivation 2010* (IMD; Department for Communities and Local Government, 2011), and categorising cases into quintile levels of deprivation. The IMD is an area-level composite measure

which assigns a deprivation score to post code areas across England, taking into consideration 7 areas: income, employment, education level, health, crime, quality of housing and living environment.

2.3. Outcome definitions

Patients included in the study started treatment with case-level depression and/or anxiety symptoms and completed treatment with PHQ-9 and GAD-7 scores below the established diagnostic cut-offs; and thus met criteria for recovery as currently applied in IAPT services (NICE, 2011). To be classed as a relapse event, post-treatment symptom scores for at least one of the outcome measures were (1) above the diagnostic cut-off and were (2) ≥5 points greater than the symptom scores at the time of the last attended treatment session. This operational definition is consistent with Jacobson and Truax (1991) criteria for reliable and clinically significant deterioration. Participants who did not meet these criteria were classed as *in remission*. Thus, our primary outcome of interest was binary and coded as 0 = in remission and 1 = relapsed.

2.4. Statistical analyses

The recruitment process for the study was summarised following the STROBE guidelines (Von Elm et al., 2007) for observational cohort studies. Demographic and clinical characteristics of the included sample were compared to the wider pool of potentially eligible participants that were approached but who did not consent to take part in the study. These comparisons were based on Mann-Whitney U tests for continuous variables and chi-square tests for categorical variables.

We used survival analysis to assess the durability of LiCBT treatment effects over 12 months following treatment completion, and to determine predictors of relapse. Observed (unadjusted) time-to-relapse (in months) was assessed using non-parametric Kaplan-Meier (KM) curves (Kaplan & Meier, 1958). KM curves plot the probability of survival (remission) over time, while taking account of censored (i.e., missing) data points (N = 154 in this study). Hazard ratios predicting time-to-relapse were estimated with a semi-parametric Cox proportional hazards model (Cox, 1992). Cox regression does not make any assumptions about the shape of the underlying hazard function; however, it assumes that the hazard functions are proportional over time (i.e., there is a constant relative hazard). Based on this assumption, the estimated hazard ratio is multiplicatively related to the underlying hazard. We evaluated the proportional hazards assumption using a log-log plot [a plot of log(-log(survival) vs log(time)] and by fitting a smooth function of model residuals (Schoenfeld residuals) over time. Rejection of the null hypothesis implies deviation from the proportional hazards assumption. Additionally, we conducted parametric survival analysis using an exponential distribution. We also compared the results using the Weibull distribution which provided similar model estimates; hence, the exponential parametric results are reported.

For variable selection, we followed the approach discussed by Machin et al. (2006). First, the residual scores for the primary variables of interest (PHQ-9 and GAD-7) at the time of treatment completion were included in the model. Next, other potential variables of interest were entered in the model, including those that were statistically significant in univariate Cox regressions. These variables were age, gender, employment status, deprivation level (as measured by the IMD variable), number of prior

treatment episodes, early change scores and residual symptoms classification for PHQ-9 and GAD-7. The final model only included variables that were either statistically significant based on an alpha of $p \leq .10$, or were hypothesised *a priori* as potential prognostic factors irrespective of statistical significance (i.e., residual symptoms). Furthermore, a sensitivity analysis was conducted entering excluded variables.

The overall model fit was assessed using Cox-Snell residuals. If the model offers a good fit to the data, then the true cumulative hazard function (conditional on covariates) should have an exponential distribution with a hazard rate of 1 (Cleves et al., 2008). We also applied a specification link test which refits the model using the predicted value and its squared term to evaluate misspecification error (which would be confirmed by a statistically significant error term). The predictive value of the regression model was assessed using Harrell's *C*, which is the proportion of all usable subject pairs where the prediction and outcome are concordant (Harrell et al., 1982). A higher value of Harrell's *C* is desirable.

A series of sensitivity analyses were also conducted. First, we included the following variables which were initially excluded from the main analysis: baseline severity, change scores for the early phase of treatment, number of LiCBT treatment sessions, family history of mental health problems, and employment status. Secondly, we used parametric survival regression to evaluate the impact on regression coefficients. We also graphically evaluated how these distributions fit the observed data. Finally, we evaluated the impact of removing any outlier observations (identified using a deviance residual plot) on model estimates. All analyses were conducted in Stata v13.1.

3. Results

3.1. Sample characteristics

A total of 439 consenting participants were included in the follow-up phase of the study, out of a pool of 2,100 potentially eligible patients who were identified from treatment discharge records (STROBE diagram available in supplementary appendix 1).

The mean age in the sample was 41.28 (SD = 14.59; range = 17 to 82); 59.7% were females; 94.2% were from a White British background; and 33.7% were unemployed. Approximately 54.8% reported having a family history of mental health problems. The most common primary presenting problems recorded in clinical assessments were mixed anxiety and depression (31.9%), depressive episode (24.8%), generalised anxiety disorder (19.3%), panic disorder (4.7%) and recurrent depression (2%). Other conditions (i.e. agoraphobia, specific phobia, health anxiety) were much less common, and approximately 13.1% of presenting problems were unspecified. The mean number of prior treatment episodes was 1.02 (SD = 1.60; range 0 to 21; 42.3% had no prior treatment episodes).

The mean number of acute-phase treatment contacts including initial assessment was 7.04 (SD = 1.99; mode = 7; range = 2 to 16). Mean baseline severity measures at the time of initial assessments were PHQ-9 = 13.60 (SD = 5.41); GAD-7 = 13.20 (SD = 4.38). Mean scores at the last acute-phase treatment contacts were PHQ-9 = 3.44 (SD = 2.40); GAD-7 = 3.19 (SD = 2.17). At the end of treatment, 31.4% of cases had residual depression symptoms (PHQ-9 = 5 to 9) and 29.6% had residual anxiety symptoms (GAD-7 = 5 to 7).

The sample of study participants and the wider pool of potentially eligible participants were comparable in all demographic and clinical characteristics except for four variables. The study participants had a higher mean age (U[1,823] = 345,288.00, p = 0.001), a higher mean number of prior treatments (U[1,289] = 159,569.00, p < 0.02), a lower mean PHQ-9 score at the final treatment session (U[1,823] = 281,739.50, p = 0.03), and a higher mean number of treatment sessions (U[1,823] = 354,203.00, p < 0.001).

[Figure 1]

3.2. Time-to-event analysis

Overall, and after accounting for censored data points, 52.8% of cases were classified as relapse events within the 12-month follow-up period (47.2% remained in remission at 12 months). Figure 1 (Panel A) displays KM survival estimates, where the curve denotes the proportion of cases remaining in remission at each monthly measurement point (with actual numbers displayed in the table underneath). The majority of relapse events occurred within the first 6 months (i.e., of those who relapsed, 49% did so by month 2; 70% by month 4; 79% by month 6) and the rate of relapse decelerates thereafter. Panel B displays a smoothed estimated hazard function (with 95% confidence intervals), showing the risk of relapse over time.

[Tables 1, 2 and 3]

3.3. Predicting relapse

Table 1 shows univariate hazard ratios for variables tested as potential predictors of time-to-relapse. The only statistically significant predictors were final treatment session PHQ-9 and GAD-7 scores, as well as the presence of residual symptoms in these measures (all $p < 0.001$). Tables 2 and 3 show the results of the Cox proportional hazards model and the parametric regression model predicting time-to-relapse using multiple variables that were endorsed in the preliminary univariate analyses. Both models show that participants with residual depression symptoms (PHQ-9 = 5 to 9; but not residual GAD-7 symptoms) at the last treatment session were approximately twice as likely (hazard ratios = 1.9) to relapse compared to those with minimal symptoms (PHQ-9 ≤ 4); $p < 0.001$. All sensitivity analyses yielded the same results and no additional variables predicted time-to-relapse (baseline severity, early response to treatment, number of LiCBT treatment contacts, family history of mental health problems, unemployment; all $p > 0.05$).

[Figure 2]

Figure 2 shows adjusted survival functions based on Cox regression (panel A) and parametric regression models (panel B). Participants who finished treatment with residual depression symptoms were at greater risk of relapse (approximately 80% probability) and tended to have a shorter duration of remission compared to those with minimal symptoms.

4. Discussion

4.1. Main findings

To our knowledge, this is the first investigation of the durability of treatment effects for low intensity psychological interventions delivered in routine stepped care practice. The main findings can be summarised in 3 points. (1) One in two patients who attained remission of symptoms after acute-phase treatment experienced a clinically significant deterioration within 12 months of completing treatment. (2) Around eight out of ten relapse events occurred within the first 6 months post-treatment. (3) Patients who reported residual depression symptoms at the end of treatment were at higher risk of relapse (80% probability) and tended to deteriorate sooner. Residual anxiety symptoms were not found to be predictive of relapse after controlling for residual depression. These observations about the effectiveness of LiCBT beyond the acute-phase of treatment are particularly sobering when we consider that all our study participants had an agreed end of treatment, with sub-clinical symptoms. Previous studies have demonstrated that the length of treatment and the presence of early symptomatic gains are predictive of remission of acute-phase symptoms in LiCBT (e.g., see Delgadillo et al., 2014), but no such association was evident with longer-term outcomes.

4.2. Strengths and limitations

Much of what is known about relapse after psychological care is derived from controlled trials, with samples typically under 100 patients (Vittengl et al., 2007). In contrast, this was a large (N > 400) and adequately powered prospective cohort study which included a sample of patients

accessing the variety of LiCBT interventions typically offered in IAPT services (Clark, 2011; Richards & Borglin, 2011). Compared to most clinical trials of psychological interventions, which undertake post-treatment follow-up at 6 or 12 months (Coull & Morris, 2011), the study design enabled us to undertake a month-by-month quantification of remission and relapse rates. This approach provided information on time-to-relapse and enabled us to learn that around 70% of relapse events are detected within the first 4 months following treatment.

There are some methodological issues to consider when interpreting these findings. We did not undertake structured diagnostic interviews to determine if participants met criteria for a common mental disorder. This is an important limitation, since it is possible that some cases that we classified as having relapsed may not meet full diagnostic criteria. The PHQ-9 and GAD-7 are well established case-finding tools for mental health problems, with upward of 80% sensitivity and specificity in primary care populations (Spitzer et al., 2006; Kroenke et al., 2007). Although it is possible that a small number of 'false positives' may have been classed as relapse events due to our reliance on self-reported PHQ-9 and GAD-7 scores, we applied a reliable change index to detect cases with a clinically significant increase in symptoms that is less likely to be explained by measurement error or chance (Jacobson & Truax, 1991). Adopting this conservative and pragmatic approach, we were able to apply a robust way to detect clinically important deterioration in a way that was intensive (monthly contacts), flexible (via post, telephone, email) and acceptable to study participants. Diagnostic interviews are seldom applied in routine practice, and thus we argue that our method offers a *proof of principle* that monthly post-treatment follow-up is feasible to embed into a routine clinical system, albeit with

expected loss to follow-up as is common in observational studies. Another limitation is that the study recruited participants from a single service. Furthermore, the study participants differed from the wider pool of potentially eligible treatment completers in some characteristics, which meant they tended to have longer treatments and marginally better depression outcomes than those in the reference sample. This is an important observation that suggests that our estimation of relapse rates may be more conservative than the rates that could be expected in the wider population of LiCBT completers (who tend to have higher levels of residual depression symptoms). Although future replication in other locations is advisable, we note that our findings are consistent with results from meta-analyses of controlled trials that assessed the longer-term durability of LiCBT treatment effects (Coull & Morris, 2011; Richards & Richardson, 2012).

As expected, some participants in our study were lost to follow-up. Survival analysis assumes *non-informative censoring*, which implies that the pattern of missing data is unrelated to the outcome of interest. Hence, survival analysis uses all available data (including that from participants who were later lost to follow-up) by partitioning the follow-up period into sub-intervals (i.e., months) at which events (i.e., relapse) or censoring occur. Subsequently, all available data within each interval is included in the analysis. It is of course possible that some cases were lost to follow-up due to relapse, and therefore our estimates of relapse are likely to be conservative.

4.3. Considerations for policy and practice

Over a decade ago, Tylee and Jones (2005) recognised how the *rule of halves* applies in depression treatment: "*only half of depressed patients seek help from doctors, half are detected in primary care, half receive treatment with only half completing it*". This aphorism seems to apply in low intensity psychological care as much as in pharmacological treatment. Richards and Borglin (2011), for example, observed that around a quarter of patients screened as suitable for IAPT treatment failed to attend their initial therapy appointment and another quarter dropped out after starting therapy. Less than half of those who start LiCBT attain reliable and clinically significant improvement (Delgadillo et al., 2014; Green et al., 2014; Firth et al., 2015). Furthermore, approximately half of treatment completers classed as recovered are likely to relapse within a year of discharge from care. Overall, taking an optimistic assumption that up to 43% of those who access LiCBT attain reliable and clinically significant improvement (Delgadillo et al., 2014), only 20% of all LiCBT cases are likely to remain in remission 12 months after treatment.

It seems hasty to consider patients 'recovered' at the point of discharge without assessing full remission of symptoms over a longer period. The wider literature in this area suggests that recovery can only be considered following an extended period (i.e., 6 to 12 months) of full remission of symptoms (Bockting et al., 2015). Current mental health policy in the UK emphasises the notion of recovery (Department of Health, 2014), reinforcing a short-term view about common mental health problems in primary care. Our findings align with the extant literature (Burcusa & Iacono, 2007; Yiend et al., 2009; Harter et al., 2007; Hardeveld et al., 2010; Gopinath et al., 2007) and suggest taking the long view, recognising that

problems like depression often have to be managed like recurrent long-term conditions.

In this regard, we would recommend two areas for future practice developments. First, the influence of residual depression symptoms on the risk of relapse is well documented (Vittengl et al., 2007; Paykel et al., 1995), and it is known that around 76% of such cases relapse within 10 months of treatment completion (Paykel, 2008). Our results show a very similar pattern for this high risk group. Patients accessing LiCBT who have shown initial response to treatment should ideally continue to access care until attaining a full remission of symptoms (scores < 5 on PHQ-9, GAD-7). It could be argued that many (31%) of the relapse cases in this study actually had a partial (rather than full) remission at the time of treatment completion, and thus never actually 'recovered'. Second, patients should have access to active and responsive follow-up by a dedicated case manager (Von Korff & Goldberg, 2001). Post-treatment follow-up appointments could be offered as less frequent booster sessions (Gearing et al., 2013) or less intensive self-management support, particularly during the critical period of 4 to 6 months after acute-phase treatment. More broadly, there are a number of evidence-based relapse prevention options that could form part of the landscape of primary care interventions. For example, continuation-phase CBT has been shown to reduce the risk of relapse by up to 29% at 12 months' follow-up compared to assessment only (Vittengl et al., 2007). Mindfulness based interventions have also been associated with a relapse risk reduction of up to 43% for patients with 3 to 4 previous episodes of depression (Piet & Hougaard, 2011). It is clear that relapse prevention is an overlooked aspect of routine stepped care practice in IAPT services and an important area for further policy and research developments.

Acknowledgements

The *West Yorkshire Low-intensity Outcomes Watch* (WYLOW) study was a longitudinal cohort project conducted between 2010 and 2016. It was funded by a Feasibility and Sustainability Funding (FSF) grant awarded by Leeds Community Healthcare NHS Trust, United Kingdom; REF: FSF2010/11/08.11.2010.

We thank Alexander Teahan, Emma Ferguson and Rachel Bos, who supported data collection and preparation for this study. Thanks also to patients who gave up their time to advise the research team during the development stage, and to all participants who agreed to maintain contact with the research team after completing an episode of treatment.

References

Andrews G, Cuijpers P, Craske MG, McEvoy P, & Titov N (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis and pilot implementation. *PLoS ONE* 5, e13196.

Barkham M, Connell J, Stiles WB, Miles JN, Margison F, Evans C, & Mellor-Clark J (2006). Dose-effect relations and responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology* 74, 160–167.

Bennett-Levy J, Richards DA, Farrand P (2010). Low intensity CBT interventions: a revolution in mental health care. In *Oxford guide to low intensity CBT interventions*. Oxford (e.d. Bennett-Levy J, Richards DA, Farrand P, Christensen H, Griffiths KM, Kavanagh DJ, Klein B, Lau MA, Proudfoot J, Ritterband L, White J and Williams C), pp. 3-18. Oxford University Press: Oxford.

Bockting CL, Hollon SD, Jarrett RB, Kuyken W, Dobson K (2015). A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clinical Psychology Review* 41, 16-26.

Bruce SE, Yonkers KA, Otto MW, Eisen JL, Weisberg RB, Pagano M, Shea MT, Keller MB (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. *American Journal of Psychiatry* 162, 1179–87.

Burcusa SL, Iacono WG (2007). Risk for recurrence in depression. *Clinical Psychology Review* 27, 959–985.

Clark DM (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry* 23, 318–327.

Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy* 47, 910-920.

Cleves M, Gould WW, Gutierrez RG, and Marchenko Y (2008). *An introduction to survival analysis using Stata*. Stata Press.

Coull G, Morris PG (2011). The clinical effectiveness of CBT-based guided self-help interventions for anxiety and depressive disorders: a systematic review. *Psychological Medicine* 41, 2239–2252.

Cox DR (1992). *Regression models and life-tables. Breakthroughs in Statistics*, pp. 527-41. Springer: New York.

Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS (2013). A Meta-Analysis of Cognitive-Behavioural Therapy for Adult Depression, Alone and in Comparison With Other Treatments. *Canadian Journal of Psychiatry* 58, 376–385.

Delgadillo J, McMillan D, Lucock M, Leach C, Ali S, Gilbody S (2014). Early changes, attrition and dose-response in low intensity psychological interventions. *British Journal of Clinical Psychology* 53, 114–130.

Department for Communities and Local Government (2011). *The English Indices of Deprivation 2010*. Author: London.

Department of Health (2014). *Closing the gap: priorities for essential change in mental health*. DH: London. www.gov.uk/government/uploads/system/uploads/attachment_data/file/281250/Closing_the_gap_V2_-_17_Feb_2014.pdf (accessed March 2015).

Eisen JL, Goodman WK, Keller MB, Warshaw MG, DeMarco LM, Luce DD, Rasmussen SA (1999). Patterns of remission and relapse in obsessive-compulsive disorder: a 2-year prospective study. *Journal of Clinical Psychiatry* 60, 346–51.

Firth N, Barkham M, Kellett S, Saxon D (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: a multilevel modelling analysis. *Behaviour Research and Therapy* 69, 54-62.

Gearing RE, Schwalbe CSJ, Lee RH, Hoagwood KE (2013). The effectiveness of booster sessions in CBT treatment for child and adolescent mood and anxiety disorders. *Depression and Anxiety* 30, 800-808.

Gellatly J, Bower P, Hennessy S, Richards D, Gilbody S, Lovell K (2007). What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. *Psychological Medicine* 37, 1217–1228.

Green H, Barkham M, Kellett S, Saxon D (2014). Therapist effects and IAPT psychological wellbeing practitioners (PWPs): a multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy* 63, 43-54.

Gopinath S, Katon WJ, Russo JE, Ludman EJ (2007). Clinical factors associated with relapse in primary care patients with chronic or recurrent depression. *Journal of Affective Disorders* 101, 57–63.

Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman AT (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica* 122, 184–191.

Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association* 247, 2543–6.

Harter M, Baumeister H, Reuter K, Jacobi F, Höfler M, Bengel J, Wittchen H-U (2007). Increased 12-month prevalence rates of mental disorders in patients with chronic somatic diseases. *Psychotherapy and Psychosomatics* 76, 354–360.

Hollon SD, DeRubeis RJ, Shelton RC, Amsterdam JD, Salomon RM, O'Reardon JP, Lovett ML, Young PR, Haman KL, Freeman BB, Gallop R (2005). Prevention of Relapse Following Cognitive Therapy vs Medications in Moderate to Severe Depression. *Archives of General Psychiatry* 62, 417–422.

Jacobson NS, Truax P (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 59, 12–19.

Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457–81.

Kendler KS, Thornton LM, Gardner CO (2000). Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the 'kindling' hypothesis. *American Journal of Psychiatry* 157, 1243–1251.

Kessler RC, Nelson CB, McGonagle KA, Liu J, Swartz M, Blazer DG (1996). Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. *British Journal of Psychiatry (Supplement)* 30, 17–30.

Kroenke K, Spitzer RL, Williams JB (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine* 16, 606-613.

Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine* 146, 317–325.

Machin D, Cheung YB, Parmar M (2006). *Survival analysis: a practical approach*. John Wiley & Sons: West Sussex.

McMillan D, Gilbody S, Richards DA (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders* 127, 122-129.

Mundt JC, Mark IM, Shear MK, Griest JM (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *British Journal of Psychiatry* 180, 461-464.

National Institute for Health and Clinical Excellence (2011). *Common Mental Health Disorders: identification and pathways to care*. National Collaborating Centre for Mental Health: London.

Otto MW, Smits JA, & Reese HE (2005). Combined psychotherapy and pharmacotherapy for mood and anxiety disorders in adults: review and analysis. *Clinical Psychology: Science and Practice* 12, 72-86.

Paykel ES (2008). Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience* 10, 431−437.

Paykel ES, Ramana R, Cooper Z, Hayhurst H, Kerr J, Barocka A (1995). Residual symptoms after partial remission: an important outcome in depression. *Psychological Medicine* 25, 1171–1180.

Piet J, Hougaard E (2011). The effect of mindfulness-based cognitive therapy for prevention of relapse in recurrent major depressive disorder: A systematic review and meta-analysis. *Clinical Psychology Review* 31, 1032–1040.

Richards DA, Borglin G (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders* 133, 51–60.

Richards D, Richardson T (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, 32, 329–342.

Richards DA, & Whyte M (2009). *Reach out: National programme educator materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions (2nd ed.)*. London, UK: Rethink.

Spitzer RL, Kroenke R, Williams JB, Lowe B (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine* 166, 1092-1097.

Tylee A, Jones R (2005). Managing depression in primary care. *British Medical Journal* 330, 800–1.

Vervliet B, Craske MG, Hermans D (2013). Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology* 9, 215-248.

Vittengl JR, Clark LA, Dunn TW, Jarrett RB (2007). Reducing relapse and recurrence in unipolar depression: a comparative meta-analysis of cognitive-behavioral therapy's effects. *Journal of Consulting and Clinical Psychology* 75, 475–488.

Von Elm E, Altman D, Egger M, Pocock SJ, Gotzsche P, Vandenbroucke JP (2007). Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 335, 806–8.

Von Korff M, Goldberg D (2001). Improving outcomes in depression: The whole process of care needs to be enhanced. *British Medical Journal* 323, 948–949.

Yiend J, Paykel E, Merritt R, Lester K, Doll H, Burns T (2009). Long-term outcome of primary care depression. *Journal of Affective Disorders* 118, 79–86.

Table 1. Univariate hazard ratios for potential predictors of time-to-relapse

Variables	Sample estimates*	Univariate hazard ratio (95% CI)	*p*
Continuous variables	mean (SD)		
Age (years)	41.28 (14.59)	0.99 (0.98 to 1.00)	0.195
PHQ-9 at final treatment session	3.44 (2.40)	1.16 (1.10 to 1.23)	< 0.001
GAD-7 at final treatment session	3.19 (2.17)	1.17 (1.10 to 1.23)	< 0.001
Previous treatment episodes	1.02 (1.60)	1.03 (0.95 to 1.12)	0.48
Categorical variables	N (%)		
Gender = female	262 (59.7)	0.94 (0.70 to 1.25)	0.650
IMD quintile			
Q1 (most affluent)	94 (21.6)	(reference category)	
Q2	94 (21.6)	0.82 (0.54 to 1.24)	0.34
Q3	102 (23.4)	0.66 (0.43 to 1.01)	0.06
Q4	88 (20.2)	0.63 (0.40 to 0.97)	0.04
Q5 (most deprived)	58 (13.3)	0.72 (0.44 to 1.15)	0.17
Residual symptoms at final treatment session			
PHQ-9 = 5 to 9	138 (31.4)	2.08 (1.56 to 2.77)	< 0.001
GAD-7 = 5 to 7	130 (29.6)	1.56 (1.83 to 2.13)	< 0.001
Early response during therapy**	181 (43.8)	1.21 (0.89 to 1.65)	0.22
Unemployed at initial assessment	148 (33.7)	0.98 (0.70 to 1.36)	0.89

*Based on chi-squared test for binary variables and *t*-test for continuous variables; ** based on reduction of symptoms greater than reliable change index between sessions 1 – 3; IMD = index of multiple deprivation; CI = confidence intervals

Table 2. Cox proportional hazards model predicting time-to-relapse

Variables	Hazard Ratio	SE	*z*	*p*	[95% CI]
Gender = female	0.888	0.131	-0.800	0.423	0.665 to 1.187
Age (years)	0.995	0.005	-1.000	0.318	0.985 to 1.005
PHQ-9 residual symptoms*	1.900	0.327	3.730	< 0.001	1.356 to 2.664
GAD-7 residual symptoms*	1.157	0.204	0.830	0.407	0.820 to 1.634
IMD quintile (reference: Q1)					
Q2	0.857	0.183	-0.720	0.470	0.565 to 1.302
Q3	0.750	0.164	-1.320	0.187	0.489 to 1.150
Q4	0.691	0.155	-1.650	0.100	0.445 to 1.073
Q5 (most deprived)	0.802	0.198	-0.900	0.370	0.494 to 1.300

* residual symptoms at last treatment session (PHQ-9 ≥ 5; GAD-7 ≥ 5); SE = standard error; CI = confidence intervals; IMD = index of multiple deprivation

Table 3. Parametric regression model predicting time-to-relapse (exponential distribution)

Variables	Hazard Ratio	SE	*z*	*p*	[95% CI]
Gender = female	0.869	0.129	-0.940	0.347	0.649 to 1.164
Age (years)	0.991	0.005	-1.870	0.062	0.981 to 1.000
PHQ-9 residual symptoms*	2.149	0.372	4.430	< 0.001	1.532 to 3.017
GAD-7 residual symptoms*	1.193	0.212	0.990	0.322	0.842 to 1.690
IMD quintile (reference: Q1)					
Q2	0.854	0.182	-0.740	0.460	0.563 to 1.297
Q3	0.711	0.155	-1.560	0.119	0.463 to 1.091
Q4	0.636	0.143	-2.020	0.044	0.410 to 0.988
Q5	0.741	0.182	-1.220	0.223	0.457 to 1.201
Constant	0.124	0.034	-7.510	< 0.001	0.072 to 0.213

* residual symptoms at last treatment session (PHQ-9 ≥ 5; GAD-7 ≥ 5); SE = standard error; CI = confidence intervals; IMD = index of multiple deprivation

Figure 1. Survival analysis of remission and time-to-relapse following low intensity CBT interventions

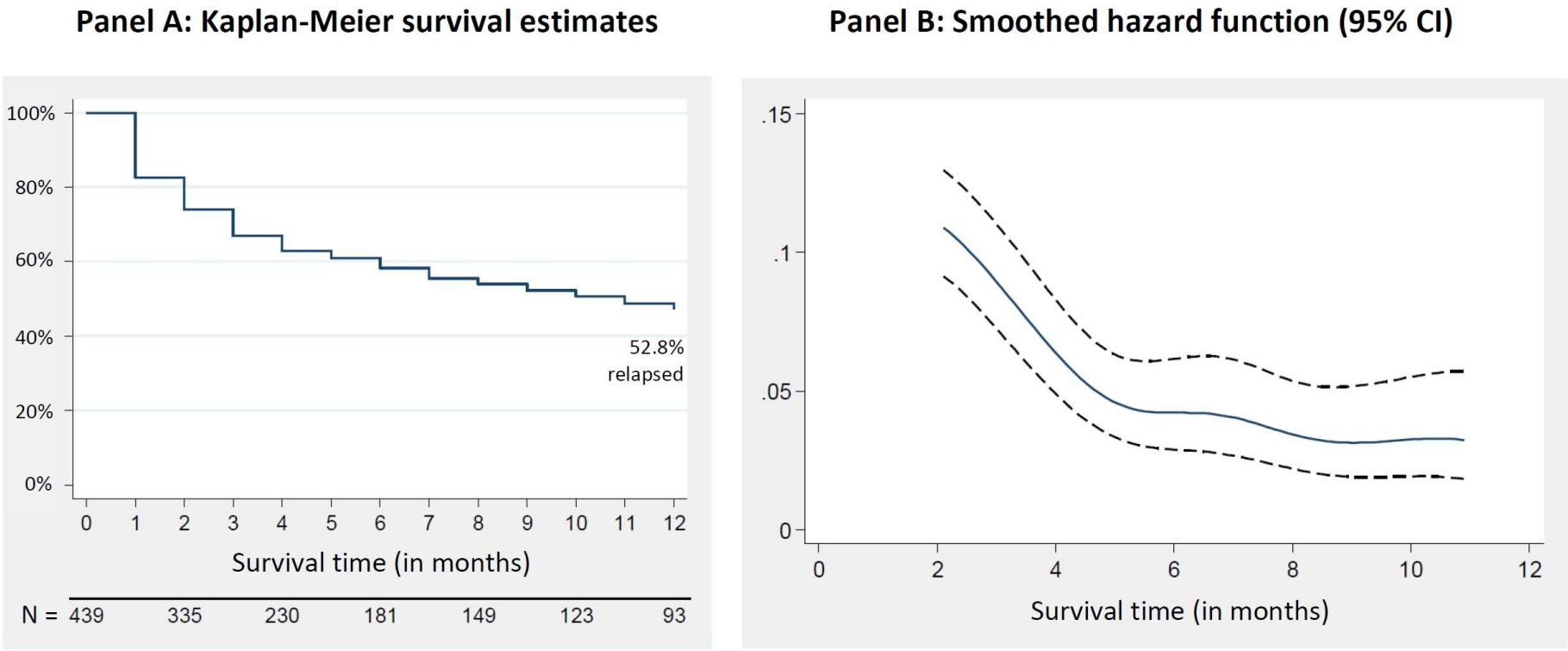

Figure 2. Adjusted survival functions for patients with and without residual depression (PHQ-9) symptoms