In vivo processes in cognitive therapy for depression: Frequency and benefits

JONATHAN W. KANTER1, JONATHAN S. SCHILDCROUT2, & ROBERT J. KOHLENBERG3

1Department of Psychology, University of Wisconsin—Milwaukee, 2Department of Biostatistics, Vanderbilt University, and 3Department of Psychology, University of Washington

(Received 26 December 2003; revised 1 March 2005; accepted 3 March 2005)

Abstract
Several recent variants of cognitive therapy (CT) for depression attempt to improve treatment with an increased focus on the occurrence of problems in the context of the therapy setting, the therapeutic process, and the therapeutic relationship. These processes are referred to as in vivo work. However, it is possible that cognitive therapists already incorporate in vivo work. One study compared standard CT to a version that emphasized in vivo work according to the guidelines outlined in functional analytic psychotherapy (FAP). The resulting treatment was called FAP-enhanced cognitive therapy (FECT). Using a turn-by-turn process-coding method, the authors examined therapist behavior in every session of this study to determine the frequency of in vivo work. Results indicated that cognitive therapists only occasionally engaged in such work, but rates increased in FECT and related to several weekly client-reported variables, including progress in treatment but not depressive symptoms.

The efficacy of cognitive therapy (CT) for depression (Beck, Rush, Shaw, & Emery, 1979) is well established (Dobson, 1989). However, it can and needs to be improved, because CT does not help all who receive it. For example, in two large-scale studies supporting the efficacy of CT for depression (Elkin et al., 1989; Hollon et al., 1992), of those clients who entered treatment, only 64% to 68% completed treatment and only 50% to 51% of those who completed treatment were classified by clinical interviewers as recovered.

One avenue for improvement is an increased focus on the occurrence of problems in the context of the therapy setting, the therapeutic process, and the therapeutic relationship. We refer to this focus as in vivo work. For example, in two large-scale studies supporting the efficacy of CT for depression (Elkin et al., 1989; Hollon et al., 1992), of those clients who entered treatment, only 64% to 68% completed treatment and only 50% to 51% of those who completed treatment were classified by clinical interviewers as recovered.

One study (Kohlenberg et al., 2002) focused on enhancing the effectiveness of CT by emphasizing in vivo work as per the guidelines outlined in functional analytic psychotherapy (FAP; Kohlenberg & Tsai, 1991). The resulting treatment was called FAP-enhanced cognitive therapy (FECT; Kohlenberg et al., 2002, 2004). In FECT, therapists are encouraged to pay particular attention to in vivo possibilities.

In the FECT study (Kohlenberg et al., 2002), which was an open trial, both CT and FECT performed well (79% of FECT clients and 60% of CT clients responded to treatment). In addition, as predicted by FECT's increased in vivo focus, FECT clients showed large improvements on measures of interpersonal functioning, whereas CT clients showed no improvements. An adherence analysis showed that CT rarely incorporated in vivo work, but FECT incorporated in vivo work more often. In addition, a specific type of in vivo work, the application of cognitive modification strategies to dysfunctional thoughts about the therapeutic relationship, predicted improved client depression and interpersonal functioning outcomes.

The FECT study suggests that increasing in vivo work in CT may be an avenue for improvement, but it is possible that FECT is nothing new. Cognitive therapists often claim that they already do in vivo work (e.g., Lazarus, 2003), and the CT manual (Beck et al., 1979) notes the importance of addressing transference reactions and provides several
examples of CT techniques specifically focused on the therapy or the therapeutic relationship. However, using a turn-by-turn coding system, Goldfried et al. (Castonguay, Hayes, Goldfried, & DeRubeis, 1995; Goldfried, Castonguay, Hayes, Drozd, & Shapiro, 1997; Goldfried, Raue, & Castonguay, 1998) showed that research CT therapists do not typically focus on in vivo work, whereas master clinicians do during significant, high-impact sessions.

In the current study, we coded all therapist turns at speech in the FECT study (Kohlenberg et al., 2002) to determine the percentage of turns with an in vivo focus. As noted, adherence ratings of CT during the FECT study suggested that cognitive therapists did not engage in this focus. However, there are several important limitations to the use of those data to answer this question. First, adherence scales ask raters to assess the form rather than the function of therapist behaviors, so many therapeutic behaviors that are functionally similar to the behavior described by the scale item often will not be identified. Similarly, it is possible that a major in vivo session focus may only marginally elevate a specific scale if the therapist did not regularly use the technical vocabulary that allows raters to correctly identify the behavior of interest. In addition, because adherence scales are very labor intensive, only a fraction of the full therapy course was rated, as is typical of most outcome studies. Thus, much therapist behavior is never rated, so selective ratings may miss important therapist activity that occurred in other sessions. For example, research on sudden gains in psychotherapy suggests that it may be a single session or two out of a full course of treatment responsible for the majority of symptomatic improvement (Tang & DeRubeis, 1999). For these reasons, it is possible that the adherence data reported by Kohlenberg et al. underestimated the true use of in vivo work by the cognitive therapists in that study.

We present an alternative turns-at-speech method for assessing in vivo work that addresses these concerns. First, ratings of almost every therapist utterance in every session of the FECT study (missing data are described later) ensured that no major therapist activity was missed. Second, raters rated the content and context of each utterance, so ratings were not dependent on the therapist’s use of a specific technical vocabulary or formally specified item descriptions. Third, raters were instructed in broad definitions that maximized the possibility that appropriate therapist behavior would be coded as in vivo focused. This method resulted in an exhaustive, simple, frequency analysis. The primary purpose of this analysis was to describe the rates of in vivo (IV) focused turns in CT and FECT. Post hoc analyses explored whether increased IV predicted several weekly client-reported process variables, including reports on progress in therapy, progress in relationships, and depression.

Because we assessed the frequency of IV in every session of the study, our statistical approach took advantage of repeated measurements over time. Specifically, we used generalized estimating equations (Liang & Zeger, 1986) assuming an independence working covariance matrix, in which valid measures of uncertainty are captured with robust standard errors. This statistical approach, used more commonly for biostatistical and epidemiological research, allowed us to compute odds ratios (ORs) that estimate the relationship between rates of IV and condition (FECT vs. CT) in addition to testing for differences in IV between conditions. Additionally, for the post hoc analyses, our statistical approach allowed us to model the increased likelihood that clients would report the occurrence of specific process variables given a specified change in the rate of IV.

Method

The FECT study

A detailed description of the exclusion and inclusion criteria, client demographic characteristics, and therapeutic procedures of the FECT study may be found in Kohlenberg et al. (2002). In brief, four experienced cognitive therapists provided therapy for both CT and FECT. Each therapist had been in practice for at least ten years and had served as a CT therapist on prior clinical trials. Three of the therapists attended a CT workshop presented by a national expert before the study, and two were board certified by the Academy of Cognitive Therapy. Therapist competency was rated during the study by Dr. Keith Dobson using the Cognitive Therapy Scale (CTS; Vallis, Shaw, & Dobson, 1986). There were no significant differences in CTS total scores between therapists or between conditions, nor was there a significant Therapist × Condition interaction. CTS scores ($M = 43.58, SD = 6.00$) were considered adequate and comparable to other studies (Shaw, 1984). Client demographic characteristics are presented in Table I.

The study design first had the therapists conduct standard CT to the best of their ability with four clients each before learning and conducting FECT with six clients each. Because it was a treatment development study, an A-B design was used to safeguard against possible treatment dispersion. The study included more FECT than CT clients to better study training effects over time in the
FECT condition. No proscriptions on in vivo work were made during the CT phase of the study. All clients met inclusion and exclusion criteria typical of randomized clinical trials for depression (e.g., Jacobson et al., 1996). The current analyses were based on the 15 CT and 23 FECT clients who completed the study. There were no significant differences between conditions in the number of clients who dropped out of the study.

Rating procedure

Coding of sessions. Individual raters rated audiotapes of whole therapy sessions with a coding scheme designed for this project. Almost every therapy session for every client in the FECT study was rated (missing data reported later). Sessions were distributed to raters at random from the data set. Each therapist turn at speech was categorized into one of two mutually exclusive categories: IV and “other.” IV was defined as talk aimed at working on client problems that occurred in therapy in relation to the therapy process, the therapy relationship, or anything else having to do with therapy, and “other” simply was defined as everything other than IV. To maximize identification of IV turns, raters were instructed to rate a turn as IV if they identified both IV and other material in a single turn. Raters were instructed to base categorizations for each turn both on the content and context of what was said. In other words, a specific turn would be rated as IV if it did not specifically refer to IV material but was part of a larger discussion that was in vivo focused. Examples of IV therapist turns include “I’ve noticed that you don’t look at me when we are discussing sensitive issues,” “What’s so important about whether I like you or not?”, and “Do you feel responsible for coming up with something to say every time we set the agenda?”

Raters. Sixteen undergraduate research assistants who received approximately 9 hr of training over the course of approximately two weeks served as raters. Training included didactic instruction, practice on how to reliably distinguish turns, and practice on how to reliably distinguish IV from other content within each turn. After three months of rating, raters received an additional 3 hr of booster training. Raters were unaware of the overall purpose of the study, and every attempt was made to keep the raters unaware of the treatment condition of the session.

Reliability. Two advanced research assistants were trained as criterion raters. The raters were trained as adherence raters for the FECT study, participated in the development of the current coding scheme and in the rater training sessions, and received approximately 15 hr of additional rating practice with the graduate student in charge of the project. Fifty-nine sessions were randomly selected from the total sample and rated by a criterion rater. Reliability was assessed by first summing the number of IV turns for each session rated for both the criterion rating and the data rating. A proportion IV score (total number of IV turns/total number of overall turns) was then calculated for each session. Intra-class correlation coefficient (ICC2,1) (Shrout & Fleiss, 1979) was calculated between the data rater’s proportion IV scores and those of the criterion raters, and results were very reliable (>.97).

Weekly client-reported measures

As part of the FECT study, clients reported on therapy session content and therapeutic progress on
a questionnaire after weekly sessions. Five items were selected from this larger assessment to capture separate key features of FECT. The first three items pertained to therapy session content and the final two to therapeutic progress and interpersonal relationships.

Therapy session content was measured by client reports of whether the following were discussed during the session: (a) inadequacies, fears, or successes in getting along personally and socially, (b) feelings about being close to or needing someone, and (c) therapy, the therapist, being a patient. These items were taken from a 19-item checklist that is a component of the Patient Therapist Session Report (Orlinsky & Howard, 1986). These items were assessed simply to establish validity for the coding scheme. Because in FECT the therapist is encouraged to emphasize interpersonal problems in the therapeutic relationship and relate these problems to other relationships, we believed that inadequacy, closeness, and therapy each would be positively related to in-session IV rates, with therapy being most strongly related.

To capture perceptions of therapeutic progress, clients were asked, "Do you feel that you made progress in this session in dealing with the problems for which you are in therapy?" When clients reported at least some progress during the session (considerable progress or some progress), then weekly progress was made, and if not (I really didn't get anywhere or in some ways my problems got worse), progress was not made.

Interpersonal relationship improvement perceptions were assessed with the item, "Were your relationships over the last week different than usual?" If a client reported that relationships were at least slightly better during the week (slightly better or much better), then relationship improvements had been made, and if not (I haven't noticed any changes, slightly worse, or much worse), relationship improvements had not been made.

On the basis of the preceding item, clients were asked, "If your relationships were different this week, is this difference due to therapy?" If clients reported that relationship differences were at least probably due to therapy (probably due to therapy or definitely due to therapy), then the relationship changes were attributed to therapy; if not (probably due to other factors or definitely due to other factors), then relationship changes were not attributed to therapy. We believed that clients who experienced positive changes in relationships would be more likely to attribute the changes to therapy than those who experienced negative changes.

Finally, clients also completed the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) as a weekly measure of depressive symptoms.

Results

Rates of IV turns in CT and FECT

Table I summarizes IV rates. IV-focused turns at speech per session were higher in FECT (M = 9.2, SD = 9.9) than in CT (M = 3.4, SD = 4.9), whereas total turns were approximately equal (79.3 and 80.1, respectively). Few (3.7%) CT sessions had more than 20% IV turns at speech, and none had more than 50%. In contrast, 21.0% of FECT sessions had more than 20% IV turns at speech and nine (2.1%) had more than 50%.

Log-linear (Poisson) regression analysis was used to establish the relative rate of IV turns in FECT compared with CT (e.g., rate of IV turns at speech in FECT divided by the rate of IV turns at speech in CT). Total number of turns per session varied from session to session, so we used log of total turns per session as an offset. In this and all regression models discussed here, we adjusted for therapist and client gender. The rate of IV turns was 2.8 times higher in FECT compared with CT (95% CI = 2.2–3.6, p < .001).

Relationship of IV turns to weekly client reports

We examined the degree to which therapy focus was related to weekly client reports on therapy session content (inadequacy, closeness, and therapy, as described previously). Clients reported such discussions in 58.6%, 31.0%, and 46.2% of sessions, respectively (Table II). Because these weekly measures were binary (dichotomous), we used logistic regression to estimate the relationships between weekly client reports and number of IV turns in the session in terms of odds ratios. The odds ratio represents the relative odds of responses one unit apart on the dependent measure given a specified change in the independent measure, and because the changes in odds is assumed to be linear the model can be described in terms of any relevant specified change. Here we report results in terms of a five-unit change because it is a simple number and is clinically meaningful in that a therapist can imagine adding five IV turns to a session. It also happens to approximate the average difference in number of IV turns per session between FECT and CT. We, therefore, present the odds of inadequacy, closeness, and therapy being reported by clients who experienced X + 5 IV turns during a session divided by the odds of such reports in equivalent clients (defined by the covariates therapist and client gender) who each
experienced XIV turns during a session. Because the total number of turns varied from session to session, we made an adjustment for it in each of these models. Results indicated an association between therapy focus and closeness (OR = 1.14, 95% CI = 1.01–1.29, p = .030), and there was clear evidence that additional IV turns were associated with client perceived discussions of therapy (OR = 1.47, 95% CI = 1.26–1.71, p = .002).

We also found that in approximately 85% of sessions (Table II) clients reported that progress was made in dealing with problems for which they were in therapy, and in 43% of sessions clients reported relationship improvements in the week after the session. Using the same logistic regression model, we found that additional IV turns were positively associated with reporting of progress and may have been associated with the reporting of relationship improvements. For progress, the odds ratio associated with five additional IV turns was 1.32 (95% CI = 1.11–1.58, p = .001), and there was a trend suggesting that IV may be associated with reporting on relationship improvements (OR = 1.10, 95% CI = 0.98–1.22, p = .10).

We next examined the odds that clients who reported positive relationship changes would attribute those changes to therapy compared with those who reported negative relationship changes. In this model, we only included sessions in which some change in relationships during the prior week was reported. This excluded 245 sessions in which relationships during the prior week were reported to have not changed, leaving 428 (65.4%) sessions for the analysis. The odds were estimated to be 34.5 times as high in the clients who reported positive changes compared with those who reported negative changes (95% CI = 14.8–81.1, p < .001). That is, clients who reported positive changes in relationships were far more likely than those who reported negative changes to attribute the changes to therapy.

### Relationship between IV turns and depression over the course of therapy

Using cross-lagged panel analysis, we examined the crossed-lagged association between (a) the IV turn rate during session t and the presession BDI scores for the same session and (b) presession BDI scores for session t and IV turns during session t + 1. We used log-linear and linear regression models, respectively, to estimate these relationships. We found no evidence to suggest any relationships between IV turns and BDI scores.

### Discussion

Because of the exhaustive nature of this study, it provides convincing evidence that cognitive therapists do not regularly focus on in vivo work in CT for depression. On average, 3.4 of 80 (4.3%) therapist turns-at-speech per CT session had an in vivo focus, whereas FECT evidenced almost three times as much in vivo focus per session. In addition, in CT there was less than one session per client over the full course of therapy in which considerable in vivo work occurred; in contrast, there were more than four sessions per client with such a focus during FECT. We note that the therapists studied in this trial were experienced and highly qualified cognitive therapists, were rated as competent by another expert, and had no proscriptions on their therapeutic technique during the CT phase. Thus, data collected on their behavior during the CT phase give a good indication for how research-quality competent CT is performed. However, they were trained and supervised in FECT by the developers of FECT; this produced increased in vivo work as predicted but may have biased therapists against these interventions during the CT phase (e.g., the allegiance effect; Luborsky et al., 1999). In addition, there are concerns about generalizing from a small sample of manual-guided research therapists to the larger population of practicing clinicians, as with all research on psychotherapy (Garfield, 1996; Goldfried & Wolfe,
work of any kind, the actual amount of active in vivo analysis that standard CT rarely incorporated in vivo identified in the current analysis. Noise of the host of theoretically inert activities likely that the effective in vivo interventions identified in the current study.

Although the current analysis suggests a possible relationship between in vivo work and relationship improvements over the course of therapy, we found no relationship between in vivo work and depressive symptoms over the course of therapy. This is consistent with earlier published results (Kohlenberg et al., 2002) in which FECT demonstrated modest depression improvements over CT but large interpersonal functioning improvements. Alternatively, traditional adherence analyses published by Kohlenberg et al. did suggest that a specific type of in vivo intervention, the use of cognitive change techniques specifically focused on thoughts and feelings about the therapeutic relationship, predicted improved depression at the end of treatment, even though this intervention occurred infrequently.

A major drawback of this study is that the types, timing, competence, and functions of in vivo interventions were not assessed; thus, only gross statements about the relations between in vivo work and process variables can be made. The primary purpose of this project was to document the frequency of in vivo work so raters were trained in a broad definition that may have resulted in the rating of theoretically inert techniques as in vivo and overinclusion of in vivo ratings as a result of misclassification. Thus, it is likely that the effective in vivo interventions identified by the earlier adherence analysis were lost in the noise of the host of theoretically inert activities identified in the current analysis.

In a related vein, although it is clear from this analysis that standard CT rarely incorporated in vivo work of any kind, the actual amount of active in vivo work engaged in during FECT is unclear and probably was less than the reported 9.2%. This may be surprising given that therapists were directly trained and supervised in these techniques by the developers of the approach. As noted in Kohlenberg et al. (2002), traditional adherence analyses also suggested that therapists added in vivo interventions to their repertoires quite slowly. In future studies, larger doses of active in vivo work may facilitate analyses of the relationship among in vivo work, weekly variables, and ultimate outcomes. However, methods for assessing the type, timing, competence, and function of in vivo interventions as well as their frequency should be developed in future studies for a full assessment of this potentially important technique. These efforts will require longer training periods and more sophisticated raters than used in the current study.

References


**Zusammenfassung**

In vivo Prozesse bei kognitiven Therapie für Depressionen: Häufigkeit und Nutzen


**Résumé**

Processus in vivo dans la thérapie cognitive de la dépression : fréquence et bénéfices

Plusieurs variantes récentes de thérapie cognitive (CT) de la dépression tentent d’améliorer le traitement par un focus renforcé sur la survenue de problèmes dans le contexte du setting thérapeutique, du processus thérapeutique et de la relation thérapeutique. Ces processus sont relatés comme du travail in vivo. Pourtant, il est possible que des thérapeutes cognitifs incorporent déjà le travail in vivo. Une étude a comparé la CT standard à une version qui soulignait le travail in vivo selon les guidelines pour la psychothérapie fonctionnelle analytique (FAP). Le traitement en résultat était nommé thérapie cognitive majorée par FAP (FECT). En se servant d’une méthode de codage du processus parole par parole, les auteurs ont examiné le comportement des thérapeutes dans chaque séance de cette étude pour déterminer la fréquence du travail in vivo. Les résultats ont indiqué que les thérapeutes cognitifs étaient engagés seulement rarement dans ce genre de travail, mais que le taux augmentait avec FECT en fonction de plusieurs des variables rapportées chaque semaine par les clients, et compris le progrès dans le traitement, mais pas par les symptômes dépressifs.

**Resumen**

Procesos in vivo en terapia cognitiva para la depresión: Frecuencia y beneficios

Diversas variantes de terapia cognitiva (CT) recientes para la depresión intentan mejorar el tratamiento con un foco más intenso en la ocurrencia de problemas en el contexto del encuadre terapéutico, el proceso terapéutico y la relación terapéutica. Estos procesos son mencionados como trabajo in vivo. Sin embargo, es posible que los terapeutas cognitivos ya hayan incorporado el trabajo in vivo. Un estudio comparó el CT estándar con una versión que enfatizó el trabajo in vivo de acuerdo con los lineamientos diseñados para la psicoterapia analítica funcional (FAP). El tratamiento resultante se llamó FAP- terapia
cognitiva mejorada (FECT, enhanced cognitive therapy). Utilizando un método de codificación de proceso turno-por-turno, los autores examinaron el comportamiento del terapeuta en cada sesión para determinar la frecuencia del trabajo in vivo. Los resultados indicaron que los terapeutas cognitivos se involucraron sólo ocasionalmente en tal trabajo, pero los puntajes aumentaron en el FECT y se relacionaron con diversas variables reportadas semanalmente por el cliente e incluyen el progreso en el tratamiento aunque no los síntomas depresivos.

**Resumo**

**Processos In Vivo na Terapia Cognitiva para a Depressão: Frequência e Benefícios**

Várias variantes recentes da terapia cognitiva (TC) para a depressão tentam melhorar o tratamento focando-se na ocorrência de problemas no contexto da terapia, do processo terapêutico e da relação terapêutica. Estes processos são referidos como trabalho in vivo. Porém, é possível que os terapeutas cognitivos já englobem trabalho in vivo. Um estudo comparou TC estendente com uma versão que enfatiza o trabalho in vivo de acordo com as linhas orientadoras estabelecidas pela psicoterapia analítica funcional (PAF). O tratamento resultante foi designado de PAF- terapia cognitiva alargada (TCFA). Usando um método de codificação das interações do processo terapêutico, os autores analisaram a frequência do trabalho in vivo. Os resultados indicaram que os terapeutas cognitivos apenas ocasionalmente se envolvem em tais trabalhos, mas os rácios aumentou na TCFA e em relação a várias variáveis relatadas semanalmente pelos clientes, incluindo o progresso no tratamento, mas não nos sintomas depresivos.