

## **GROUP MEMBERSHIP AND STAFF TURNOVER AFFECT OUTCOMES**

### **IN GROUP CBT FOR PERSISTENT PAIN**

Amanda C de C Williams & Henry W W Potts

Amanda C de C Williams, PhD CPsychol

*Corresponding author*

Research Department of Clinical, Educational and Health Psychology,

University College London,

Gower Street,

London WC1E 6BT, UK

Tel: +4420 7679 1608

Fax: +4420 7916 1989

E-mail: [amanda.williams@ucl.ac.uk](mailto:amanda.williams@ucl.ac.uk).

Henry W W Potts PhD CStat

Centre for Health Informatics and Multiprofessional Education (CHIME),

Whittington Campus (University College London),

Highgate Hill,

London N19 5LW, UK

Tel: +4420 7288 3383

Fax: +4420 7288 3322

E-mail: [h.potts@chime.ucl.ac.uk](mailto:h.potts@chime.ucl.ac.uk)

30 pages altogether including 4 tables, 1 figure.

## **Abstract**

The effects of two contextual factors, group membership and staff turnover, on the outcome of group cognitive behavioral therapy (CBT) for persistent pain were investigated. The data came from end of treatment and one month follow-up assessments of 3050 individuals who attended an intensive group programme over sixteen years. Intraclass correlations (*ICC*) showed significant intragroup effects on self-efficacy (*ICC* = 0.16 at end of treatment; 0.12 at one month), catastrophizing (*ICC* = 0.06; 0.13) and distance walked (*ICC* = 0.20; 0.19). This underlines the importance of modelling group membership when analyzing data from group interventions. Linear regression showed that high periods of staff turnover were significantly related to poorer outcomes on self-efficacy and distance walked at end of treatment, with the effect on self-efficacy persisting to one month follow-up. Having demonstrated significant contextual effects in an existing data set, further research is needed to explore the mechanisms by which these effects operate.

## **Introduction**

Just as no two therapies are ever the same, no two groups are ever the same, yet identification of common factors in effective groups remains largely descriptive [13,22,40]. We still know little about the contextual factors affecting outcome of group cognitive behavioral treatment (CBT) of persistent pain [24,25,27,29], which is widely recommended for persistent non-cancer pain [11,17,23]. CBT pain management groups tend to fall somewhere between psychoeducation or work groups and group or milieu therapy [10,13,33]. Patients' focus is on individual not group goals, although they may collaborate and draw on collective resources, and relationships within the group are not themselves a target for change.

A historical data set from a specialist treatment centre for group CBT for pain management offered us the possibility of investigating intragroup effects and turnover in the staff team on common patient outcomes. These are not the only possible contextual effects but those on which good data were available.

### *Intragroup effects*

Many psychological treatments show benefits of group delivery as good as those from individual treatment, with possible cost savings [5,13,36]. Common factors in groups include universality and acceptance [12,13,36,37,40], and patients consistently comment on feeling supported and enabled by their peers. It is likely that such characteristics would vary across the groups studied here; additionally, groups were observed to

establish and enforce on their members very different norms about behavior, most noticeably about talking/not talking about pain outside therapeutic sessions. This led us to suspect intragroup correlations in the data, whereby patients in the same group have more similar outcomes than patients in different groups. Where this is found, it represents a problem for statistical analysis of group-based treatment outcomes; Baldwin et al. [2] demonstrated how correction for dependence of observations can substantially reduced the number of outcomes that meet statistical criteria for significant change.

#### *Staff team effects*

Therapists vary in efficacy [38]; so we suggest might therapeutic staff teams. In our centre, each staff team included clinical psychologists, physiotherapists, doctor (consultant anesthesiologist), nurse, and occupational therapist, most working daily with their patient group. Staff worked as tutors or coaches as well as therapists. Our interest was in change within teams and whether it affected the outcome of groups treated during times of change compared to times of stability. In general, experienced staff left and were replaced by less experienced staff who worked under supervision for several weeks, thus the change stretched over a longer time than patients' stay of four weeks. Whatever the skills level of staff leaving and arriving, we anticipated that changes would reduce cohesion within the staff team resulting in a negative impact on outcome.

## Method

### *Sample*

The data were taken from an existing anonymized database of 3050 patients in the period from September 1992 to June 2004, with permission from the appropriate UK NHS Research Ethics Committee. Data were routinely collected with informed consent shortly before the start of a patient's treatment, at the end of treatment, one month after the end of treatment, and at longer term follow-up (not used here due to attrition from follow-up). Patients who were unable to return for follow-up were asked to complete self-report measures and return them by mail.

Patients were referred to the specialized treatment centre from primary care and from specialist care all over the UK, and assessed by a consultant anaesthetist and clinical psychologist. They were accepted if they reported two or more of the following: interference in non-work and/or work activity due to pain; high levels of affective distress; habitual over-activity leading to increased pain; use of high levels of pain medication with little reported benefit; unnecessary use of medical aids; high levels of pain behavior. Exclusion criteria were pain duration less than one year; suitability for further medical treatments following examination; inability to speak English; inability to climb stairs; current major psychiatric disorder (active psychosis, severe depression with high risk of suicidal attempt); current opioid misuse (either chronic illegal opioid use or in a methadone maintenance programme). These are largely similar to those of similar programmes in the USA and Australia [17,36].

Patients were predominantly female (58.9%) and 81.5% were White, 5.2% Afro-Caribbean and 3.8% Asian. Median age was 45 years (range 18 to 84). Just under one quarter (24.6%) were classified as professional or managerial, 40.2% as skilled manual/clerical, and 21.8% as unskilled manual workers. However, only a minority (19.2%) were employed full or part-time, with most permanently sick or disabled (33.0%) or unemployed because of pain (24.6%), and the remainder retired (9.6%), homemakers (4.3%), unemployed for other reasons (1.1%) or students (1.0%). The main site of pain was the lower back (45.1%), and the median chronicity was 6 years.

Treatment consisted of 16 (or occasionally 8) days residential intensive cognitive and behavioral pain management [39], during which patients were taught and rehearsed new behaviors, and in the evenings were accommodated in single rooms that were previously a staff residence, sharing some facilities. Patients socialized in the evenings as much or as little as they wished, and weekends were spent at home or away from the programme. Groups usually consisted of 10 patients. Data were available on 330 groups in total, giving an overall average group size of 9.2. (Average group sizes are reported in Tables 1 and 2, but note that these represent the number of individuals with sufficiently complete data for the analyses reported there and are thus lower than the actual group sizes.)

Staff members were organized into multidisciplinary teams of psychologists, physiotherapists, occupational therapists and nurses, with doctors working across teams. This meant that patients usually related mainly to a core team of four staff members, but

up to seven if staff were part-time. The inpatient treatment programme remained largely consistent in content and process over the years covered by this study.

### *Measures*

Routine assessment covered domains of pain experience, mood, cognitive content and process, activity limitation and social roles, physical performance, and medication use. It was carried out in a standardized manner, mainly by staff not involved in treatment. For this study of effects on change over treatment, a selection is made of physical performance, and two cognitive measures, catastrophizing and self-efficacy. This selection represents diverse outcomes that showed consistent positive change over treatment [24]; there were very small numbers for negative change [20,26].

Physical performance consisted of the metres walked in 5 minutes over a 20 metre indoor course, at the patient's chosen speed, resting if required. Standard instructions were given and a member of staff kept time and measured the distance [14]. The interrater reliability of a similarly conducted 10 minute walk, with which the 5 minute walk is highly correlated, is an *ICC* of 0.994 [14].

Catastrophizing was measured in the first few years using the catastrophizing subscale of the Cognitive Strategies Questionnaire (CSQ [30]), consisting of 6 statements, for example "I feel I can't stand it anymore", to which the patient responds in terms of frequency of thinking it when in pain, from 0 = *never* to 6 = *always*. Its internal consistency was previously reported as  $\alpha = 0.78$  [30]. This was superseded by the Pain Catastrophizing Scale [34], consisting of 13 statements such as "It's awful and I feel that

it overwhelms me”, with the response options of 0 = *not at all*, 1 / 2 / 3 = *to a slight/moderate/great degree*, and 4 = *all the time*. Cronbach’s alpha for the PCS was 0.93 at baseline and 0.95 post-treatment in a partly overlapping data set. Because the CSQ and PCS have similar items, with the latter scale being developed from the former, we created a standardized catastrophizing index. This was done by fitting a Normal distribution to the CSQ data and PCS data separately: for each group, values were converted into the number of standard deviations from the mean. This index was then used across both groups.

Self-efficacy was measured by the Pain Self-Efficacy Questionnaire (PSEQ [1]), this asks patients to rate their confidence in achieving activities in ten different areas despite pain. Each statement (e.g., “I can enjoy things despite the pain”) is scored from 0 = *not at all confident* to 6 = *completely confident*, producing a total score from 0 to 60. Internal consistency has previously been reported as  $\alpha = 0.92$  [1].

#### *Intragroup effects*

Analysis was done using Stata 8.2 for Windows. The presence of intragroup effects is measured by the intraclass correlation (*ICC* [7]). Non-zero *ICCs* are clinically interesting, but also have implications for the further analysis of the data. Unless taken into account by using appropriate techniques to model the variance structure of the data, non-zero *ICCs* lead to inflated significance levels in regression analyses. The design effect, or inflation factor, measures this and is based on the *ICC* and the average cluster size (*m*) [10].



$$\text{design effect} \approx 1 + (m - 1) \times ICC$$

Thus, if the group size equals one (i.e. no groups) or if the *ICC* equals zero, the design effect is one. As a rule of thumb, a design effect greater than two indicates that the clustering in the data needs to be taken into account during estimation. Given the possible presence of group effects, all the regression analyses performed in this paper took into account the clustering.

#### *Staff turnover effect*

Records provided the start and end date of employment of members of staff. From these, the number of staff members starting or ending in each month, and the total staff numbers, were calculated. This total number of staff starting or ending was our raw figure of staff turnover. A linear regression analysis was performed applying the monthly turnover scores to those patients whose treatment was predominantly in that month. This metric is an approximation of any real effect of turnover. Given the exploratory nature of this study, we also calculated three further metrics that we thought might model turnover differently. The raw figure was divided by total staff numbers to give a proportion, on the grounds that the effect of a change in staff may be less noticeable when the staff complement was larger.

The raw figure therefore presumes that the effect of a staff member joining or leaving is immediate and short-lived. Informal consultation with experienced staff produced a consensus that the disruption of a staff member joining or leaving extended before and

after the end of employment. Staff often left before their recorded official end date to use outstanding leave entitlements; there was frequently a short gap before they were replaced, and then there was an observation and training phase for the new staff member. Anxiety in the team rose while incorporating the new member and then diminished when the new member discharged responsibilities reliably. It was thus decided to calculate a smoothed figure, where the effect of someone's starting or leaving was spread over the month before, the month s/he left or started in and the three months afterwards. This smoothing function was calculated as one-twelfth of the effect in the month before a staff member arrived or left; one third of the effect in the month of starting or leaving; one quarter in the first month afterwards; one sixth in the second month afterwards; and one twelfth in the third month afterwards.

## Results

### *Intragroup effects*

The outcome measures all approximately follow Normal distributions. It seemed reasonable to presume that any ICCs would be highest for the data collected on the last day of treatment, so analyses were carried out on those data first.

*Table 1 about here*

Significantly nonzero intraclass correlations were seen for three of the four variables, with substantial design effects (greater than two) seen for self-efficacy (PSEQ) and the walk test. The result for the number of good days was not significantly different from zero, although note the smaller sample size and, thus, reduced power. Having seen some non-zero intraclass correlations on the last day, we repeated the analyses on the one month follow-up data. Significant ICCs persisted at one month follow-up (Table 2).

*Table 2 about here*

### *Staff turnover effects*

The outcome data were collected from patients admitted between September 1992 and June 2004. Figure 1 shows the total number of staff and the staff turnover per month over this period. In all, there were 21 psychologists, 11 physiotherapists, 10 occupational therapists and 6 nurses who were employed over the period, and one doctor. Several other medical staff were peripherally involved for short stretches of time and their data are not included. One psychologist (AW), one physiotherapist and the doctor were employed

over the entire period. At the other extreme, the shortest total time any health professional was employed at the unit was 10 months; while the shortest period worked was 5 months (for an individual who had returned to the unit after a break).

*Figure 1 about here*

The raw measure of turnover was used to predict the four outcome measures on the last day and at one month follow-up using univariate linear regressions taking into account the group variance structure (Table 3). The regression coefficients, B, are given.

*Table 3 about here*

Self-efficacy (PSEQ) and the distance walked showed statistically significant relationships with the turnover metric at the end of treatment. The effect is weaker but can still be seen at one month follow-up, although the result for distance walked just fails to reach significance ( $p = 0.055$ ). Catastrophizing shows a significant relationship at the end of treatment, but not at one month follow-up.

Table 3 gives values for the regression coefficient, B. The (raw) turnover metric is the number of changes in staff (counting both departures and arrivals) in a month, so the B values are the average change in the outcome score for one more person joining or leaving that month. Thus, for each change in staffing, patients on average had a PSEQ

score that was 1.2 points lower, a Normalized catastrophizing score that was 0.04 standard deviations higher, and walked 4.6 metres less at the end of treatment.

The analyses were repeated using the three supplementary metrics, namely the turnover per total staff size; a turnover figure smoothed asymmetrically; and a smoothed turnover figure per total staff size. These different metrics are inherently correlated: the smallest relationship was between proportional turnover and smoothed proportional turnover (Kendall's rank correlation  $\tau_b = 0.50$ ,  $p < 0.001$ ); the highest was between raw turnover and proportional turnover (Kendall's  $\tau_b = 0.93$ ,  $p < 0.001$ ). Each of the turnover statistics was tested as a predictor of the key outcomes at end of treatment (Table 4). Broadly, the different turnover statistics show the same pattern of results for self-efficacy and distance walked. However, the result is not significant for distance walked with the smoothed metric. With catastrophizing, the significant result seen with the raw figure is not seen with any of the other measures.

The analyses were repeated for one month follow-up. Broadly, the different turnover statistics again show a similar pattern of results, although there are a number of differences. The self-efficacy result held for the proportional figure, but fell just outside statistical significance for the smoothed, proportional figure ( $p = 0.051$ ) and not at all significant for the smoothed figure. Distance walked was not quite significant on the raw figure at one month follow-up, but was on the two proportional figures.

It is difficult to conclude from these results that one or other measure of turnover better represents the clinical reality. Smoothed results show smaller associations, which may imply that smoothing is an inappropriate way to represent the effect that we are trying to model.

## Discussion

We found clear evidence of intragroup effects on three out of the four outcome variables (self-efficacy, catastrophizing and distance walked), both at the end of treatment and at one month follow-up. The intraclass correlations observed are large enough to warrant group-corrected statistics being used to avoid Type I error, as argued by Baldwin et al. [2], and to underline the need for clinicians to be aware of how the dynamics of different groups can affect individual participants. We also found significant relationships between measures of staff turnover and two of the outcome variables (self-efficacy and distance walked, with marginal results for catastrophizing), persisting weakly at follow-up. These are modest in size: that for self-efficacy is about 14% of the mean change seen over treatment (see [30]) for each change in staff in the month. This bears out the attention given to the context of group treatment by Moos [21,22] and others, although we were unable to apply any of the measures developed in their studies.

We approached the analysis of this exploratory study with a set of a priori hypotheses based on experience of conducting group treatment, but this is a post hoc analysis. As such, we were constrained by the data available and the observational nature of the study, and by our choice of outcome measures.

A measure of turnover was a greater challenge. Data were calculated based on employment records, but as discussed above, these may not reflect the actual dates individuals left the unit. Additionally, we necessarily made assumptions that the effect of any staff member leaving was the same as another, regardless of discipline or length of

service. As our turnover measure is only a rough approximation, it is possible that our effect sizes underestimate the true effect (cf. [9]). Broadly, the proportional model gave similar results to the raw turnover, and the smoothed versions of raw and proportional turnover produced smaller effects, so perhaps the effects of turnover were briefer than experienced by the staff who suggested a smoothed model.

Direct rather than proxy measures would have been preferable for testing our hypotheses [27]: we did attempt to measure patients' self-rated therapeutic alliance and group cohesion. Therapeutic alliance, to patient group and to staff team, was measured with an adapted Empathy Scale [6], and group cohesion with a Group Climate Scale short form [18], but both produced high ratings with a ceiling effect and insufficient variance to allow meaningful analysis. This may have been because, when sampled near the end of the inpatient programme, patients were generally very positive about both staff and fellow patients; sampling a few weeks after the end of the programme might have obtained more moderate scores and greater variance. We would hypothesise that, as in individual CBT, therapeutic alliance is an important vehicle for therapeutic effects [5,20,28,32].

There are several possible confounding effects. Turnover may have had an effect for the reasons discussed, but it could also be a symptom of some other process with a direct effect. For staff turnover, we checked for but found no seasonal effect in data plots. What may explain the results observed, however, is some common factor affecting the unit that both increased staff turnover and impacted on patient outcomes. Therapist emotional



wellbeing has been reasonably consistently related to better outcomes [3]. As is not uncommon in a clinical setting, the treatment unit had periods of strengthened institutional support and growth and others when its future was uncertain. It is likely that these difficult periods were reflected in poorer team cohesion as well as in turnover, and that staff who were expending energies on the unit's survival had less to offer patients. These circumstances could not be captured but were likely to depress patient outcomes. Likewise, periods of high staff morale may have had a positive effect on patient outcomes through factors such as better engagement and therapeutic alliance, and more team discussion of patients' needs and progress. If these different periods were also associated with differential turnover, we would observe a correlation between turnover and outcomes. Higher turnover during difficult periods would be unsurprising, but it should also be noted that periods of high morale were often associated with growth, that also entailed increased turnover. Additionally, some turnover resulted from positive personal life events: staff getting married; having children; taking a career break. Even if turnover were just a symptom of staff morale, we would not be replacing one contextual effect for another: the effect of unit milieu on outcome stands, demonstrating important processes that are largely neglected in research on group interventions of this sort.

While detrimental effects of increased turnover seem plausible, how such an effect might occur in detail remains unclear. The event of experienced staff leaving has multiple effects, most obviously on other staff and on therapeutic content. New staff may have delivered treatment differently: they usually brought enthusiasm and perhaps a tendency to adhere more closely to therapeutic techniques, but had to adapt these techniques to

programme content, and were often observed to stick to a script rather than following patients' leads (as described in individual therapy by new therapists: 29, chapter 16). New staff may simply be less skilled when it comes to this particular treatment approach, but they can also be more flexible and imaginative about skills teaching and practice. The arrival of new staff has been associated with more disruption in inpatient psychiatric wards, but this is not consistently found [4].

It may also be the case that it is not the staff who have left or joined that matter, but how the remaining team members feel. While a staff member with low morale leaving an otherwise healthy team, and being rapidly replaced, may have little effect, the loss of good staff and the problems in replacing them may increase staff members' subjective sense of stress. Anecdotally, there was also a unit "wisdom" that staff in their last few weeks at work became less mindful of the possible consequences for other team members of their individual clinical decisions.

If our observations of contextual effects are correct, what does this imply? As has been repeatedly prescribed in the literature, but less often implemented, group effects must be considered in the analyses of studies of group treatments [2]. Analysis of variance models will not identify sources of variability within groups, and direct tests of mediation are needed to understand processes of change better [29]. We would also encourage the exploration of staff-patient processes in group treatment that might be loosely described in terms of therapeutic alliance (or alliances – within the patient group and between staff and patients), and staff team cohesion.

Such effects may not be seen in prospective studies or trials, as these are often brief, with staff generally employed for the duration of the trial and adhering to a detailed protocol for treatment. But there are many established programmes using CBT in the treatment and rehabilitation of people with persistent pain, and process research so far has tended to focus on cognitive change during treatment in relation to longer term outcomes, emotional and functional, but has ignored the context in which that change occurs [24]. Such programmes could include measures of staff morale and team functioning, and patient sampling for therapeutic alliance and group cohesion [15]. In particular, given how common such groups are in clinical health settings, we suggest that others examine their data for similar effects over patient cohorts and, where possible, across staff changes. We have done no more than demonstrate effects of clinically meaningful size and speculate on possible mechanisms. The next step, identifying some of those mechanisms, would need to be guided by stronger theory about group processes, and ideally would allow for the possibility of interactions with other processes. For instance, Ogradniczuk and Piper [28] found engagement of group members was related to outcome only when conflict levels were low, and suggest, as did Yalom [40], that low conflict may be a necessary condition for engagement in work within the group.

There are implications here for the training and support of staff. Health system policies and organizational practices have mixed effects on retention or loss of experienced staff. Turnover and morale issues are widely relevant, for managers as well as for clinicians,

since change is a constant in health systems and healthcare services, but the possible costs in decrements to patient outcomes are rarely calculated.

The evidence presented here, showing that contextual factors can impact on patient outcomes, has clinical implications for designing good interventions as well as implications for those testing or investigating group interventions. The mechanisms by which the milieu of the patient group and the staff team can affect outcomes are worth investigating in other group contexts, and we would encourage greater recognition of these processes in all patient groups, not only those conforming to conventional group therapy principles [35].

**Neither author has any conflict of interest to declare.**

## References

- 1 Asghari A, Nicholas MK. Pain self-efficacy beliefs and pain behavior: a prospective study. *Pain* 2001;94:85-100.
- 2 Baldwin SA, Murray DM, Shadish WR. Empirically supported treatments or Type I errors? Problems with the analysis of data from group-administered treatments. *J Cons Clin Psychol* 2005;73:924-935.
- 3 Beutler LE, Malik M, Alimohamed S, Harwood TM, Talebi H, Noble S, Wong E. Therapist variables. In M.J. Lambert, editor, *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* 5<sup>th</sup> ed. New York, NY: John Wiley & Sons, 2004. pp. 227-306
- 4 Bowers L, Jeffrey D, Simpson A, Daly C, Warren J, Nijman H. Junior staffing changes and the temporal ecology of adverse incidents in acute psychiatric wards. *J Adv Nursing* 2007;57:153-160.
- 5 Burlingame GM, MacKenzie KR, Strauss B. Small group treatment: evidence for effectiveness and mechanisms of change. In M.J. Lambert, editor, *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* 5<sup>th</sup> ed. New York, NY: John Wiley & Sons, 2004. pp. 647-696).
- 6 Burns DD, Auerbach A. Therapeutic empathy in cognitive-behavioural therapy: does it really make a difference? In PM Salkovskis, editor. *Frontiers of Cognitive Therapy*. NY: Guilford Press, 1996. pp. 135-164.
- 7 Campbell M K, Grimshaw JM, Elbourne DR. Intraclass correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. *BMC Med Res Methodol* 2004;4:9.

- 8 Castle NG. Measuring staff turnover in nursing homes. *The Gerontologist* 2006;46:210-219.
- 9 Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol* 1991;114:906–914.
- 10 Elwyn G, Koppel S, Greenhalgh T, Macfarlane F. *Groups: a guide to small group work in healthcare, management, education and research*. Oxford, Oxon: Radcliffe Medical Press, 2001.
- 11 European Union (2004). *European Guidelines for the management of chronic non-specific low back pain*. Retrieved 11<sup>th</sup> January 2009 from [http://www.backpaineurope.org/web/html/wg2\\_results.html](http://www.backpaineurope.org/web/html/wg2_results.html)
- 12 Frank JD, Frank JB. *Persuasion and healing: a comparative study of psychotherapy*. Baltimore, MD: Johns Hopkins University Press, 1991.
- 13 Hajek P. Group therapy. In S Ayers, A Baum, C McManus, S Newman, J Weinman, R West, editors. *Cambridge Handbook of Psychology, Health & Medicine*, 2<sup>nd</sup> ed. Cambridge, Cambs: Cambridge University Press. pp. 352-355.
- 14 Harding VH, Williams ACdeC, Richardson PH, Nicholas MK, Jackson JL, Richardson I, Pither CE. The development of a battery of measures for assessing physical functioning of chronic pain patients. *Pain* 1994;58:367-375.
- 15 Joyce AS, Piper WE, Ogrodniczuk JS. Therapeutic alliance and cohesion variables as predictors of outcome in short-term group psychotherapy. *Int J Group Psychother* 2007;17:269-296.
- 16 Keefe FJ, Beupré PM, Gil KM, Rumble ME, Aspnes AK. Group therapy with patients with chronic pain. In DC Turk & RJ Gatchel, editors. *Psychological*

- approaches to pain management: A practitioner's handbook, 2nd ed. New York, NY: Guilford Press, 2002. pp. 234-255
- 17 Keefe FJ, Rumble ME, Scipio CD, Giordano LA, Perri LM. Psychological aspects of persistent pain: current state of the science. *J Pain* 2004;5:195-211.
  - 18 MacKenzie KR. The clinical application of a group measure. In RR Dies & KR MacKenzie, editors. *Advances in Group Psychotherapy: integrating research and practice*. NY: International Universities Press, 1983. pp 159-170.
  - 19 Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *J Cons Clin Psychol* 2000;68:438-450.
  - 20 Mohr DC. Negative outcomes in psychotherapy: a critical review. *Clin Psychol Science Practice* 1995;2:1-27.
  - 21 Moos R. Context and coping: towards a unifying conceptual framework. *Am J Community Psychol* 1984;12:5-25.
  - 22 Moos RH. Social contexts: transcending their power and their fragility. *Am J Community Psychol* 2003;31:1-13.
  - 23 Morley SJ, Eccleston C, Williams ACdeC. Systematic review and meta-analysis of randomized controlled trials of cognitive behavior therapy and behavior therapy for chronic pain in adults, excluding headache. *Pain* 1999;80:1-13.
  - 24 Morley SJ, Keefe FJ. Editorial: Getting a handle on process and change in CBT for chronic pain. *Pain* 2007;127:197-8.
  - 25 Morley S, Williams ACdeC. Conducting and Evaluating Treatment Outcome Studies. In: DC Turk & R Gatchel, editors, *Psychological Approaches to Pain*

- Management: A Practitioners Handbook, 2<sup>nd</sup> ed. New York, NY: Guilford Press, 2002.pp 52-68.
- 26 Morley S, Williams ACdeC, Hussain S. Estimating the clinical effectiveness of cognitive behavioral therapy in the clinic: evaluation of a CBT informed pain management programme. *Pain* 2008;137:670-680.
  - 27 Newton-John TR, Geddes J. The non-specific effects of group-based cognitive-behavioral treatment of chronic pain. *Chron Illness* 2008;4:199-208.
  - 28 Ogrodniczuk JS, Piper WE. The effect of group climate on outcome in two forms of short-term group therapy. *Group Dynamics: Theory, Res Practice* 2003;7:64-76.
  - 29 Pachankis JE, Goldfried MR. On the next generation of process research. *Clin Psychol Rev* 2007;27:760–768.
  - 30 Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment, *Pain* 1983;17:33-44.
  - 31 Roth AD, Fonagy P. What works for whom: a critical review of psychotherapy research, 2<sup>nd</sup> ed. New York, NY: Guilford Press, 2005.
  - 32 Safran J, Segal Z. *Interpersonal process in cognitive therapy*. New York, NY: Basic Books, 1996.
  - 33 Smith P. *Group processes and personal change*. London, UK: Harper Row, 1980.
  - 34 Sullivan MJL, Bishop S, Pivik J. The Pain Catastrophizing Scale: development and validation. *Psychol Assess* 1995;7:524-532.
  - 35 Thompson S. *The Group Context*. London, UK: Jessica Kingsley Publishers, 1999.



- 36 Thorn BE, Kuhajda MC. Group cognitive therapy for chronic pain. *J Clin Psychol* 2006;62:1355-1366.
- 37 Toro PA, Rappaport J, Seidman E. Social climate comparison of mutual help and psychotherapy groups. *J Cons Clin Psychol* 1987;55:430-431.
- 38 Wampold BE, Brown GS. Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. *J Cons Clin Psychol* 2005;73:914-923.
- 39 Williams ACdeC, Nicholas MK, Richardson PH, Pither CE, Justins DM, Chamberlain JH, Harding JR, Ralphs JA, Jones SC, Dieudonne I, Featherstone JD, Hodgson DR, Ridout KL, Shannon EM. Evaluation of a cognitive behavioral programme for rehabilitating patients with chronic pain. *Brit J Gen Practice* 1993;43:513-518.
- 40 Yalom ID. *The theory and practice of group psychotherapy*, 4th ed. New York, NY: Basic Books, 1994.

### *Acknowledgments*

The support of the INPUT Special Purposes Fund, Grant 815 from the Guys & St Thomas' Charitable Foundation, is gratefully acknowledged, as is Guy's & St Thomas' NHS Hospitals Trust for permission to use the data. Thanks to Chris Barker for helpful comments on an earlier draft; to André Lynam-Smith for painstaking work on the staff change and group data; and to Stephen Morley for discussion of the research plan.

Figure 1 *Staff numbers and turnover September 1992 to June 2004*

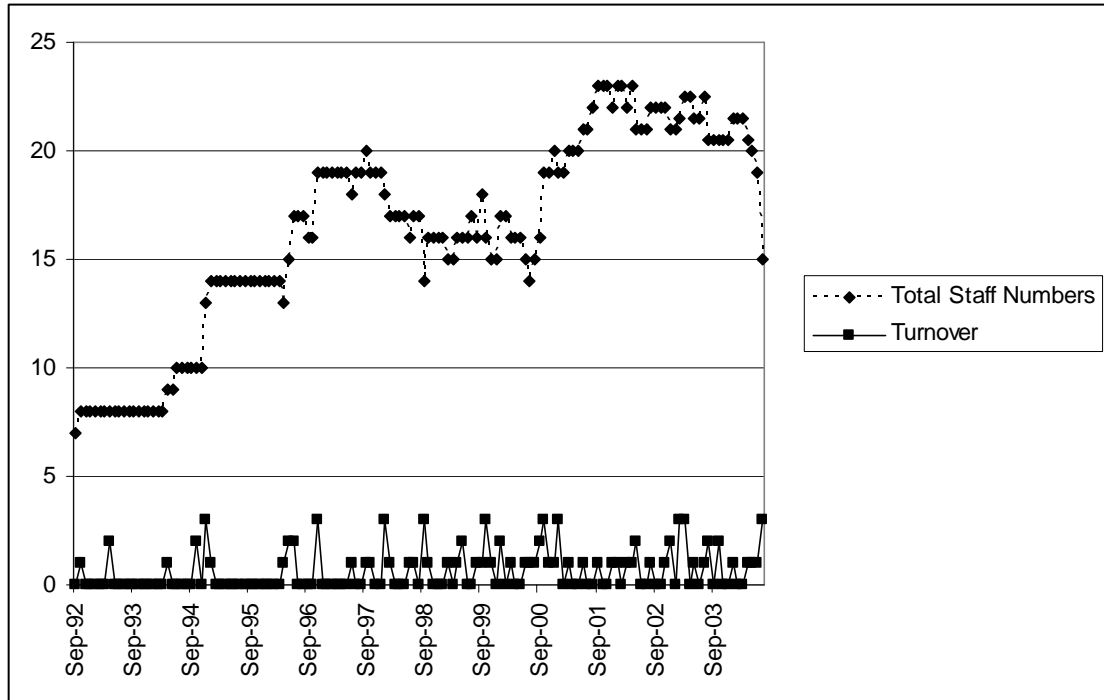


Table 1 *Intraclass correlations of groups for outcomes at the end of treatment*

| Variable        | n    | average cluster | intraclass correlation | standard error | 95% CI for ICC | design effect | using 95% CI |
|-----------------|------|-----------------|------------------------|----------------|----------------|---------------|--------------|
| Catastrophizing | 2915 | 8.89            | 0.06                   | 0.013          | 0.04, 0.09     | 1.5           | 1.3, 1.7     |
| PSEQ            | 2737 | 8.37            | 0.16                   | 0.018          | 0.12, 0.19     | 2.2           | 1.9, 2.4     |
| Walk test       | 2647 | 8.14            | 0.20                   | 0.020          | 0.16, 0.24     | 2.4           | 2.1, 2.7     |

Table 2 *Intraclass correlations of groups for outcomes at one month follow-up*

| Variable        | n    | average cluster | intraclass correlation | standard error | 95% CI for ICC | design effect | using 95% CI |
|-----------------|------|-----------------|------------------------|----------------|----------------|---------------|--------------|
| Catastrophizing | 2350 | 7.32            | 0.13                   | 0.018          | 0.09, 0.16     | 1.8           | 1.6, 2.0     |
| PSEQ            | 2381 | 7.37            | 0.12                   | 0.018          | 0.08, 0.15     | 1.8           | 1.5, 2.0     |
| Walk test       | 2212 | 6.96            | 0.19                   | 0.022          | 0.15, 0.24     | 2.1           | 1.9, 2.4     |

Table 3 *Results of the univariate linear regressions, each with raw turnover as the independent variable*

| Variable        | Time                | n    | B     | 95% CI for B | p       |
|-----------------|---------------------|------|-------|--------------|---------|
| Catastrophizing | End of treatment    | 2914 | 0.04  | 0.00, 0.08   | 0.045   |
|                 | One month follow-up | 2351 | 0.02  | -0.03, 0.06  | 0.5     |
| PSEQ            | End of treatment    | 2737 | -1.20 | -1.70, -0.71 | <0.0001 |
|                 | One month follow-up | 2382 | -0.94 | -1.51, -0.36 | 0.0014  |
| Walk test       | End of treatment    | 2647 | -4.61 | -8.74, -0.47 | 0.029   |
|                 | One month follow-up | 2213 | -4.45 | -8.99, 0.1   | 0.055   |

Table 4 *Results of the univariate linear regressions, four each with the three further turnover metrics as the independent variables; end of treatment*

| Dependent variable   | Independent variable  | Effect of turnover at end of treatment |       |              |          |
|----------------------|-----------------------|--|-------|--------------|----------|
|                      |                       | N                                      | B     | 95% CI for B | p        |
| Catastrophizing      | smoothed              |  | 0.04  | -0.04, 0.11  | 0.3      |
|                      | proportional          | 2914                                   | 0.49  | -0.19, 1.18  | 0.2      |
|                      | smoothed proportional |  | 0.57  | -0.72, 1.86  | 0.4      |
| PSEQ (self-efficacy) | smoothed              |  | -1.16 | -2.12, -0.20 | 0.018    |
|                      | proportional          | 2737                                   | -22.6 | -30.7, -14.4 | < 0.0001 |
|                      | smoothed proportional |  | -26.9 | -43.6, -10.2 | 0.0016   |
| Walk test            | smoothed              |  | -4.13 | -11.97, 3.70 | 0.3      |
|                      | proportional          | 2647                                   | -115  | -187, -42    | 0.0021   |
|                      | smoothed proportional |  | -179  | -320, -39    | 0.013    |