Towards evidence-based clinical supervision: the development and evaluation of four CBT guidelines

Derek Milne* **and Chris Dunkerley**

Newcastle University, and Northumberland, Tyne and Wear NHS Trust

Abstract. Clinical supervision is central to evidence-based practice (EBP) and continuing professional development (CPD), but the evidence base has made little impact on supervision, a major form of CPD. We unite the two by developing four evidence-based guidelines for cognitive behavioural therapy (CBT) supervision. The guidelines were designed to address the supervision cycle (i.e. collaborative goal-setting; methods of facilitating learning; evaluation and feedback) within the context of the supervision alliance. Guideline development followed the National Institute for Clinical Excellence approach, including a representative stakeholder working group (with local service users and supervisees), a national group of supervisors and supervisor trainers, plus an expert reference group. A total of 106 such participants completed an *ad-hoc* guideline evaluation tool, designed to provide a multi-dimensional reaction evaluation of the guidelines. The guidelines were all rated favourably, satisfying the key initial criteria of accuracy and acceptability, and were judged to represent a CBT approach to supervision. It is concluded that the use of the guidelines might help CBT supervisors to better meet demands for CPD (including specialization in supervision) and EBP.

Key words: CBT supervision, evidence-based practice, guidelines.

Introduction

Clinical supervision lies at the heart of the modernized National Health Service (NHS), providing an instrument for delivering clinical governance [Department of Health (DoH), 1999, fostering evidence-based practice (EBP; DoH, 2001*a*) and promoting continuing professional development (CPD) (DoH, 2001*b*). This is reflected in the core national standard that 'clinical care and treatment is carried out under supervision' (DoH, 2004, p. 29). However, implementing these policies has proved problematic. In the case of EBP, many clinicians view tools like clinical guidelines and treatment manuals as relatively unhelpful (Lucock *et al.* 2006), and perhaps for this reason EBP has had little impact on essential professional activities like clinical supervision (Cape & Barkham, 2002).

---

*Author for correspondence: Dr D. Milne, Doctorate in Clinical Psychology, Ridley Building, Newcastle University, Newcastle NE1 7RU, UK. (email: d.l.milne@ncl.ac.uk)

This divide between policy and practice is clearly a problem for central government, but many practitioners are also placed in an uncomfortable position by the 'divisive concept' of EBP (Dyer, 2008), leading some of them to call for a solution. To illustrate, in the case of cognitive behavioural therapy (CBT), Pretorius (2006) argued that, 'It is not sufficient to provide supervision, without consideration of what constitutes good practice and current best evidence in the field' (p. 413). The author of the first supervision manual ever produced also recognized the need for such information, as it can help practitioners to be aware of intervention options, information which seemed to improve their clinical results (Neufeldt, 1994). In general, guidelines can provide the kind of instruction that results in improved learning outcomes (Kirschner *et al.* 2006).

However, these appear to be lone voices, as few formal supervision guidelines have appeared in the past 15 years. According to the preliminary review in Milne (2009), there have only been three further published guidelines (Henggeler *et al.* 2002; Baltimore & Crutchfeld, 2003; Fall & Sutton, 2004), although bodies such as the British Association for Behavioural and Cognitive Psychotherapies (BABCP) have produced guidance leaflets (e.g. Lewis, 2005), and its journal (*Behavioural and Cognitive Psychotherapy*) has published papers with general guidelines (e.g. Townend *et al.* 2002; Pretorius, 2006). In summary, clinical supervision appears to be highly valued, yet is paradoxically under-developed.

There are a number of plausible reasons for this paradox. First, the evidence base has been criticized on methodological grounds (Ellis *et al.* 1996), with the result that Ellis & Ladany (1997) cautioned against using this evidence base as a guide to supervisory practice. There have also been few expert consensus statements of the kind that could at least interpret the best-available research and theory. A further reason is the dearth of rigorous evaluations of supervisor training. For instance, Kilminster & Jolly (2000) noted that training programmes are rarely empirically or theoretically grounded. As noted above, yet another factor is the generally negative attitude that many practitioners appear have towards EBP (Dyer, 2008) and their preference for a minimally guided, constructivist approach (Kirschner *et al.* 2006). Not least, introducing guidelines represents a form of organizational change, which tends to be an exacting challenge (see, e.g. Northcott, 1998). These are some of the factors that help us to understand this paradox. How might it be addressed?

Latterly the situation has improved, with the recognition that there are useful seams of sound supervision research (e.g. Milne & James, 2000); the availability of expert guidance statements (Falender *et al.* 2004); the publication of better training studies (e.g. Bambling *et al.* 2006); a softening of EBP, with the growing recognition that practice recommendations must include both empirical evidence and professional consensus (Parry, 2000); and a recognition that supervision can do harm if allowed to proceed without guidance (Ladany *et al.* 1997). It is also worth noting that not all practitioners hold equally negative attitudes to EBP: adherents to CBT were found to be significantly more positive about EBP (including guidelines) than were comparable groups (Lucock *et al.* 2006). Consequently, as Green & Youngson (2005) have argued, 'the reasonable goal of "science-informed practice"... should not be beyond our reach' (p. 2), at least in CBT supervision.

Guidelines

One way to promote science-informed practice is to develop evidence-based guidelines for clinical supervision (Watkins, 1997). The American Psychological Association (APA) defined practice guidelines as 'a set of statements that recommend specific professional conduct'

(APA, 2002, p. 1048). The APA emphasized that these guidelines are not mandatory, nor are they intended to take precedence over clinical judgement. Rather, guidelines are a tool for bringing the evidence base to bear on practice, with the aim of assisting practitioners to form appropriate judgements, thereby improving the quality of healthcare (Grilli *et al.* 2000).

However, although there are now many guidelines regarding mental health practice, there have been few guidelines that address clinical supervision, and those that do exist have had significant limitations. For example, Neufeldt (1994) developed and evaluated a manual for new counselling supervisors. The manual presented 26 supervision strategies, each with a description, step-by-step procedures, and notes on usage. Although the simple reaction evaluations that were reported were positive, the manual drew on the evidence base partially and unsystematically, and it may also lack generalizability to non-counsellors. Perkins & Mercaitis (1995) developed a guide to session planning and feedback, evaluating it by means of a control group design. However, the guide was addressed primarily to supervisees. Northcott (1998) generated a set of practice guidelines within an acute medical ward in the NHS, but these were developed informally and were not evaluated at all. Fall & Sutton's (2004) clinical supervision manual is well-grounded in both the evidence base and professional consensus, but the manual was designed for counselling supervisors working in the USA, and as such lacks direct relevance to the NHS. Baltimore & Crutchfield (2003) developed a manual that was more broadly aimed at new clinical supervisors in the helping professions. It provided good coverage of process and ethical issues, but little on supervisory techniques such as modelling.

In summary, there are some difficulties with existing guidelines, such as the need for guidelines to be adapted to local/national needs (such as the NHS; Parry, 2000) and to be evaluated and evidence-based. We therefore took three steps in developing the present guidelines, based upon the advice of NICE (2003) and Parry (2000). First, we conducted a systematic review of the evidence for clinical supervision; second, we developed a model of clinical supervision that was broad enough to be acceptable to mental health practitioners within the NHS; and third, we sought professional consensus and evaluation at every stage of the guideline development process. Therefore, at the outset we sought integrative guidelines derived inductively from what was known, guidelines that would be highly relevant to CBT supervision.

Literature review

Dunkerley *et al.* (2004) undertook a systematic review of evidence-based supervision. The review aimed to assess the evidence base for clinical supervision and its component parts and practices, to evaluate the trustworthiness of this evidence through an assessment of methodological rigour, and to provide the evidence base for developing guidelines for clinical supervision. NICE (2003) provided a format for systematic reviews that feed into guideline development. Although we based our approach upon this format, in its pure form it was inappropriate for a primarily psychological literature. Therefore, we adapted it with reference to an evaluative manual used by Milne & James (2000) and guidance produced by the Centre for Reviews and Dissemination (2001).

Supervision model

The second step was to develop a model of supervision. The tandem model aims to be accessible, integrative and evidence-based (Milne & James, 2005). It conceptualizes
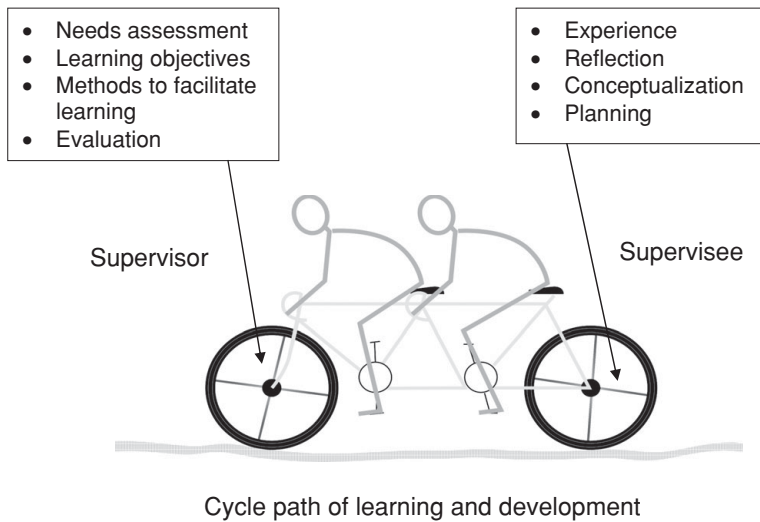
Fig. 1. The tandem model of clinical supervision.

supervision as being like two cyclists on a tandem (see Fig. 1). By analogy, as the leader, the supervisor steers the tandem, the supervisee follows. The front wheel, appropriately under the immediate control of the supervisor, represents the interrelated steps of needs assessment, agreeing learning objectives, use of methods to facilitate learning, and evaluation. The guidelines essentially concern this wheel plus the relationship between the cyclists. The rear wheel represents the experiential learning cycle (Kolb, 1984) with its four modes: experience, reflection, conceptualization, and planning. This is therefore closer to the supervisee's sphere of operation. If these wheels are operating effectively, the tandem proceeds on a path of learning and development. This is an analogy for what was later referred to as the Evidence-Based Clinical Supervision model (Milne, 2009). This subsumes the CBT approach to supervision, but suggests enhancements (e.g. a greater emphasis on experiential learning than is customary in CBT supervision), as detailed in Milne (2008).

Professional consensus

The third step was to seek professional consensus. We did this in two ways (following NICE, 2003, and Shekelle *et al.* 1999). First, we formed a Guideline Development Group (GDG) comprised of key stakeholders (practising clinical supervisors, clinical tutors from a local training programme for mental health professionals, trainees from that programme, and a service user). The GDG aimed to advise on the scope of the guidelines, to assist in the process of guideline development (e.g. the identification and prioritization of topics), and to advise on evaluation. Second, we asked both experts and potential users of the guidelines to formally evaluate them. This evaluation is the primary focus of the present study.

Aim and objectives

Our general aim was to foster evidence-based clinical supervision (Milne, 2009).

Our objectives were to:

(1) Develop four guidelines that operationalized the tandem model, which combined the evidence base with professional consensus, and which were responsive to NHS needs.
(2) Evaluate the degree to which the guidelines were accurate and acceptable.

Method

Design

The acceptability of the guidelines was assessed within a cross-sectional survey design, intended to obtain a reaction evaluation (Kirkpatrick, 1967). A reaction evaluation taps the likes and dislikes of participants with regard to an intervention like a guideline. The dominant methodology was action research, in the sense that we took a participative and iterative approach to defining and solving locally relevant problems (Boog, 2003). Given the 'real world' nature of the evaluation, we used an opportunity sample of participants.

Participants

The guidelines were formally evaluated by users (i.e. clinical supervisors), their tutors (i.e. colleagues who organized and delivered supervisor training workshops, supported the supervisors, etc.), by national experts, and by others (the GDG), giving a total sample size of $n = 106$. This approach follows NICE (2003) and Parry (2000). Specifically, the participants were:

(1) *The Guideline Development Group* (*GDG*; $n = 13$). The group comprised three clinical psychologists who were practising supervisors, two clinical tutors from the University of Newcastle Doctorate in Clinical Psychology programme, one member of a service user group (Launchpad), one mental health nurse, three clinical psychology trainees, and one assistant psychologist.
(2) *Practising clinical psychology supervisors*. A convenience sample of those who supervised trainees from the Newcastle Doctorate in Clinical Psychology ($n = 30$). These were recruited through two routine supervisor training workshops run by the Doctorate.
(3) *Tutors*. This group comprised clinical tutors (and similarly specialist clinical psychologists) involved in the training of clinical psychologists across the UK ($n = 49$).
(4) *Experts*. Four national experts who conducted training, consultancy and research in clinical supervision ($n = 14$ completed ratings).

Therefore, groups 2–4 were mostly made up of qualified clinical psychologists, both males and females, drawn from the four core specialisms in clinical psychology (i.e. adult mental health, learning disability, child and family, and older adults), with a modal age range of 30–45 years.

Guidelines

Four guidelines were developed which reflected the available empirical and theoretical literature, including the recommendations for CBT supervision in Pretorius (2006) and Townend *et al.* (2002). They were labelled: *Developing the Supervision Contract*

(including collaborative agenda-setting), *Methods of Facilitating Learning* (including making supervision an active process with experiential methods, such as reviewing tapes), *Evaluation in Supervision* (e.g. reviewing one's competence), and the *Supervisory Alliance* (the relational context). Each guideline broadly follows the same NICE-derived format (NICE, 2003):

(1) An introduction that covers the context and scope of the guideline.
(2) Key practice recommendations.
(3) The principles (conceptual rationale) for the detailed practice recommendations that follow. An example of a principle from the Evaluation guideline is that of supervisee involvement, through self-evaluation.
(4) Practice suggestions. This is the core of each guideline, as it lays out the practical steps recommended for the area of supervision covered by the guideline. For example, in the guideline Developing the Supervision Contract the steps are: conducting a needs assessment, setting goals, establishing the learning contract, and reviewing the contract at regular intervals.
(5) A review of the evidence base.
(6) A rating for the strength of the evidence upon which the guideline is built. The rating scheme was also based upon the NICE (2003) system for grading strength of evidence.

Implementation notes. These stress that, for example, the guidelines were designed for use by qualified mental health practitioners and were primarily intended to be compatible with CBT supervision. The guidelines have been published online (see Milne, 2009; they are part of a training manual that can be found at http://bcs.wiley.com/he-bcs/Books?action=index&bcsId=5119&itemId=1405158492).

Guideline Evaluation Form (GEF)

A nine-item GEF was developed especially for the present study, as no suitable instrument was found in the literature. Initial guideline evaluation questions were generated by the GDG. The general evaluation literature was then consulted (e.g. Rossi *et al.* 2004), as was the more specialized literature about the development and evaluation of clinical guidelines (Cluzeau *et al.* 1999; Shekelle *et al.* 1999; Grilli *et al.* 2000; Parry 2000; NICE, 2003; Parry *et al.* 2003). We believe that this gave the GEF content validity. The resulting GEF instrument measures the acceptability of the following dimensions of a guideline: its readability, content, directiveness, relevance, face validity, capacity to empower supervisor and supervisee, capacity to raise standards of supervision, coverage of ethical issues, credibility of the information presented, and capacity to develop capability in supervisors. Responses were invited on a three-point scale: 'Not yet acceptable', 'Acceptable' and 'Good'. A final question asks for qualitative comments to clarify the quantitative ratings, or to suggest improvements (a copy of GEF may be obtained on request from the first author).

Procedure

Following NICE (2003), the steps were first the identification, refinement, and prioritization of topics that the guidelines would address. This iterative process was undertaken by the authors, in conjunction with the GDG. There then followed the collection and interpretation of relevant evidence. This included the literature review of Dunkerley *et al.* (2004) together with

Table 1. *A summary of the acceptability ratings for the four guidelines*

Question	Relationship guideline	Contract guideline	Methods guideline	Evaluation guideline	All guidelines together
1. Easy to read?	2.73 (0.46)	2.14 (0.65)	2.23 (0.68)	2.42 (0.64)	2.38 (0.61)
2. Acceptability of content?	2.73 (0.46)	2.90 (0.30)	2.43 (0.63)	2.78 (0.46)	2.71 (0.46)
3. Acceptably . . .					
(*a*) Directive?	2.50 (0.50)	2.43 (0.68)	2.37 (0.67)	2.56 (0.51)	2.47 (0.59)
(*b*) Relevant?	2.33 (0.62)	2.33 (0.66)	1.87 (0.68)	2.36 (0.57)	2.22 (0.63)
(*c*) Face valid?	2.53 (0.52)	2.67 (0.48)	2.50 (0.63)	2.76 (0.44)	2.62 (0.52)
4. Empowering?	2.53 (0.52)	2.43 (0.60)	2.43 (0.77)	2.75 (0.44)	2.54 (0.58)
5. Standard raising?	2.40 (0.51)	2.57 (0.60)	2.52 (0.58)	2.77 (0.43)	2.57 (0.53)
6. Address ethics?	2.53 (0.52)	1.76 (0.70)	1.76 (0.69)	1.98 (0.44)	2.01 (0.59)
7. Credibility?	2.60 (0.51)	2.48 (0.51)	2.29 (0.60)	2.52 (0.59)	2.47 (0.55)
8. Develop capability in supervisor?	2.20 (0.41)	2.52 (0.51)	2.37 (0.63)	2.50 (0.51)	2.40 (0.52)
Overall mean score for each guideline	2.51 (0.51)	2.42 (0.64)	2.27 (0.69)	2.54 (0.55)	2.44 (0.56)

Values are mean (S.D.)
Rating scale: 1 = Not yet acceptable; 2, Acceptable; 3, Good.

three linked reviews that formed part of the wider research project (Milne, 2009), existing guidelines (e.g. British Psychological Society, 2003), literature specifically addressing the development of clinical guidelines (e.g. Centre for Reviews and Dissemination, 2001; NICE, 2003), popular texts on clinical supervision (e.g. Watkins, 1997), expert consensus statements (Falender *et al.* 2004; Kaslow, *et al.* 2004), and the staff development literature (e.g. Goldstein, 1993). This led to the drafting of four guidelines, each reviewed initially by the GDG. They were then evaluated by the practising supervisors, followed by the expert group (these two groups utilized the GEF). With regard to the supervisors, guidelines were presented at two of a series of routine CPD workshops linked to the Newcastle Doctorate in Clinical Psychology. Supervisors were given a short presentation about the present project and then asked to read the guideline and complete a GEF independently. Tutors were asked to read guidelines and complete the GEF at two seminars held during a national conference of the Group of Trainers in Clinical Psychology. The four experts were sent guidelines and GEF forms through the post. The project was approved by the Research and Development Department of the NHS Trust that employed the authors.

Results

Mean ratings and standard deviations for each GEF question across each guideline are shown in Table 1. It can be seen that ratings for the Supervision Contract guideline were all in the 'acceptable' to 'good' range, apart from question 6 (coverage of ethical issues), which was rated 'not yet acceptable'. Ratings for the Methods guideline were all in the 'acceptable' to 'good' range, apart from question 3(b) (relevance to all orientations and specialisms) and question 6 (coverage of ethical issues), which were both rated 'not yet

acceptable' (i.e. the respondents viewed the guidelines as squarely covering CBT supervision, at the expense of other orientations). Ratings for the Evaluation guideline were all in the 'acceptable' to 'good' range. However, question 6 (coverage of ethical issues) was again rated lowest, indicating the need to improve the treatment of ethical issues. Collapsed together, the ratings for all the guidelines were all in the 'acceptable' to 'good' range, apart from question 6 (coverage of ethical issues), which remained 'not yet acceptable'. The total mean rating for each guideline was in the 'acceptable' to 'good' range. The Methods guideline was rated lowest, while the Evaluation guideline was rated marginally the highest. The total mean rating for all the guidelines collapsed together was in the 'acceptable' to 'good' range.

Statistical tests were performed to determine whether the GEF ratings differed significantly between guidelines. The data were not truly independent, since in some cases participants discussed a guideline in small groups before individually completing GEFs. Therefore, a series of non-parametric Kruskal–Wallis tests were used to compare ratings for each guideline, question by question. A significance level of $\alpha = 0.05$ was adopted. There was a significant difference between guidelines for question 1 ('Was the guideline easy to read?'; $H(2) = 8.81$, $p = 0.03$), and question 2 ('Did the content seem factually accurate?'; $H(2) = 11.43$, $p = 0.01$). There was also a significant difference between guidelines for question 3(b) ('Was the guideline ... relevant to all major theoretical orientations and specialisms?'; $H(2) = 9.80$, $p = 0.02$), and question 6 ('Does the guideline address ethical issues appropriately'; $H(2) = 15.66$, $p = 0.001$). All other differences were non-significant.

Each of these findings was followed up with Mann–Whitney tests. A Bonferroni correction was applied and therefore all *post-hoc* effects are reported at a 0.025 level of significance. Results indicated that the Relationship guideline was rated significantly higher for question 1 (i.e. was judged easier to read) than both the Contract guideline ($U = 81$, $p = 0.013$) and Methods guideline ($U = 134.50$, $p = 0.015$). The Relationship guideline was also rated significantly better for question 6 (ethical issues) than the guidelines for Contract ($U = 68$, $p = 0.003$), Methods ($U = 93$, $p = 0.001$) and Evaluation ($U = 85$, $p = 0.008$). The Methods guideline was rated significantly lower for question 3(b) (relevance) than both Contract ($U = 203.50$, $p = 0.02$) and Evaluation ($U = 232.50$, $p = 0.008$) guidelines; and this guideline was also rated significantly lower for question 2 (content) than the Contract guideline ($U = 185.50$, $p = 0.003$). Overall, the Method guideline was rated significantly lower than the guidelines for Relationship ($U = 18220.5$, p $= 0.002$), Contract ($U = 27150$, $p = 0.02$) and Evaluation ($U = 28893$, $p = 0.0001$). The differences between the Method guideline and the other guidelines represent effect sizes that are conventionally described as 'medium to large' (Cohen, 1988).

In summary, the quantitative (GEF) ratings of the four guidelines indicate that they all fell within the 'acceptable' category, with means ranging from 2.3 to 2.5 (i.e. mid-way between 'acceptable' and 'good'). However, the Methods guideline was rated significantly lower than one or more of the other three, in terms of its readability, limited ethical content, relevance and factual accuracy. Comparisons between the GEF questions suggested that the guidelines as a whole were weakest on tackling ethical aspects of supervision, and best on their factual accuracy. Overall, these findings satisfy the first two conditions for an effective guideline, validity and acceptability (Marriott & Cape, 1995), although it seems that the Methods guideline requires further attention on both counts.

Ordered themes

	FIRST		SECOND	THIRD
1.	Guideline would benefit from being easier to read (less 'dry'; too repetitive; too much jargon-needs explaining). Length & wordiness (make more concise).	[13]	Readability (46%)	
2.	Structure: add a contents section & headings; reduce	[9]		
3.	significant variations across all sections. Would like a clear conclusion/broader summary	[8]		
4.	More (case) examples would be nice.	[6]		
5.	Not always clear if talking about supervisee or	[3]		
6.	supervisor.	[2]		
7.	Specific changes/inclusions: add/more on: 'process', i.e. methods of facilitation; skills teaching; supervision of supervision; research. Case presentation needs to have a richer account. Over-emphasis on teaching specific skills.	[7]	Content (20%)	Feed-back
8.	Not enough on supervisory relationship & what facilitates it (e.g. shift focus on 'being with' as opposed to 'doing to' in supervisory relationship)	[6]		
9.	Supervisees' perspectives: e.g. how to deal with struggling.	[3]		
10.	Perhaps guidelines themselves could be SMARTER?	[1]		
11.	Could do with these references.	[1]		
12.	Address an 'integrative approach' (seemed overly-CBT; & clarify if applies to all professions?).	[10]	Relevance (17%)	
13.	Some adjustment to cover 'research supervision'	[3]		
14.	Content felt a bit divorced from processes I follow	[2]		
15.	More on boundaries & BPS code of conduct.	[5]		
16.	Did not address ethical issues appropriately; needs to be more explicit & a subject of discussion	[5]	Ethics (17%)	
17.	More attention to dealing with diversity.	[1]		
18.	More specifics for new supervisors.	[1]		

Fig. 2. A summary of the qualitative feedback on all four guidelines (from question 9 of the Guideline Evaluation Form; numbers in square brackets represent the frequency of comments).

Qualitative results

Question 9 of the GEF asked for comments, in order to clarify the quantitative ratings or to suggest improvements to the guidelines. We have summarized the 90 comments received within Fig. 2, a content analysis. This lists the main themes that were mentioned, alongside the frequency of each mention. (Note that these are 90 comments, not participants, so several comments may have been made by one participant.) It can be seen that the most frequently noted comment concerned problems over the readability of the guidelines (46% of the comments made), with the suggestion that jargon and repetition could usefully be reduced (a comment made 13 times). The second most common comment was about the content of the guidelines (20% of comments), including suggested additions, especially heightening the emphasis on the supervisory alliance (six comments). Reflecting the GEF data, 17% of comments also identified a need to strengthen the treatment of ethical issues. The same proportion of comments was made about increasing the relevance of the guidelines, so that they might be more applicable to non-CBT supervision, etc. In summary, the guidelines

satisfied the 'essential' criteria established for treatment manuals (a similar type of guidance document), but fell down in terms of the 'desirable' criterion of being applicable to a wide range of approaches (Duncan *et al.* 2004).

Summary and discussion

Our first objective was to develop four guidelines to address the supervision cycle (assessing learning needs and collaborative agenda-setting; facilitating learning; and evaluation) in the context of the supervisory alliance. This was achieved with the help of a local guideline development group, supplemented by advice from four national experts. We attempted to follow established good practice in this process (NICE, 2003; Parry, 2000), including drawing on the best available evidence (including systematic reviews and expert consensus), then seeking feedback on draft guidelines from different stakeholder groups (the supervisors and tutors), leading to refined guidelines. Contrary to our aim of creating integrative guidelines, this feedback indicated that the guidelines clearly represented a CBT approach to supervision. With hindsight, this can be seen as a consequence of adopting an evidence-based, pragmatic strategy for developing the guidelines.

Second, we sought to evaluate the guidelines in terms of their acceptability, including their accuracy. Taken together, the overall rating for all guidelines was firmly in the 'acceptable' range, mid-way to the top available rating of 'good'. Further, the overall rating for each guideline was 'acceptable'. The criterion of accuracy (assessed by the item 'acceptability of content') was met by all four guidelines (mean ratings ranged from 2.43 to 2.90, which approximates to 'good', the top of our three-point rating scale). Individually, the Methods guideline was rated lowest (mean 2.27), and the Evaluation guideline highest (mean 2.54). We found that the Methods guideline was rated significantly lower for content and relevance than the other guidelines, and these differences represented medium-to-large effect sizes. When all guidelines were taken together, every dimension was rated as 'acceptable' apart from coverage of ethical issues, which was rated 'not yet acceptable'.

Critical review

The 'acceptable' ratings for all guidelines suggest that the guidelines are broadly appropriate for use by new CBT supervisors. A similarly positive result was obtained by Neufeldt (1994) in a rare reaction evaluation of a manual for new supervisors. The findings also provide indirect support, in the form of professional consensus, for the Evidence-Based Clinical Supervision model (Milne, 2009). In particular, it supports the presence of all the factors represented in the model and in recommended CBT supervision (Lewis, 2005): needs assessment, learning objectives, methods to facilitate learning, and evaluation, in the context of a supervisory alliance.

The low ratings for the coverage of ethical issues may simply reflect inadequate attention to this topic. Indeed, qualitative responses (Fig. 1) suggested that the complaint was one of omission or scarcity, rather than the inappropriate treatment of ethics. Ethics are of course a crucial aspect of supervision, as indicated by their inclusion in a USA consensus statement regarding the minimal competencies of a clinical supervisor (Falender *et al.* 2004). The guidelines were therefore redrafted, so that each included an explicit section on ethical aspects

of supervision (these revised versions are the ones that are available from the above-noted website).

With regard to the low ratings that the Methods guideline received for content, qualitative information strongly indicated that the major concern was the omission of process issues, e.g. transference. This is an example of the consequence of adopting an evidence-based strategy for the guidelines. Similarly, qualitative responses suggested that the dominant concern about the relevance of the Methods guideline was that it had limited relevance to orientations other than CBT. Both complaints may simply reflect the use of a theoretically divergent group of psychologists. However, they may also stem from a difficulty with the exceptionally broad scope of this guideline. Methods attempted to encompass most of what goes on in most supervision sessions, from case discussion to modelling. Since some methods are largely specific to an orientation (e.g. live supervision to systemic approaches and interpretations to psychodynamic psychotherapy), it may not be possible to write a methods guideline that is acceptable to most orientations, yet sufficiently comprehensive and detailed to be of practical use. It follows that it may make more sense to break the guideline down into many smaller guidelines. These issues are not limited to the Methods guideline. Relevance to all orientations and specialisms was the dimension that received the second lowest rating for all guidelines together. It was second lowest for Evaluation and third lowest for Contract. This is important, given that the guidelines were originally intended to be compatible with most theoretical orientations, and given that they are based on a model that attempts to be relatively pan-theoretical. Again, this broad deficit may reflect an error of omission. But it may also stem from the attempt to make the guidelines evidence-based, an emphasis that generally appeals to CBT adherents, whilst alienating others (Lucock *et al.* 2006; Dyer, 2008). There is generally more published, methodologically sound, empirical evidence for CBT supervision than for other approaches (Milne & James, 2000). A recommended solution is to temper the evidence-based nature of the guidelines with professional consensus (Parry, 2000; DoH, 2001*a*). This was attempted here, but clearly did not go far enough towards achieving consensus.

Three methodological limitations should be noted. First, there were ceiling effects for some GEF questions, which suggests that the response format was too narrow: expanding it from a three-point to a five-point rating scale would probably provide greater sensitivity. Second, the GEF did not define 'acceptable'. Ellis *et al.* (1996) noted the importance of defining constructs in supervision research. 'Acceptable' may have meant different things to different people. However, it is unlikely that interpretations differed so widely as to seriously confound the results obtained. Third, the sample was almost entirely made up of clinical psychologists and included a representative range of theoretical orientations, so it is possible that (for example) a BABCP sample would produce different results, although the presence of three experienced CBT supervisors in the expert group would suggest not. As these guidelines will in any case require periodic updating to reflect developments, it might make sense to undertake that development work with a suitable sample.

Finally, a more general critical reflection is to recognize that guidelines are not necessarily the best way to develop EBP. Iberg (1991) has argued persuasively that simply providing corrective feedback (as part of an ongoing intervention process; i.e. a 'shaping' approach) may be more efficient and can overcome the various problems listed in the Introduction (e.g. weak knowledge base). This empirical emphasis is also congruent with CBT, but arguably the optimal approach is to combine corrective feedback with well-established knowledge, as in the guidelines themselves (Townend *et al.* 2002; Pretorius, 2006).

Implications

Whilst it is conventional to start by evaluating stakeholders' reactions to new interventions (to first ensure acceptability and social validity), future research should evaluate the guidelines at the more exacting levels of learning, transfer and impact (Kirkpatrick, 1967), preferably using the augmented framework developed by Kraiger *et al.* (1993). To illustrate, the extent to which learning is facilitated by these guidelines could be assessed with a quiz; the generalization of any such learning to the workplace could be assessed through supervisees' ratings; and the impact on the service system could be assessed through archival data, such as committee minutes and external audits (Milne, 2007). In a related way, it would be useful to evaluate how the guidelines are rated alongside the other methods that are normally used in a CPD workshop. This would have ecological validity, since the guidelines are written to be used in conjunction with other workshop methods, including didactic presentations and experiential learning exercises (e.g. educational role play).

Conclusion

As far as we are aware, the present project is the first to develop and evaluate guidelines for clinical supervision that are evidence-based, relevant to the NHS, and which follow the NICE procedure. Indeed, although supervision guidelines are recognized as potentially useful, they have rarely been developed or evaluated systematically (Watkins, 1997), unlike CBT manuals more generally (Duncan *et al.* 2004). The consistent approval given to the four supervision guidelines presented above suggests that it is possible to produce evidence-based guidelines for clinical supervision that are valid and acceptable to a range of stakeholders, two necessary conditions for an effective guideline (Marriott & Cape, 1995). The guidelines were deemed suitable for CBT supervision, so future research might advantageously draw on them to enhance the training of CBT supervisors. Use of the guidelines may also help clinicians to meet demands for EBP (DoH, 2001*a*), can support their CPD (DoH, 2001*b*), and might contribute to the recent efforts to research CBT supervision in manualized ways (e.g. Sholomskas *et al.* 2005; Bambling *et al.* 2006).

Recommended follow-up reading

Falender CA, Shafranske EP (2004). *Clinical Supervision: A Competency-Based Model.* Washington DC: American Psychological Society.

References

American Psychological Association (2002). Criteria for practice guideline development and evaluation. *American Psychologist* **57**, 1048–1051.

Baltimore ML, Crutchfield LB (2003). *Clinical Supervisor Training*. Boston, MA: Allyn and Bacon.

Bambling M, King R, Raue P, Schweitzer R, Lambert W (2006). Clinical supervision: its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research* **16**, 317–331.

Boog BWM (2003). The emancipatory character of action research, its history and the present state of the art. *Journal of Community & Applied Social Psychology* **13**, 426–438.

British Psychological Society (2003). *Policy Guidelines on Supervision in the Practice of Clinical Psychology*. London: British Psychological Society.

Cape J, Barkham M (2002). Practice improvement methods: conceptual base, evidence-based research, and practice-based recommendations. *British Journal of Clinical Psychology* **41**, 285–307.

Centre for Reviews and Dissemination (2001). Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews (CRD Report Number 4). York: NHS Centre for Reviews and Dissemination.

Cluzeau FA, Littlejohns P, Grimshaw J, Feder G, Moran SE (1999). Development and application of a generic methodology to assess the quality of clinical guidelines. *International Journal for Quality in Health Care* **11**, 21–28.

Cohen J (1988). *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

DoH (1999). Clinical Governance in the New NHS. London: Department of Health.

DoH (2001*a*). Treatment Choice in Psychological Therapies and Counselling: Evidence Based Clinical Practice Guideline. London: Department of Health.

DoH (2001*b*). Working Together – Learning Together: A Framework for Lifelong Learning for the NHS. London: Department of Health.

DoH (2004). The Ten Essential Shared Capabilities. London: Department of Health.

Duncan EAS, Nicol MM, Ager A (2004). Factors that constitute a good cognitive-behavioural treatment manual: a Delphi study. *Behavioural and Cognitive Psychotherapy* **32**, 199–213.

Dunkerley CJ, Milne DL, Wharton S (2004). A Systematic Review of Evidence-based Clinical Supervision. Presentation given at the annual conference of the European Association for Behavioural and Cognitive Psychotherapy, Manchester, UK.

Dyer K (2008). Practice guidelines in clinical psychology: an uncertain future for a clinical profession. *Clinical Psychology Forum* **190**, 17–20.

Ellis MV, Ladany N (1997). Inferences concerning supervisees and clients in clinical supervision: an integrative review. In: *Handbook of Psychotherapy Supervision* (ed. C. E. Watkins), pp. 447–507. New York: Wiley.

Ellis MV, Ladany N, Krengel M, Schult D (1996). Clinical supervision research from 1981 to 1993: a methodological critique. *Journal of Counseling Psychology* **43**, 35–50.

Falender CA, Cornish JAE, Goodyear R, Hatcher R, Kaslow NJ, Leventhal G, Shafranske E, Sigmon ST, Stoltenburg C, Grous C (2004). Defining competencies in psychology supervision: a consensus statement. *Journal of Clinical Psychology* **60**, 771–785.

Fall M, Sutton JM (2004). *Clinical Supervision: A Handbook for Practitioners*. Boston, MA: Allyn and Bacon.

Goldstein IL (1993). *Training in Organizations*. Pacific Grove, CA: Brooks/Cole.

Green D, Youngson S (2005). *DCP Policy on Continued Supervision*. Leicester: British Psychological Society.

Grilli R, Magrini N, Penna A, Mura G, Liberati A (2000). Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* **355**, 103–106.

Henggeler SW, Schoenwald SK, Liao JG, Letourneau EJ, Edwards DL (2002). Transporting efficacious treatments to field settings: the link between supervisory practices and therapist fidelity in MST programmes. *Journal of Clinical Child Psychology* **31**, 155–167.

Iberg JR (1991). Applying statistical control theory to bring together clinical supervision and psychotherapy research. *Journal of Consulting and Clinical Psychology* **59**, 575–586.

Kaslow NJ, Borden KA, Collins FL, Forrest L, Illfelder-Kaye J, Nelson PD, Vasquez MJ, Willmuth MEL (2004). Competencies conference: future directions in education and credentialing in professional psychology. *Journal of Clinical Psychology* **60**, 699–712.

Kilminster SM, Jolly BC (2000). Effective supervision in clinical practice settings: a literature review. *Medical Education* **34**, 827–840.

Kirkpatrick DL (1967). Evaluation of training. In: *Training and Development Handbook* (ed. R. L. Craig and L. R. Bittel), pp. 87–112. New York: McGraw- Hill.

Kirschner PA, Sweller J, Clark RE (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist* **41**, 75–86.

Kolb DA (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.

Kraiger K, Ford JK, Salas S (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology* **78**, 311–328.

Ladany N, Brittan-Powell CS, Pannu RK (1997). The influence of supervisory racial identity interaction and racial matching on the supervisory working alliance and supervisee multicultural competence. *Counselor Education and Supervision* **36**, 284–304.

Lewis K (2005). The supervision of cognitive and behavioural psychotherapists. *BABCP Magazine Supplement* (May 2005). Accrington: BABCP.

Lucock MP, Hall P, Noble R (2006). A survey of influences on the practice of psychotherapists and clinical psychologists in training in the UK. *Clinical Psychology and Psychotherapy* **13**, 123–130.

Marriott S, Cape J (1995). Clinical practice guidelines for clinical psychologists. *Clinical Psychology Forum* **81**, 2–6.

Milne DL (2007). Evaluation of staff development: the essential 'SCOPPE'. *Journal of Mental Health* **16**, 389–400.

Milne DL (2008). CBT supervision: from reflexivity to specialisation. *Behavioural and Cognitive Psychotherapy* **36**, 779–786.

Milne DL (2009). *Evidence-Based Clinical Supervision*. Chichester: BPS/Blackwell.

Milne D, James IA (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology* **39**, 111–127.

Milne D, James IA (2005). Clinical supervision: ten tests of the tandem model. *Clinical Psychology Forum* **151**, 6–9.

NICE (2003). Guideline Development Methods: Information for National Collaborating Centres and Guideline Developers. London: National Institute for Clinical Excellence.

Neufeldt SA (1994). Use of a manual to train supervisors. *Counselor Education and Supervision* **33**, 327–336.

Northcott N (1998). The development of guidelines on clinical supervision in clinical practice settings. In: *Clinical Supervision in Practice* (ed. V. Bishop), pp.109–142. London: Macmillan.

Parry G (2000). Developing treatment choice guidelines in psychotherapy. *Journal of Mental Health* **9**, 273–281.

Parry G, Cape J, Pilling S (2003). Clinical practice guidelines in clinical psychology and psychotherapy. *Clinical Psychology and Psychotherapy* **10**, 337–351.

Perkins JM, Mercaitis PA (1995). A guide for supervisors and students in clinical practicum. *The Clinical Supervisor* **13**, 67–78.

Pretorius WM (2006). Cognitive behavioural therapy supervision: recommended practice. *Behavioural and Cognitive Psychotherapy* **34**, 413–420.

Rossi PH, Freeman HE, Lipsey MW (2004). *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage.

Shekelle PG, Woolf SH, Eccles M, Grimshaw J (1999). Developing guidelines. *British Medical Journal* **318**, 593–596.

Sholomskas DE, Syracuse-Siewert G, Rounsaville BJ, Ball SA, Nuro KF, Carroll KM (2005). We don't train in vain: a dissemination trial of three strategies of training clinicians in CBT. *Journal of Consulting and Clinical Psychology* **73**, 106–115.

Townend M, Iannetta L, Freeston MH (2002). Clinical supervision in practice: a survey of UK cognitive-behavioural psychotherapists accredited by the BABCP. *Behavioural & Cognitive Psychotherapy* **30**, 485–500.

Watkins CE (ed.). (1997). *Handbook of Psychotherapy Supervision.* Chichester: Wiley.

Learning objectives

By studying this paper, readers will be able to:

(1) Summarize the role of guidelines within evidence-based practice.
(2) Describe the four presented supervision guidelines.
(3) Note three strengths and three weaknesses regarding the evaluation of these guidelines.