Feature Article
Empirical studies of clinical supervision in psychiatric nursing: A systematic literature review and methodological critique

Niels Buus¹ and Henrik Gonge²
¹Institute of Public Health, University of Southern Denmark, Odense, and ²Århus University Hospital, Risskov, Århus, Denmark

ABSTRACT: The objective of this paper was to systematically review and critically evaluate all English language research papers reporting empirical studies of clinical supervision in psychiatric nursing. The first part of the search strategy was a combination of brief and building block strategies in the PubMed, CINAHL, and PsycINFO databases. The second part was a citation pearl growing strategy with reviews of 179 reference lists. In total, the search strategy demonstrated a low level of precision and a high level of recall. Thirty four articles met the criteria of the review and were systematically evaluated using three checklists. The findings were summarized by using a new checklist with nine overall questions regarding the studies’ design, methods, findings, and limitations. The studies were categorized as: (i) effect studies; (ii) survey studies; (iii) interview studies; and (iv) case studies. In general, the studies were relatively small scale; they used relatively new and basic methods for data collection and analysis, and rarely included sufficient strategies for identifying confounding factors or how the researchers’ preconceptions influenced the analyses. Empirical research of clinical supervision in psychiatric nursing was characterized by a basic lack of agreement about which models and instruments to use. Challenges and recommendations for future research are discussed. Clinical supervision in psychiatric nursing was commonly perceived as a good thing, but there was limited empirical evidence supporting this claim.

KEY WORDS: clinical supervision, mental health, psychiatric nursing, systematic review.

INTRODUCTION
There is a substantive body of literature on clinical supervision in psychiatric nursing. However, the literature on clinical supervision has been marked by several conflict-ridden issues on the theory, practice, and research of clinical supervision. Most notably, there has been no general consensus on a definition of clinical supervision. Relationships and roles, tasks, theories and methods, goals, and contextual influences continue to be at issue in scholarly debates (Hytkas et al. 1999; Jones 2006).

The variety in the definitions is probably related to the historical emergence of clinical supervision in psychiatric nursing in the latter half of the 20th century. In this period of time, inherited institutional practices of control and discipline merged with new ideas on training and education (Kelly et al. 2001), following in the wake of nurses’ increased professional autonomy (Butterworth 1992), and with a more widespread and increasing concern about the strain effects related to working with people who are mentally distressed. Thus, the emerging practices of ‘clinical supervision’ in psychiatric nursing...
transformed the traditional controlling practices of ‘super-vision’ and subsumed several new institutional and professional issues, such as reflection, education, and recreation. The term has since been defined and redefined by scholars and clinicians to fulfill several professional and educational aims, which has further added to the congested meaning of the term.

Even after accepting that the term ‘clinical supervision’ may be the only common denominator between papers on the issue, the literature on clinical supervision in psychiatric nursing is not easy to summarize and evaluate. Arguments and reviews are often constrained by referencing a mixture of comments, theoretical discussions, and empirical studies from both psychiatric and general nursing. This mixture of theoretical and empirical references may add to the impression that our knowledge of clinical supervision in psychiatric nursing is more robust and trustworthy than it in fact is. Moreover, clinical supervision in psychiatric nursing may well appear and work differently compared to clinical supervision in other nursing specialities. This is because interpersonal and psychosocial interventions are all important and fundamental in psychiatric nursing, and stress experienced in relation to learning and practicing these interventions to a large extent is different from the stress experienced in other specialities. Thus, research findings are probably not directly transferable between psychiatric nursing and other nursing specialities (Yegdich & Cushing 1998).

In this paper, we focus on the empirical research papers of clinical supervision of nurses in psychiatric settings regardless of study designs and operationalizations of the term. Hitherto, reviews of clinical supervision in psychiatric nursing have not been systematic and exhaustive. The aim of the following systematic literature review was to summarize and critically evaluate all empirical studies of clinical supervision in psychiatric nursing and to identify and discuss issues that would benefit from additional research in the future.

MATERIALS AND METHODS

Two overall strategies were used to search for empirical research papers on clinical supervision in psychiatric nursing: (i) a combination of ‘brief’ and ‘building block’ search strategies (searching databases); and (ii) a citation pearl growing strategy (reviewing reference lists) (Harter 1986). This particular combination of search strategies demonstrated a low level of ‘precision’ and a high level of ‘recall’ (Harter 1986).

The combined ‘brief’ and ‘building block’ search strategy was used to identify references in the following databases: CINAHL, PubMed, and PsycINFO. These three databases were chosen for the search because a wide range of potentially relevant journals was indexed in these databases. The searches for literature took place in November 2007.

CINAHL’s thesaurus included the controlled heading ‘clinical supervision, mental health’, and this brief search identified 89 references. Furthermore, and surprisingly, the combination of the two controlled brief searches ‘clinical supervision’ and ‘psychiatric nursing’ identified 44 references hardly overlapping the references from the first search.

PubMed’s controlled headings (MeSH terms) did not include ‘clinical supervision’. Two brief free text searches ‘clinical supervision’ and ‘nurs’ were combined with a block of controlled headings: ‘Psychiatric Nursing’ or ‘Mental Health Services’. This search identified 114 references.

PsycINFO’s controlled heading ‘Professional Supervision’ was combined with a block of controlled headings: ‘Psychiatric Hospital Staff’ or ‘Psychiatric Nurses’. This search identified 29 references in peer-reviewed journals.

There was a substantial overlap between the references identified in the different databases, and in order to minimize errors in the search and review process, the references were handled with the software program Reference Manager (Thomson Reuters, New York, NY, USA). The first author reviewed the abstracts or full-text articles of all the references identified in the database search.

All the relevant references identified using the building block strategy were used as the outset for a laborious and comprehensive ‘citation pearl growing strategy’ (Harter 1986). Reference lists from the articles were reviewed to identify references that were not included in the three databases (e.g. grey literature) or references that were not indexed correctly in the databases. The reference lists of 179 full-text articles were examined as part of the citation pearl growing strategy.

The review’s inclusion criterion was:

English language research papers reporting empirical studies of clinical supervision in psychiatric nursing.

The review’s exclusion criteria were:

Papers that did not explicitly aim at studying clinical supervision in psychiatric nursing, but dealt with it indirectly; for example, studies of occupational stress that conclude that clinical supervision may reduce stress

Brief research reports and reports from ongoing research
<table>
<thead>
<tr>
<th>Study</th>
<th>Aim</th>
<th>Design (points of measuring)</th>
<th>Method/instruments</th>
<th>Setting and participants</th>
<th>Analyses</th>
<th>Result</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hallberg (1994)</td>
<td>To investigate how nurses perceive the impact of systematic CS and the effects of CS on nurses’ degree of burnout, boredom, and job satisfaction</td>
<td>Before and after study with no control (baseline, 6 and 12 months)</td>
<td>An open-ended interview guide (only at 6 and 12 months). TM, MBI. A Swedish instrument for measuring satisfaction with nursing care (SNCW).</td>
<td>A child psychiatric ward in Sweden. 9 beds for children aged 9-16 years. 11 nurses (all nurses available on the ward)</td>
<td>Open coding of qualitative data. ANOVA</td>
<td>The nurses felt more understood by other professionals and understood other professionals better. The nurses felt that their knowledge base had changed and broadened. There were some significant improvements of burnout (1/3 TM subscales) and of satisfaction with nursing care (2/8 factors). All other findings were not significant</td>
<td>Small sample size, no control group, no control of confounding variables, the rate of CS participation was not reported, supervisor and researcher were identical, and one of the instruments was new and not validated or described in sufficient detail. The intervention included both supervision and supervised individually planned care</td>
</tr>
<tr>
<td>Berg and Hallberg (1999) (Part of the same project as Berg and Hallberg (2000))</td>
<td>To explore the effects of 1 year of systematic clinical supervision and supervised planning of individual care on nurses’ sense of coherence, creativity, work-related strain, and job satisfaction</td>
<td>Before and after study with no control (baseline, 6 and 12 months)</td>
<td>SOC, CCQ, WRSI, SNCW (Welander-Hansson et al. 1995). 14 additional statements</td>
<td>A general psychiatric ward in Sweden. 36 beds. 22 nurses (all nurses on the ward). 18 participated throughout the study</td>
<td>ANOVA. Correlation analyses. Wilcoxon rank sum test. Factor analyses</td>
<td>There were significant improvements of the CCQ (3/10 subscales). The nurses’ view on the effects of CS (14 statements) became significantly more positive. All other findings were not significant</td>
<td>Small sample size. No control ward. No control of confounding variables. Supervisor and researcher were identical. SNCW was new and not validated or described in sufficient detail. The scores of the 14 statements were summarized without explanation. The intervention included both supervision and supervised individually planned care</td>
</tr>
</tbody>
</table>
To evaluate the impact of 1 year of clinical supervision and individually planned nursing care on nurses' experiences (Berg et al. 1994; Edberg et al. 2006, 1999; Hallberg and Norberg 1993; Hallberg et al. 1994b). Part of the same project as (Holst et al. 1999).

A controlled before-and-after study (baseline: June and December 2001 and 2002) with a follow-up after 12 months. The groups differed significantly at baseline.


ANCOVA

A. Group (6-8 nurses: the staff on duty). B. 1.5–2 h every third week during the first 6 months; 2 h every second week during the last 6 months. A total of 30 h. C. Reflective discussions of a patient with a focus on emotions, views on the patient and actions towards the patient (Hallberg & Norberg 1993). D. RN and long experience and training in psychiatric care. E. External interactions with patients (Edberg et al. 1996) and the health status of patients (Edberg et al. 1999).

1. The climate was perceived more creative and innovative on the EW (4/10 subscales on the CCQ, one of these differed significantly at baseline). Neither BMI nor MBI indicated any differences between EW and CW at follow up (Berg et al. 1994).

2. This study reports no comparisons between EW and CW. Based on a before and after comparison, the nurses on the EW felt increased satisfaction with several aspects of their work and the care they provided (Hallberg et al. 1994b).

3. The nurses on EW found patients to be more cooperative (2/44 factors identified in the SNC) and nurses on the EW found it easier to handle the patient characteristics agony and obedience (2/6 factors identified in the SNC). There were no significant changes between the wards in the 4 factors identified in the ERNC (Hallberg & Norberg 1990).

4. On the EW, the patients were less sad and more cooperative (2/44 variables). Nurses on the CW reported demanding physical behaviour as more constantly present. There were no significant changes in the reports of verbal behaviours (Edberg & Hallberg 2001).

5. The cooperation style between the wards changed. On EW patients' resistance decreased and mutual activities increased. On CW the opposite changes were observed. Among the surviving patients, the nurse-patient cooperation was better on the EW (Edberg et al. 1996).

6. There were no differences in patients' cognitive functions (MMSE and the OBS orientation subscale) and behaviour (GRAS). There were significant differences (baseline-12 months) in 1/7 OBS subscales: there was a deterioration of the patients' speech performance on CW. At the CW the patients' mood decreased (1/3 factors identified in the PMAS) (Edberg et al. 1999).

Notes taken during non-participatory observations of morning care (Hallberg et al. 1995; Edberg et al. 1999).

The climate was perceived more creative and innovative on the EW (4/10 subscales on the CCQ, one of these differed significantly at baseline). Neither BMI nor MBI indicated any differences between EW and CW at follow up (Berg et al. 1994).

This study reports no comparisons between EW and CW. Based on a before and after comparison, the nurses on the EW felt increased satisfaction with several aspects of their work and the care they provided (Hallberg et al. 1994b).

The nurses on EW found patients to be more cooperative (2/44 factors identified in the SNC) and nurses on the EW found it easier to handle the patient characteristics agony and obedience (2/6 factors identified in the SNC). There were no significant changes between the wards in the 4 factors identified in the ERNC (Hallberg & Norberg 1990).

On the EW, the patients were less sad and more cooperative (2/44 variables). Nurses on the CW reported demanding physical behaviour as more constantly present. There were no significant changes in the reports of verbal behaviours (Edberg & Hallberg 2001).

The cooperation style between the wards changed. On EW patients' resistance decreased and mutual activities increased. On CW the opposite changes were observed. Among the surviving patients, the nurse-patient cooperation was better on the EW (Edberg et al. 1996).

There were no differences in patients' cognitive functions (MMSE and the OBS orientation subscale) and behaviour (GRAS). There were significant differences (baseline-12 months) in 1/7 OBS subscales: there was a deterioration of the patients' speech performance on CW. At the CW the patients' mood decreased (1/3 factors identified in the PMAS) (Edberg et al. 1999).

The climate was perceived more creative and innovative on the EW (4/10 subscales on the CCQ, one of these differed significantly at baseline). Neither BMI nor MBI indicated any differences between EW and CW at follow up (Berg et al. 1994).

This study reports no comparisons between EW and CW. Based on a before and after comparison, the nurses on the EW felt increased satisfaction with several aspects of their work and the care they provided (Hallberg et al. 1994b).

The nurses on EW found patients to be more cooperative (2/44 factors identified in the SNC) and nurses on the EW found it easier to handle the patient characteristics agony and obedience (2/6 factors identified in the SNC). There were no significant changes between the wards in the 4 factors identified in the ERNC (Hallberg & Norberg 1990).

On the EW, the patients were less sad and more cooperative (2/44 variables). Nurses on the CW reported demanding physical behaviour as more constantly present. There were no significant changes in the reports of verbal behaviours (Edberg & Hallberg 2001).

The cooperation style between the wards changed. On EW patients' resistance decreased and mutual activities increased. On CW the opposite changes were observed. Among the surviving patients, the nurse-patient cooperation was better on the EW (Edberg et al. 1996).

There were no differences in patients' cognitive functions (MMSE and the OBS orientation subscale) and behaviour (GRAS). There were significant differences (baseline-12 months) in 1/7 OBS subscales: there was a deterioration of the patients' speech performance on CW. At the CW the patients' mood decreased (1/3 factors identified in the PMAS) (Edberg et al. 1999).
Papers concerning clinical supervision of student nurses or professional groups other than psychiatric nurses, including psychiatric nurses mixed with other professionals

Articles on practice development

Articles with insufficient accounts of research method. This criterion was used in three instances after an additional independent review by the second author. Both authors found that the accounts of method were inadequate for a meaningful reconstruction of the method.

In total, 34 articles were eligible and thus reviewed. The authors reviewed the articles systematically using specialized checklists from CONSORT (for effect studies; http://www.consort-statement.org.), STROBE (for survey studies; http://www.strobe-statement.org.), and COREQ (for qualitative data analysis) (Tong et al. 2007). The research methods varied significantly in the 34 papers and it was not feasible to sum up the results in a meta-analysis or metasynthesis. Instead, we summarized the articles using a checklist with nine items designed for this particular review. The checklist included generic questions about the particular research method (aim, design, data collection, analysis, and findings) and the particular supervision under research (context, participants, and method). The checklist comprised of the following items:

1. What was the aim of the study?
2. What was the research design?
3. Which methods were used to collect data?
4. In which type of setting did the study take place?
5. Who were the participants (subjects of research)?
6. How were data analyzed?
7. The type of supervision? (i) form (e.g. individuals, groups); (ii) scope (e.g. frequency and/or duration); (iii) supervisors’ education/training; (iv) methodological/theoretical frameworks for supervision; (v) supervisors’ association with the clinic (e.g. internal or external).
8. What were the key findings?
9. What were the limitations of the conclusions reached in the study?

In the review, we have emphasized the most fundamental limitations of the each study with regards to conclusions, design, and the study’s practical accomplishment. Both authors reviewed the papers independently and discussed their findings until consensus was reached.

RESULTS

The papers were divided into four categories defined by research designs and/or methods for data collection.

Several papers reported findings from the same projects: The 34 papers reported from 25 empirical projects. The larger amount of papers added to an impression that the scope of empirical studies on clinical supervision was wider than it was in reality. To avoid a further contribution to this biased impression, we will mention when papers present data originating from the same research projects. In situations where it is relevant, we will present and review these papers together in the tables.

The categories included: (i) effect studies (n = 9). These studies had a (quasi-)experimental ‘before and after’ design and were characterized as having a well-defined clinical supervision intervention; (ii) survey studies (n = 12). This category included cross-sectional surveys using questionnaires and the application of statistical analyses; (iii) interview studies (n = 6). Studies in this category were characterized by the use of semistructured interviews to collect data and by the use of an interpretative approach; (iv) case studies (n = 7). This relatively broad category included interpretative analyses of data generated during clinical supervision sessions or collected during fieldwork.

We have summarized the papers in four tables in line with our checklist and the four categories of papers. We have omitted answers to some of the questions in the tables below, as the answers were redundant or irrelevant in relation to some of the categories of studies.

Effect studies

The reported findings from the four projects designed to measure the effect of clinical supervision did not provide convincing empirical evidence to support the assumption that clinical supervision in psychiatric nursing settings had an effect on the nurses and/or the patients in their care (see Table 1). The three primary reasons for reaching this conclusion were related to the projects’ results and design: (i) there was a relatively limited number of significant changes following the introduction of the clinical supervision intervention schemes; (ii) only two projects (Berg et al. 1994; Bradshaw et al. 2007; Edberg & Hallberg 2001; Edberg et al. 1996; 1999; Hallberg & Norberg 1993; Hallberg et al. 1994b) had controlled designs and none randomized the subjects of the research. Furthermore, none of the studies included multivariate statistical analyses and/or other controls for confounding variables that may have influenced the outcome measures; and (iii) the intervention schemes in three of the four projects included a clinical supervision intervention as well as an intervention on ‘individually planned care’ (Berg & Hallberg 1999; Berg et al. 1994; Edberg & Hallberg 2001; Edberg et al. 1996; 1999; Hallberg 1994; Hallberg &
Norberg 1993; Hallberg et al. 1994b). This made it impossible to conclude that clinical supervision alone had an effect on the outcome measures.

Two secondary reasons for the conclusion were related to the quality of data and the data analyses: (i) the roles of supervisor, supervisee, researcher, and data collector coincided in several ways. This may have biased measurements in the preferred direction; (ii) all studies included one or more instruments for data collection that were new and not validated. In two of the four projects (Berg & Hallberg 1999; Berg et al. 1994; Edberg & Hallberg 2001; Edberg et al. 1996; 1999; Hallberg & Norberg 1993; Hallberg et al. 1994b), conclusions were based on the identification of factors in the dataset and it was very difficult to grasp what the underlying phenomena (factors) referred to and what the results indicated. Thus, the instruments had no normative scores to compare results with and it was impossible to compare results between the different projects.

Not only were the significant effects of clinical supervision limited, the particular quasi-experimental designs and the basic statistical analysis may have prompted the acceptance of statistical differences when there were in fact none (type I errors). In other words, the observed effects of the clinical supervision interventions were limited, and the identified effects have not been challenged sufficiently to determine if they were in fact spurious.

Survey studies
The reported findings from the 12 cross-sectional surveys provided some indications about the prevalence of clinical supervision among psychiatric nurses and nurses’ perceived benefits and attitudes towards clinical supervision (see Table 2). The ratio of respondents participating in clinical supervision was presented in six projects and ranged from 33% to 81% with a median of 73% (Edwards et al. 2005; 2006; Jones & Bennet 1999; Kelly et al. 2001; Magnusson et al. 2002; Rask & Levander 2002; White & Roche 2006). The relatively wide range was probably caused by the particular way the research was carried out, but may also reflect differences in the prevalence of clinical supervision. A positive attitude towards clinical supervision was reported by staff as well as management. Various benefits of clinical supervision were reported; however, the ratio of positive findings was not convincing. In several studies, the hypotheses, discussions, and/or conclusions included more or less explicit assumptions about causal relations between clinical supervision and other variables. These assumptions were invalid as cross-sectional data do not allow interpretations regarding causality.

A number of methodological problems and flaws were identified regarding the survey studies’ design, instruments, response rates, and the statistics applied. First, only cross-sectional studies were reported. Longitudinal designs with collection of data at different points in time allowing analyses of development or process were not applied. Second, in six studies the questionnaires were developed locally and they suffered from inadequate validity and reliability testing (Magnusson et al. 2002; Pesut & Williams 1990; Rice et al. 2007; Severinson & Hallberg 1996; Veeramah 2002; White & Roche 2006). In three studies the questionnaires were modified versions of relatively new and untested instruments (Major’s attitudes to Clinical Supervision Scale and Satisfaction with Nursing Care and Work scale (SNCW) (Jones & Bennet 1999; Kelly et al. 2001; Rask & Levander 2002). Only two projects (three papers) applied internationally recognized and validated instruments (Manchester Clinical Supervision Scale (MCSS) and Maslach Burnout Inventory (MBI)) (Edwards et al. 2005; 2006; Hyrkas 2005). These latter instruments differed from the former by collapsing several items into indexes. This provided more rigid measurements of the phenomena in question compared to single item measures and measurements based on factor analyses. Third, the response rate could not be retrieved in two studies (Hyrkas 2005; Pesut & Williams 1990) and was very low (<50%) in two additional projects (three papers) (Edwards et al. 2005; 2006; White & Roche 2006). None of the surveys included examinations of non-respondents or comparisons between the sample and non-respondents. Such additional analyses could determine if a study was addressing a biased sample of the population. Fourth, five studies were based on descriptive statistics only (Jones & Bennet 1999; Kelly et al. 2001; Rice et al. 2007; Veeramah 2002; White & Roche 2006), and the application of more advanced statistics was sparse in the remaining studies. All the significant associations identified in the studies relied on univariate analyses ignoring the possible effect of relevant confounding variables which might have been examined through multivariate analyses.

Interview studies
The interview studies were difficult to compare and sum up because they drew on different, more or less explicit theories about human experience and aimed at creating different types of understanding. Two of the five interview projects were concerned with developing an understanding of the meaning of concepts related to clinical supervision (Arvidsson et al. 2000; 2001; Berg & Hallberg 2000) and three projects were concerned with description...
<table>
<thead>
<tr>
<th>Study</th>
<th>Aim</th>
<th>Method/instruments</th>
<th>Setting and participants</th>
<th>Analyses</th>
<th>Results</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edwards et al. (2005); Edwards et al. (2006)</td>
<td>To establish which factors influence the effectiveness of CS and how CS influences burnout</td>
<td>MCSS, MBI, A demographic questionnaire</td>
<td>260 community mental health nurses in Wales; Response rates: 260/817 = 32%; MBI: 26%; MCSS: 23%</td>
<td>Mann-Whitney U-test, ANOVA, Correlation analyses</td>
<td>A. 73% received one-to-one supervision, 17% received supervision in groups, and 8% a combination. B. 57% of the respondents had monthly sessions. Sessions lasted 46–60 min (44%) and &gt;60 min (32%). C. = D. = E. =</td>
<td>Low response rates. No examination of non-respondents. No control of confounding variables</td>
</tr>
<tr>
<td>Severinson and Hallberg (1996)</td>
<td>To analyze nurses' view of the effect of CS, their experiences of working milieu and their influences on their duties</td>
<td>An instrument with 62 items measuring effect of supervision, work environment, and influences on duties</td>
<td>2 acute wards in a psychiatric clinic in Sweden; All 26 nurses on the wards</td>
<td>Factor analysis, Mann-Whitney U-test, Correlation analyses</td>
<td>A. Groups of 4 supervisees and a supervisor. B. Weekly for 1.5 h. C. 2 registered nurses with 2 years' special training in CS. D. Supervisor raises questions about the patients' situation based on Kate Eriksson's theory about caring. The aim was to start reflections on the professional role and the caring situation. E. =</td>
<td>The effect of CS showed no relation to how the working milieu or the possibilities of influencing duties were experienced. Small sample size. No control of confounding variables. The instrument was new and not validated</td>
</tr>
<tr>
<td>Kelly et al. (2003)</td>
<td>To establish the degree to which CS occurs in Northern Ireland (methods, attitudes, and standardization)</td>
<td>A modified version of Major's attitude towards CS scale (20 items) (Major 1993). Data on demographics and CS practices</td>
<td>151 staff members in 10 (of 11) trusts delivering community mental health in Northern Ireland; Response rate: 151/255 = 61%</td>
<td>Descriptive statistics</td>
<td>A. 74% received individual CS, 7% in groups, and 19% a mixture of these types. B. 57% of the CS occurred every 4–5 weeks and 27.4% with an irregular frequency. C. 37% of the participants had received training in supervising. D. = E. =</td>
<td>124 (53%) respondents were involved in CS and there was broad support for CS. Respondents found that managers were not the best supervisors. No analyses of non-respondents. Dataset was divided into very small subgroups for further comparisons</td>
</tr>
<tr>
<td>Rice et al. (2007)</td>
<td>To explore ways to make CS available to all mental health nurses</td>
<td>Questionnaire with 7 questions regarding CS in the respondents' organization</td>
<td>Mental health services in Northern Ireland; 12 mental health trust directors and all heads of education</td>
<td>Summaries of responses</td>
<td>–</td>
<td>CS was supported by trust directors and a list of recommendations was formulated based on the statements of intent. The analyses were not described in detail</td>
</tr>
<tr>
<td>Hyrkas (2005)</td>
<td>To explore and evaluate the current state of CS and ascertain the levels of stress, burnout and job satisfaction</td>
<td>Demographics, MCSS, MBI, short version of MJSS</td>
<td>500 mental health nurses at 12 sites in Finland (14 sites were invited; Response rate: unclear)</td>
<td>Comparisons to normative scores χ²-tests</td>
<td>A. 64% received one-to-one CS, 36% group CS. B. CS was most likely to occur every 3–4 weeks (65%) and up to 60 min (66%). C. 47% of supervisors were specialized psychiatric nurses, and 28% psychologists. D. = E. 64% responded that the supervisor belonged to the same organization as the respondent</td>
<td>There were significant relationships between MCSS scores and burnout (2/3 MBI subscales: de-personalization and personal accomplishment) and job satisfaction (total job satisfaction and 1/2 MJSS subscales: intrinsic). The representativeness of the participants was unknown, including non-respondents. Some of the statistical methods were not reported. No analysis of confounders</td>
</tr>
</tbody>
</table>
Veenmah (2002) To explore the use and perceived benefits of CS
A new questionnaire 105 community mental health nurses working within 2 trusts in south-east England. Response rate: 105/165 = 64%
Descriptive statistics
A. 80% received one-to-one CS, 6% group CS, and 12% a mixture of these. B. Sessions were most likely to occur monthly (60%) for 30–120 min (98%). C. 79% of 52 supervisors had received some form of training. D. 52% reported using eclectic approaches. E. –
The most frequently reported goal of CS was 'To assist in developing clinical competency'. The most frequently reported benefit was: 'It helps practitioners to reflect on their dilemmas, difficulties and successes and how they reacted to, solved or achieved them'.

Pesut and Williams (1990) To survey clinical nurse specialists' experiences with and attitudes towards CS
A new questionnaire: NCSQ (54 items). Demographic questions 61 volunteering clinical nurse mental health specialists at a conference in the USA. Response rate: unclear
C²-tests t-tests Factor analyses
A. Majority of respondents reported other individual CS or individual CS in combination with another modality. B. –. C. –. D. –. E. –
66% believed that supervisors should be certified. 63% believed that themes of the therapy relationship paralleled themes of the supervisory relationship. 65% believed that graduate programmes do not provide enough preparation for the supervisory role.
The instrument was new, not described, and not validated. Applied statistics were not always described. Sampling was haphazard.

White and Roche (2006) To review practices and experiences among mental health nurses in particular in relation to CS
A new individual questionnaire: A new area mental health services questionnaire 601 Australian mental health nurses and 17 Area Health Services. Response rate (median Area Health Service): 21%
Descriptive statistics Comparisons with other datasets
A. –. B. CS was most likely to be on a fortnightly or monthly basis for 1–2 h. C. –. D. Three-quarters of the nurses receiving CS did not know which model was used. E. –
33% of respondents received CS. 51% of the nurses responding that they received CS. There were significant differences in perceptions of ethical issues between the group of nurses receiving CS and those who did not in 10/20 items.
The instruments were new, not described, and not validated. Sampling was haphazard. Very low response rate.

Magunsson et al. (2002) To examine the perceived influence of CS
A new questionnaire (20 items). Data on demographics and CS practices 660 nurses, district nurses and mental health care workers in Sweden. Response rate: 660/1139 = 58%
C²-tests – 51% of the nurses responded that they received CS. There were significant differences in perceptions of ethical issues between the group of nurses receiving CS and those who did not in 10/20 items
The instrument was new and not validated. No control for confounding variables.

Rask and Levander (2002) To investigate nurses' satisfaction with nursing care and work
A modified SNCW (30 items) 246 nursing care staff in 5 Swedish psychiatric forensic units. Response rate: 246/550 = 70%
t-tests Factor analyses ANOVA Correlation analysis Mann–Whitney U-test
A. –. B. 27% (out of 75% receiving CS) received CS weekly, 35% once a fortnight, and 19% once a month. C. 44% were educated as nurse and 33% as psychologist. D. –. E. –
75% of respondents received CS. The most common perceived focus of CS was 'patients' mental problems' and how to interact with the patient'. FOX of CS (346 items) were correlated with job satisfaction factors (3/7)
No control of confounding variables. Variables were not totally independent.

Jones and Bennett (1999) To examine views and practices of CS
A modified version of Major's attitudes towards CS scale (24 items) (Major 1993). 2 focus groups 57 nurses in acute psychiatry in England. Response rate: 56/80 = 70%
Descriptive statistics Content analysis
A. One-to-one CS. B. –. C. –. D. –. E. –
73% of the respondents received CS. There was an overall positive attitude towards CS.
No control of confounding variables.

CS, clinical supervision; MBI, Maslach Burnout Inventory; MCSS, Manchester Clinical Supervision Scale; MJSS, Minnesota Job Satisfaction Scale; NCSQ, Nursing Clinical Supervision Questionnaire; SNCW, Satisfaction with Nursing Care and Work Scale.
and evaluations of practices related to clinical supervision (Olofsson 2005; Scanlon & Weir 1997; Sloan 1999) (See Table 3). The studies indicated that the psychiatric nurses generally found clinical supervision beneficial and facilitating, but that nurses also could have contradictory experiences and attitudes towards clinical supervision. These contradictions came forth because participation in supervisory sessions could be personally challenging and stressful.

However, the confirmability/dependability (reliability/ objectivity in ‘quantitative’ terms) (Miles & Huberman 1994) of the studies supporting this conclusion was questionable because of incomplete descriptions of the methods used to generate and analyze text. The most questionable issue was related to the interpretative processes in the analyses, in particular the effect the researchers’ interpreters’ preconceptions had on the interpretative process. First, data were generated through a dialogue between the interviewer and the respondents. The interviewer invariably brought structure into this dialogue by asking questions informed by his or her preconceptions about clinical supervision. The authors did not critically reflect on how these preconceptions influenced the particular interview questions and probes, and subsequently, the findings of the study in any of the papers. Second, stepwise and gradual methods for analyzing text were described in all of the papers, and four papers touched briefly on strategies for validating the emerging themes, peer review (Arvidsson et al. 2000; 2001; Scanlon & Weir 1997), and participant validation (Sloan 1999). However, none of the papers described strategies for testing interpretations; for example, attempts to challenge or falsify emerging themes by examining outliers (Miles & Huberman 1994). Such testing could have facilitated reflections on conceptions that the researchers took for granted. Thus, it remained unclear to what extent the papers’ findings were affected by the researchers’ preconceptions of clinical supervision.

The validity/credibility (Miles & Huberman 1994) of the studies was questionable because of incomplete or missing descriptions of the datasets in relation to the identified themes and subthemes. All the studies had a fixed number of informants, and the amount of data in the studies was not reported. In two projects, the range (length in minutes) between the shortest and longest interview was reported, but none of the studies reported the absolute amount of data (text) actually used in the analysis. Further, none of the studies discussed at which point themes and subthemes were regarded as saturated with data, but one project (two studies) (Arvidsson et al. 2000; 2001) provided an account of the distribution of text units across themes and participants. Summing up this issue of validity/credibility, the results of all the themes and subthemes identified in the dataset, had to be saturated with data from a limited number of informants, but how this analytical task was actually performed was not discussed in any of the studies and could not be evaluated by readers on the basis of the background information about the studies.

Case studies

Like the interview studies, the case studies used interpretative approaches. The case studies differed from the interview studies by utilizing different and varied methods for generating data (See Table 4). One study analyzed transcriptions of video recordings (Stevenson & Jackson 2000), two studies analyzed transcriptions of audio recordings (Severinson 1995b; Sloan & Watson 2001), two studies analyzed notes taken during clinical supervision sessions (Hallberg et al. 1994a; Holst et al. 1999), and two studies analyzed data generated during prolonged fieldwork (Cleary & Freeman 2005; Grant 2000). The former five studies, which were concerned with generating data during clinical sessions, gave no account of the premises for this data construction (e.g. were, and if so, how were non-verbal expressions systematically interpreted and put into writing?). Without this information it remained unclear what type of text the following interpretations were based upon. Furthermore, and surprisingly, data were analyzed almost without explicit considerations of the social context for text construction. This meant that the analyses were not concerned with the relationship between what was said and the particular social dynamics during the sessions. In other words, the social dynamics during the sessions were disregarded as data in the analyses. The pilot study by Sloan and Watson (Sloan & Watson 2001) had some tendencies towards analyzing supervisory dynamics as John Heron’s intervention analysis (Heron 1989) was used to categorize supervisor interventions. However, the links between supervisor interventions and supervisee responses were not explicitly examined.

As data were produced and interpreted this way, the analyses in the five case studies resembled the analyses from the interview studies, and like the interview studies, these studies could be challenged regarding the accounts of the relationship between delimited data sets and the identified themes, as well as the researcher’s effects on the findings (see the critique above).

The two studies that drew substantially on data generated through fieldwork offered contextually sensitive interpretations of both nurses’ attitudes to clinical
<table>
<thead>
<tr>
<th>Study</th>
<th>Aim</th>
<th>Design</th>
<th>Method/instruments</th>
<th>Setting and participants</th>
<th>Analyses</th>
<th>Results</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berg and Hallberg (2000). (Part of the same project as Berg &amp; Hallberg 1999)</td>
<td>To reveal the meaning and significance of psychiatric nurses' experiences with CS and individually planned care</td>
<td>Qualitative interviews after 1 year with a new CS and care-planning programme</td>
<td>Semistructured interviews. An interview manual with open-ended questions focusing on 2 overall issues</td>
<td>A general psychiatric ward in Sweden. 16 beds. 22 nurses: all nurses on the ward</td>
<td>21 transcriptions. The analysis was inspired by latent content analysis and themes were identified</td>
<td>2 themes: 1. 'Confronting the complexities of ongoing life in daily nursing care'. CS gives the opportunity to speak out, but can also create confrontations within the group. 2. 'Strengthening the foundation for nursing care'. CS leads to increased professional awareness, but also to antagonism and displeasure among the staff</td>
<td>The interpretative process was inadequately described and the level of influencing preconceptions was not reflected on. The extent of the interviews was not reported and the follow-up questions were not reported</td>
</tr>
<tr>
<td>Scanlon and Weir (1997)</td>
<td>To examine to what extent nurses experience CS as helpful and what they regard as hindering and facilitating effective CS</td>
<td>Qualitative interviews</td>
<td>Semistructured interviews. A 'loose' interview guide</td>
<td>10 psychiatric nurses (supervisees) from south-east England (4 from community settings, 4 from in-patient settings, and 2 from a residential therapeutic community)</td>
<td>10 transcriptions. The constant comparative method</td>
<td>The nurses perceived CS as valuable, but stated that CS did not take place to a sufficient degree</td>
<td>The participants were identified among the investigators' own professional contacts. The extent of the interviews was not reported and the follow-up questions were not reported</td>
</tr>
<tr>
<td>Arvidsson et al. (2000); Arvidsson et al. (2001)</td>
<td>To describe conceptions of how group CS influences professional competence</td>
<td>Qualitative interviews after the first and second year of 2 years of group CS and then 4 years after</td>
<td>Semistructured interviews. Open-ended questions</td>
<td>10 strategically-sampled psychiatric nurses in Sweden</td>
<td>10 transcriptions from each period of data collection. Phenomenographic method: Thematic analysis of conceptions and narratives</td>
<td>1. 4 description categories: 1. A feeling of job satisfaction. 2. Gaining knowledge and competence. 3. Gaining a sense of security in nursing situations. 4. A feeling of personal development (Arvidsson et al. 2000). 2. 2 additional categories were identified in the follow-up study: 5. Realizing the value of supervision. 6. A sense of professional solidarity (Arvidsson et al. 2001)</td>
<td>The interview questions probe only for positive experiences of CS, not negative. There was no strategy for challenging conceptions and descriptive categories. The analysis of changes between data collections is not described sufficiently</td>
</tr>
<tr>
<td>Sloan (1999)</td>
<td>To identify the characteristics of a good supervisor from a supervisee perspective</td>
<td>A questionnaire and a focus group</td>
<td>Nominal group technique based on a questionnaire (developed from Fowler 1995) and a focus group interview</td>
<td>A convenience sample of 33 staff nurses working in a community health care trust in the UK. 8 completed the questionnaire and 6 participated in the focus group</td>
<td>Nominal group technique and thematic content analysis of the data from the focus group</td>
<td>3 categories: 1. What happens during GSI? CS sessions are divided between clinical business, emotional support and professional development. 2. Good qualities of a clinical supervisor: the most important qualities were to make the supervisee comfortable and develop supportive relationships. 3. Limitations: There was no result of how CS is conducted: the allocation of supervisor and the documentation of the sessions were most problematic</td>
<td>The questionnaire, which was the outset for the focus group discussion, was very short. The length of the focus group and the amount of information was unclear. The interpretative process was inadequately described and the level of influencing preconceptions was not reflected on</td>
</tr>
<tr>
<td>Olofsson (2005)</td>
<td>To describe psychiatric nurses' experiences of participating in CS reflection groups</td>
<td>A retrospective interview after a period with group supervision</td>
<td>A structured interview with 3 overall topics</td>
<td>23 nurses at 2 Swedish units (general psychiatry and elderly psychiatric care) participated in 11 reflection groups focusing on the use of coercion; 21 of these nurses were interviewed</td>
<td>Content analysis of notes taken during the 21 interviews</td>
<td>2 domains: 1. Psychiatric nurses' views of systematic CS and staff support. The nurses valued CS. 2. Psychiatric nurses' experience of participating in reflection groups on the use of coercion; there were both negative and positive aspects to participating in the groups</td>
<td>Most participants participated in only 1 reflection group. Interviews were retrospective, looking as far back as 11 months. The amount of data was not reported and data were recorded by hand. The analysis was not described in any detail</td>
</tr>
</tbody>
</table>

CS, clinical supervision.
<table>
<thead>
<tr>
<th>Study</th>
<th>Aim</th>
<th>Method/instruments</th>
<th>Setting and participants</th>
<th>Analyses</th>
<th>Results</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stevenson and Jackson (2000)</td>
<td>To implement and evaluate democratic CS (egalitarian meetings)</td>
<td>Video recordings of 6 CS sessions</td>
<td>6 (5–8) community psychiatric nurses from an English mental health trust</td>
<td>Transcriptions of the recordings were analyzed with a modified version of grounded theory</td>
<td>The participants found the democratic CS meetings to be a liberating alternative to traditional, hierarchical CS. The group allowed them to question existing practices, but left the participants anxious about returning to traditional practices of CS</td>
<td>It was unclear when the analyses were concerned with reflections on CS or the actual CS taking place. The researchers were also informants and it was unclear to which degree the findings were a result of the researchers’ wish to promote democratic CS. The actual process of democratic CS was insufficiently described</td>
</tr>
<tr>
<td>Severinson (1995b)</td>
<td>To analyze how nurses experience CS</td>
<td>Audio recordings of 10 CS sessions from the beginning, middle, and end of a 2-year intervention</td>
<td>5 nurses (supervisees) and a nurse supervisor at a Swedish psychiatric clinic</td>
<td>The transcribed recordings (130 pages) were analyzed with grounded theory and interpreted hermeneutically</td>
<td>4 sets of factors influence the nurses’ professional growth. The nurses were influenced by guilt related to fear and responsibility. They talked about understanding, forgiveness, and reconciliation. Faith, love, and confirmation were mediated by the supervisor</td>
<td>The results were presented incoherently. It remained unchallenged to which extent the researcher’s preconceptions influenced the results</td>
</tr>
<tr>
<td>Sloan and Watson (2001)</td>
<td>To explore which CS interventions are facilitating or constraining</td>
<td>An audio recording of a CS session</td>
<td>2 nurses from a community mental health team in Scotland</td>
<td>Categorization of the recording (a mixture of inductively and deductively derived categories)</td>
<td>7 supervisor interventions were identified. The research approach was appropriate</td>
<td>It was a pilot study and only a reduced amount of data and results were presented. Both the analytical process and the origin of data were unclear</td>
</tr>
<tr>
<td>Holst et al. (1999) (Part of the same study as Berg et al. 1994; Edberg &amp; Hallberg 2001; Edberg et al. 1996, 1999; Hallberg &amp; Noebel 1993; Hallberg et al. 1994b)</td>
<td>To explore nurses’ reflections on caring and patients during CS</td>
<td>Test analysis of notes taken during CS</td>
<td>10 patients at 2 wards at a psychogeriatric clinic in Sweden. 11 beds on each ward</td>
<td>A phenomenological hermeneutic approach. 48 situations were categorised according to the type of encounter and the how it affected the participants’ identity</td>
<td>CARE for people with severe dementia included more or less meaningful encountering that could confirm or threaten the person’s identity</td>
<td>The procedure for note taking and the amount of data were not accounted for. The interpretative process was not clearly described, including eventual strategies for validating finding</td>
</tr>
<tr>
<td>Hallberg et al. (1994a)</td>
<td>To understand a nurse’s experiences of working with a person suffering from schizophrenia</td>
<td>Test analysis of notes taken during 15 CS sessions</td>
<td>A Swedish male nurse</td>
<td>Thematic analysis</td>
<td>The nurses’ description of himself and the patient during CS suggested parallel emotional processes</td>
<td>The amount and character of data not described. The interpretative process was not clearly described. The descriptions of the patient’s feelings were made by the nurse and not the patient himself</td>
</tr>
<tr>
<td>Cleary and Freeman (2005)</td>
<td>To explore nurses’ perceptions of professional attitudes and support in relation to CS</td>
<td>5 months of part-time ethnographic fieldwork including observations, 10 interviews, and 5 open discussion groups</td>
<td>A 22-bed admission ward in an Australian public hospital</td>
<td>Thematic analysis</td>
<td>Nurses viewed CS as beneficial, but did not find it practically feasible. Many nurses viewed ad hoc reflection with peers and other supportive systems just as (and more) appropriate than formal CS and thus they were reluctant towards implementing CS</td>
<td>The interpretative process was not clearly described, including eventual strategies for validation/falsification of findings</td>
</tr>
<tr>
<td>Grant (2000)</td>
<td>To explain the mismatch between the perceived importance of CS and the failure of its implementation</td>
<td>Unstructured interviews, organizational documents and texts on CS</td>
<td>43 senior executive officers, managers, and staff in an English mental health trust</td>
<td>An amalgam of qualitative data analyses and participant validation of categories</td>
<td>A set of related assumptions: That the trust board was instrumental in starting CS; however, the board was not really committed. CS was good and the board needed persuasion about the importance of CS, but no one could or saw the point in marshalling the necessary arguments for CS</td>
<td>There were procedures for verification of categories, but none for falsifying</td>
</tr>
</tbody>
</table>

CS, clinical supervision.
supervision during everyday work and the organizational decision making regarding clinical supervision (see Table 4). These two case studies indicated that it was difficult to implement regular, formalized clinical supervision sessions because nobody in the organization, nurses or administrators, felt the need and supported the implementation of clinical supervision.

**DISCUSSION**

On the basis of the evidence presented in this systematic review of peer-reviewed empirical studies on clinical supervision in psychiatric nursing, it was only possible to draw very tentative conclusions regarding the effect of clinical supervision, the prevalence of clinical supervision, and/or psychiatric nurses’ attitudes towards clinical supervision. In general, studies were relatively small scale, used relatively new and basic methods for data collection and analysis, and did not include sufficient strategies for identifying confounding factors and how the researcher’s preconceptions influenced the analyses. The review indicated a general potential for advancing empirical research on clinical supervision through a wider adherence to more sophisticated and established research methods.

The search for peer-reviewed references on clinical supervision in psychiatric nursing had two limitations. First, the search was based on the assumption that the studies of the practices of clinical supervision could be identified by the search term ‘clinical supervision’. However, there were different conceptualizations of these practices in different parts of the world, notably between Scandinavia, the UK, and the USA (Cutcliffe & Lowe 2005), there were several practices resembling clinical supervision, for example, mentorship and preceptorship (Lynch & Happell 2008a,b,c; Lynch et al. 2008), and some authors locate clinical supervision within the wider context of ‘reflective practice’ (e.g. Johns & Freshwater 2005). Therefore, searching for references using the search term ‘clinical supervision’ may not have identified all references on the clinical phenomenon/practice. Second, the review took an outset in databases that did not include all types of monographs, chapters in books, and grey literature. However, this limitation could be regarded as minor because of the citation pearl growing strategy. We found that all the encountered monographs were also published in peer-reviewed journals (e.g. Severinsson 1995a; Sloan 2006), and no substantial empirical studies on clinical supervision in psychiatric nursing were solely published as chapters in edited books.

A major challenge for carrying out convincing effect studies of clinical supervision in psychiatric nursing is that it is very difficult to organize a randomization of participants and to control the experimental setting for confounding factors for prolonged periods of time. Variables such as shifting workloads, stressful events, interpersonal and interprofessional conflicts, and patients’ degenerating or volatile conditions can confound measurements of the effect of clinical supervision. Furthermore, it is also very difficult to standardize and measure the clinical supervision intervention because the supervisor’s persona is a significant part of the supervisory effect. These difficulties hamper most attempts to carry out convincing trials of clinical supervision. These problems of organizing the study settings may explain why the few studies in the review that used the same well-documented measures, for instance, the MBI yielded contradictory results and why there were so few significant changes following supervision. Finally, because of the difficulties related to the practical organization of the research setting, convincing effect studies may be easier to carry out in outpatient settings where the relationships between the supervision intervention, the nurse(s), and the patient(s)/client(s) are more direct (one on one), and thus more controllable compared to inpatient settings.

The review of the survey studies revealed ambiguous results regarding associations between clinical supervision and other clinical phenomena. Future survey research in this field would benefit from more large-scale studies, preferably with longitudinal designs, with high response rates. The lack of consistent research findings was not just related to a complex research subject, as described earlier, there was also a general lack of consensus among the researchers regarding what research instruments to use in the studies. Instruments have been made available specifically for measuring attitudes towards clinical supervision (e.g. Major’s attitudes towards CS scale) (Major 1993), and the self-reported quality and effect of clinical supervision (e.g. MCSS) (Winstanley 2000); however, the application of these instruments was not widespread. Further, there was some overall agreement on which phenomena were relevant to examine in relation to clinical supervision. This was typically work-related phenomena, such as burnout, stress, and job satisfaction. Still, there was no consensus on which particular instruments to utilize, and this lack of consensus regarding research instruments complicated comparisons between studies. A number of the survey studies utilized the same instruments (e.g. MCSS and MBI). In one instance, comparable associations regarding one (depersonalization) out of three MBI factors were found (Edwards et al. 2006; Hyrkas 2005). To determine whether such findings are coincidental, there is a need for more widespread...
agreement regarding concepts and models of clinical supervision and how these concepts should be operationalized and measured in empirical research.

The review of the interview studies indicated that they were concerned with developing an understanding of the meaning of concepts related to clinical supervision and with describing and evaluating practices. In other words, the researchers were primarily interested in the meaning of clinical supervision and what the respondents thought about clinical supervision. The researchers only had a secondary interest in gaining an understanding of how the respondents felt about participating in clinical supervision; for example, ‘how does it feel trying to find time for a session?’ or ‘how does it feel to engage with an external supervisor from another profession?’ A wider focus for interview studies may, along with a combination of observational data, create deeper insights into why nurses may be ambivalent about clinical supervision. Two case studies did combine interview data with observational data and yielded situational sensitive interpretations of supervisory practices. However, the review showed that the five case studies that generated data during supervision had in effect disregarded interpretative attention to the social dynamics of clinical supervision sessions. An inclusion of appropriate methods for generating and interpreting data about discursive and social interaction would allow researchers to examine supervisory interventions in further detail.

Proctor’s model (Proctor 1987) was the most widely used model in the studies. The model is a theoretical statement. The supervisor and supervisee(s) have three joint roles and tasks in supervision: normative (controlling/instructive), formative (educative/reflective), and restorative (recreational) roles and tasks. The model does not consider clinical outcomes of clinical supervision or situational/organizational factors impinging on clinical supervision. In spite of widespread references, the model could not be used to categorize the studies in the review because many of the studies had explorative aims and aims addressing issues outside Proctor’s three functions. The largest group of studies that could be fitted into the model concerned the restorative (stress-reducing function). More research needs to be done on how to extend and refine the model, that is, how to include dynamic relations between the tasks and roles and clinical and situational/organizational factors.

CONCLUSION

Clinical supervision in psychiatric nursing is commonly perceived as a good thing (Mullarkey et al. 2001), but the present review unequivocally indicates that the empirical evidence supporting this claim is limited. Empirical research studies of clinical supervision in psychiatric nursing were few in numbers, relatively small scale, and used new and relatively basic methods for data collection and analysis. Moreover, a substantial part of the studies suffered from serious methodological flaws. Clinical supervision may have a beneficial effect on psychiatric nurses, but if this effect is to be identified through research, then future studies need to be larger in scale and concur with general conventions for valid and reliable research.

The most serious obstacle for developing the field was probably the general lack of consensus regarding which definitions and models should guide empirical research of clinical supervision. This includes which research questions to pose, which variables and issues to examine, and what instruments to use in the study of clinical supervision. The lack of agreement hampers the possibilities for comparisons of results, and as a further consequence, an accumulative field of knowledge.

REFERENCES


Bradshaw, T., Butterworth, A. & Mairs, H. (2007). Does structured clinical supervision during psychosocial intervention education enhance outcome for mental health nurses and
the service users they work with? Journal of Psychiatric and Mental Health Nursing, 14 (1), 4–12.


