

Effects of clinical supervision of mental health professionals on supervisee knowledge, skills, attitudes and behaviour, and client outcomes: protocol for a systematic review

Robert Allan, Alan McLuckie, Lillian Hoffecker

Submitted to the Coordinating Group of:

| | |
|-------------------------------------|--|
| <input type="checkbox"/> | Crime and Justice |
| <input type="checkbox"/> | Education |
| <input type="checkbox"/> | Disability |
| <input type="checkbox"/> | International Development |
| <input type="checkbox"/> | Nutrition |
| <input checked="" type="checkbox"/> | Social Welfare |
| <input type="checkbox"/> | Methods |
| <input type="checkbox"/> | Knowledge Translation and Implementation |
| <input type="checkbox"/> | Other: |

Plans to co-register:

| | | | |
|-------------------------------------|-------|-----------------------------------|--------------------------------|
| <input checked="" type="checkbox"/> | No | | |
| <input type="checkbox"/> | Yes | <input type="checkbox"/> Cochrane | <input type="checkbox"/> Other |
| <input type="checkbox"/> | Maybe | | |

Date Submitted: 13 March 2016

Date Revision Submitted: 6 July 2017

Approval Date:

Publication Date: 26 July 2017

BACKGROUND

The problem, condition or issue

Clinical supervision is increasingly being recognized as a core professional competency within the mental health field (Brosan, Reynolds, & Moore, 2008). Supervision is also considered an essential component of modern effective health care systems (Kadushin, 2002) and training programs for mental health professionals (Berger & Mizrahi, 2001; Milne, Sheikh, Pattison, & Wilkinson, 2011; Watkins, 2011). The term mental health professional is used here to refer to all clinicians working in the mental health field, including psychotherapists, counsellors, social workers, psychologists, psychiatrists, nurses, pastoral counsellors and couple/marriage and family therapists as well as trainees in each of these professions. Similarly, the term mental health work refers to the clinical work for each of these professions, which includes, but is not limited to the provision of direct psychosocial interventions (i.e., non-pharmacological interventions), such as mental health therapy, couple and family counselling, psychotherapy and psychosocial interventions designed to improve, enhance, maintain socio-emotional, psychological, behavioural, interpersonal functioning and/or reduce risk factors, such as psychiatric symptoms and/or addictions. Despite the culture of evidence-based practices in mental health settings, the practice of supervision of mental health professionals lags behind in its use of evidence-informed practices (Schoenwald et al., 2009). Further exacerbating the problem is that most supervisors assume, perhaps erroneously, that the supervision they provide is effective (Kilminster & Jolly, 2000).

There are at least three clinical reasons (Milne, 2009) to develop a better understanding of supervision: one, effective supervision may be essential to enhance and maintain mental health professionals' competencies; two, supervision can increase fidelity to evidence-based treatment models; and third, effective supervision can reduce unnecessary interventions (e.g., mental health professionals making referrals to multiple services when contraindicated or failing to disengage services when goals are achieved) and reduce waitlist times and health care costs. Although the quality of the evidence emerging from outcome research on the effectiveness of supervision is limited (Schoenwald, Sheidow, & Chapman, 2009; Waller, 2009) due to methodological issues, there is indication that supervision results in improved patient care outcomes (Bambling, King, Raue, Schweitzer, & Lambert, 2006; Bradshaw, Butterworth, & Mairs, 2007; Callahan, Almstrom, Swift, Borja, & Heath, 2009; Milne, Aylott, Fitzpatrick, & Ellis, 2008; Watkins, 2011) and that it acts as a quality assurance mechanism (Schoenwald et al., 2009). Without supervision, the quality control of mental health work depends on the ability of mental health professionals to self-evaluate their competencies (Hansen et al., 2006). Self-evaluations prove to be difficult with early career and lower skilled clinicians who are found to typically over-rate their competencies (Vallance, 2005), which can have negative implications for patient outcomes and safety.

The intervention

Clinical supervision is an ongoing supportive learning process for clinicians of all levels to develop, enhance, monitor, and, when necessary, remediate, professional functioning (Bernard & Goodyear, 2014). Supervision is a distinct professional practice with knowledge, skills, and attitudes components. For some professions (e.g. Marriage and Family Therapy), supervisors require specific training to be recognised as an “approved supervisor” while other professions promote experienced clinicians into the role of “supervisor” after some time and clinical experience (Falender, Burnes, & Ellis; 2013; Falender, Ellis, & Burnes, 2013; Bernard & Goodyear, 2014; Reiser & Milne, 2012). The chief function of supervision is to minimize non-purposeful activity and maximize intentionality with the goal of directly optimizing clinician competencies, ensuring quality control, and enhancing confidence for the end goal of improving patient outcomes (Milne, 2009). Supervision is provided in a variety of formats including one-on-one supervision, small group supervision, peer-based consultation, and facilitated team-based consultation. Supervision can include presentations via case discussion, video review or live presentation/demonstrations (Todd & Storm, 2002).

There are a number of definitions of supervision put forth by the various mental health professions. For example, the American Psychological Association (2014) defines supervision as:

... a distinct professional practice employing a collaborative relationship that has both facilitative and evaluative components, that extends over time, which has the goals of enhancing the professional competence and science-informed practice of the supervisee, monitoring the quality of services provided, protecting the public, and providing a gatekeeping function for entry into the profession. (p. 5)

The Association for Counsellor Education and Supervision (2011) provide a list of guidelines in twelve areas for addressing the ethical and legal protection of the rights of supervisors, supervisees, and clients; and meeting the professional development needs of supervisees while protecting client welfare. While the Council for Accreditation of Counseling and Related Educational Programs (2009) offer a definition of supervision in an educational setting as:

A tutorial and mentoring form of instruction in which a supervisor monitors the student’s activities in practicum and internship, and facilitates the associated learning and skill development experiences. The supervisor monitors and evaluates the clinical work of the student while monitoring the quality of services offered to clients. (p. 62)

Further definitions are offered by the American Association for Marriage and Family Therapy (AAMFT), the National Association of Social Work (NASW), the British Association of Counselling and Psychotherapy (BACP), and the Royal Australian and New Zealand College of Psychiatrists (RANZCP).

AAMFT (2014) describe the process of supervising marriage and family therapists (MFT) as:

...evaluating, training, and providing oversight to trainees using relational or systemic approaches for the purpose of helping them attain systemic clinical skills. Supervision

is provided to an MFT or MFT trainee through live observation, face-to-face contact, or visual/audio technology-assisted means. (p. 5)

Another definition of supervision is offered from the NASW (2013). This organisation defines professional supervision as:

the relationship between supervisor and supervisee in which the responsibility and accountability for the development of competence, demeanor, and ethical practice take place. The supervisor is responsible for providing direction to the supervisee, who applies social work theory, standardized knowledge, skills, competency, and applicable ethical content in the practice setting. The supervisor and the supervisee both share responsibility for carrying out their role in this collaborative process. (p. 6)

The BACP (2016) defines supervision as a “formal arrangement for therapists to discuss their work regularly with someone external in order to maintain adequate standards of therapy”. The BACP recommends that supervisors not know the identity of clients and be members of a recognized professional body and preferably accredited. A final definition of supervision is offered by the RANZCP (2012) that outlines an apprenticeship process that includes a minimum of four hours per week for 40 weeks of supervision by a College-accredited supervisor. The RANZCP define supervisors as clinically competent in the area they are supervising in and familiar with the use of a competency-based assessment tool. Each of the definitions developed by the different professional organisations share developmental, ethical, and supportive roles while each offer different emphases based on the epistemological roots of their professions and understandings of supervisory relationships.

There are also different approaches to **defining the intervention** of clinical supervision. Milne (2007), for example, proposed four criteria for developing a definition of supervision. The first criterion is *precision* which can be accomplished by clarifying what is distinct about supervision and also in comparison with what it is not (e.g. not therapy). Secondly, a definition of supervision must include *specification*. This will include the elements that make up supervision (e.g. video review). The third criterion proposed by Milne is that of *operationalization* which he suggests must be stated in the form that permits measurement. The fourth and final criterion is *corroboration* which can be substantiated with adequate support from the research. There are a number of challenges for this research to applying all four of these criteria for a single definition of clinical supervision or to promote as Milne (2007) suggests an “empirical” definition of clinical supervision.

One challenge is the varied understanding of supervision across the mental health professions. Psychology is the only mental health profession that explicitly adheres to a science-practitioner model (Frank, 1984) and places an emphasis on what is measurable by the senses. While each profession engages in evidence-based practices there is a richer dialogue of what constitutes evidence among the different professions as well as a critique of the narrow band of empiricism that drives the science-practitioner approach (e.g. Coulter, 2011; Elliott, 1998; Gambrill, 2010; Holmes, Murray, Perron, & Rail, 2006; Slife, Wiggins, &

Graham, 2005; Staller, 2006; Wendt, Jr., 2006). To be systematic, it is important that we review research from across the mental health professions.

For the purpose of this systematic review we will include research that defines supervision as involving a supportive learning process for clinicians when carrying out psychotherapy, counselling and/or the provision of psychosocial mental health interventions of all levels to develop, enhance, monitor, and when necessary, remediate, professional functioning. We will include research where the supervision sessions may have: case-presentations, presentation of video or audio-tapes from a therapy session, exploration of self-of-the-therapist issues, or process recordings (interpersonal process recall). And the supervision can take place: one-to-one, triadic (1 supervisor and dyad), in a group format, live (with call-in and/or with bug in ear), consultation teams, reflecting teams, or online. One of the objectives of this review is to assess the research about supervision of multicultural competencies. A further clarification is required to address the understanding of multicultural competencies.

Developing multicultural competence is integral to the formation of clinical competence (Falender, Shafranske, & Falicov, 2014). Multicultural competencies in the mental health professions include a range of attitudes, beliefs, knowledge, skills, and actions (Ratts, Singh, Nassar-McMillan, Butler, & McCullough, 2015) that provide a framework to optimize client engagement, participation and benefit from psychotherapeutic intervention and research. Developmental domains of multicultural competencies include mental health professional self-awareness, the client's worldview, the therapeutic relationship, and mental health and advocacy interventions (Ratts et al., 2015). There are a number of obstacles to integrating cultural perspectives in supervision including the need to clarify the role of understanding what cultural heritage and sociopolitical context have to do with human suffering and critically examining the epistemological foundations of the psychotherapies that are used (Falicov, 2014). Multicultural competence is considered an ethical and practice imperative and there is a need to clarify the best research-based approaches to the supervision of mental health professionals in this area (Falender, Shafranske, & Falicov, 2014). The aim of this part of the systematic review is to identify the research pertaining to the supervision of multicultural competencies in the mental health therapies. This will not include training or workshops about multicultural competencies.

How the intervention might work

The logic model for this intervention outlines a series of themes as opposed to a singular intervention, as clinical supervision takes many forms and can be informed by a myriad of models ranging from behavioral to psychodynamic approaches (often mirroring mental health intervention approaches). There is, however, a shared objective for supervision across professions to minimize non-purposeful activity and maximize intentionality with the goal of directly optimizing clinician competencies, ensuring quality control, and enhancing confidence for the end goal of optimizing care (Milne, 2009). Supervisors accomplish one or

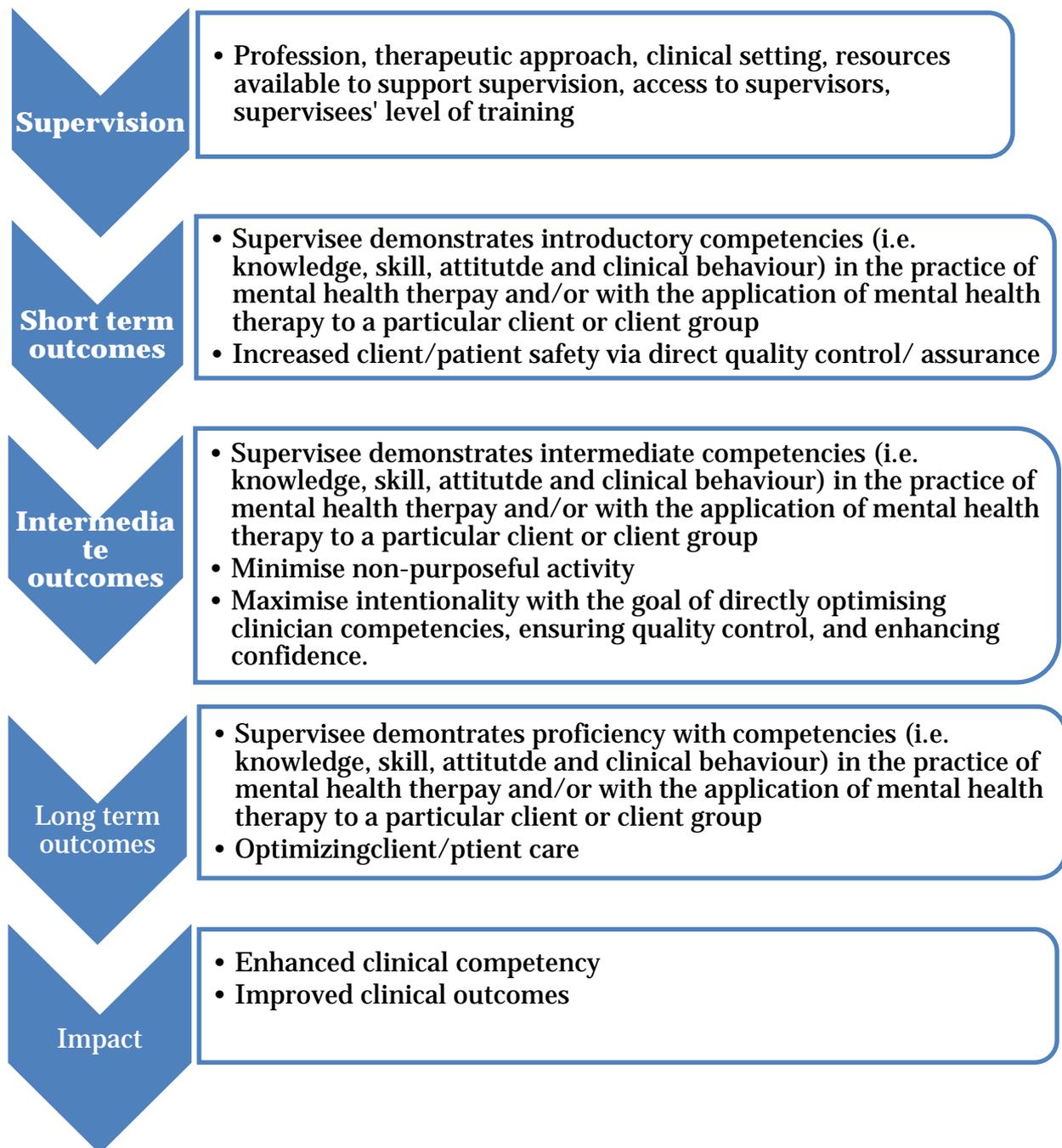
more of these objectives by targeting supervisee knowledge, skills, attitudes and/or behaviours.

Educational settings responsible for supervising mental health trainees/students share the duties as outlined by Council for Accreditation of Counseling and Related Educational Programs (2009). These supervision interventions include a tutorial and mentoring form of instruction in which a supervisor monitors the student's activities and facilitates the associated learning and skill development experiences. The measurement of this form of supervision happens with direct verbal feedback, grading, and feedback forms. Also important to consider when discussing supervision interventions is the clinical approach such as family or systemic therapy.

For systemically oriented supervisors and mental health professionals, the self is unavoidably a part of the therapeutic system (Cheon & Murphy, 2007; Lee & Everett, 2004). Integrating the self into the therapeutic system and supervision generates self-of-the-therapist issues that are to be explored in supervision (Aponte et al., 2009). One of the purposes of this self-exploration is to learn how to use these emotional materials to enhance the effectiveness of the professional use-of-self (Timm & Blow, 1999). Aponte et al. (2009) report that the degree to which mental health therapists commit to exploring the challenges in their lives and engage in personal growth and development is proportionate to the ability to relate to clients' efforts to deal with their challenges. Progress in supervision is measured by supervisees' reflective journaling, review of video of therapy sessions, and exploration of challenges in group and individual supervision. Assessment of the supervision is through supervisee self-reports and review of clinical work by supervisors.

The intervention of supervision includes knowledge, skill, attitude, and/or behavioural components and the priority placed on each will vary across supervision approaches for reasons previously noted. Moreover, supervision of mental health professionals can vary according to the: supervisor, therapeutic approach, clinical setting, mental health worker, resources available to support supervision, access to supervisors, profession, as well as other factors. The means by which supervision is evaluated varies from experimental designs to qualitative approaches. Experimental and quasi-experimental designs typically employ quantitative measures, compared before and after supervision and with comparison groups in order to determine whether supervision is effecting improvements on the supervisee's skills and/or knowledge, as well as on clinical outcomes with the supervisee's clientele. Qualitative methods are often harnessed to help illuminate the experiences and perspectives supervisees and/or supervisors have with the supervision process as well as help identify the key facets of the supervision mechanism.

Logic model:



Why it is important to do the review

There are no systematic reviews on this topic registered with the Campbell Collaboration or the Cochrane Collaboration to date. However, a number of previous reviews have addressed clinical supervision, but have been somewhat restricted and have not looked at it from an overall mental health professional perspective. For example, Bogo and McKnight (2006) published a review of the research and non-research literature pertaining to clinical supervision within the field of social work and social work trainee field education. Due to the

broad scope of practice of social workers Bogo and McKnight did not focus on the practice of mental health therapy by social workers or the supervision of this practice. A systematic review of social work supervision practices specific to the field of child welfare was conducted by Carpenter, Webb and Bostock (2013), however their scope did not include mental health therapies.

Brunero and Stein-Parbury (2008), Francke and Graaff (2012), Butterworth, Bell, Jackson, and Pajnikihar (2007), as well as Cummins (2008) also conducted reviews of the literature examining the effectiveness of clinical supervision for nurses. The focus of the reviews included a broad range of nursing interventions and included few studies pertaining to nursing supervision specific to the practice of mental health therapy. Dawson, Phillips and Leggat (2013) broadened the population of their review to include all allied health professions, however, similarly they did not focus on mental health therapy.

Faman and colleagues (2012) completed a systematic review of outcomes of supervision on patient care and medical residents' competencies. Although this review included a targeted population for psychiatry, this study did not search for, or identify studies specific to supervision for mental health therapies. Other reviews of studies within the field of medicine, such as those by Kilminster and Jolly (2000) identified key mechanisms of supervision leading to positive outcomes including the nature of the supervisory relationship. Other reviews looking at allied health professionals (Barak, Travis, Pyun, & Xie, 2009) focused on positive outcomes for supervision arising through the supervision process including enhanced job satisfaction and reduced burnout. Follow-up reviews are needed as these reviews did not specifically examine supervision pertaining to mental health therapy. A review conducted by Milne and James (2000) did focus on supervision for mental health therapy, however, mental health therapy was narrowly defined as cognitive behavioural therapy (CBT).

This review is important for a number of reasons. First, little is known about what evidence exists for supporting the use of supervision or what supervision practices may be help/harmful for supervisees, which ultimately impacts the level of care received by patients accessing mental health services (Wheeler & Richards, 2007). Second, little is known about the evidence for multicultural and/or culturally sensitive supervision practices and how these may impact supervisee competence and patient care (Falendar et al., 2014). Third, there is indication in the research literature that competent supervision results in improved patient care outcomes and that it acts as a quality assurance mechanism (Tracey, Wampold, Lichtenberg, & Goodyear, 2014; Watkins, 2012). Without supervision, the quality control of mental health interventions depends on the ability of mental health therapists to self-evaluate their competencies. Self-evaluations prove to be difficult with early career and lower skilled therapists who are found to typically over-rate their competencies, which can have negative implications for patient outcomes and safety. Finally, supervision is considered an essential component of modern effective health care systems and health care training programs in general (Kilminster & Jolly, 2000). This systematic review will contribute to

enhancing our knowledge of effective supervision practices, including the impact of multicultural supervision, which will lead to improved care, better training, and better management of care.

OBJECTIVES

The primary objectives of this review are to summarize and synthesizes the available evidence on the effects of supervision for enhancing the clinical and multicultural competencies of mental health professionals. The specific research questions guiding our review include:

- 1) What effect does clinical supervision (versus peer consultation or no supervision) have on mental health professionals' clinical competencies (i.e., knowledge, skill, attitude & clinical behaviour) in the practice of mental health therapy?
- 2) What effect does clinical supervision (versus peer consultation or no supervision) have on mental health professionals' multicultural competencies (i.e., knowledge, skill, attitudes and clinical behaviour) in the practice of mental health therapy?
- 3) What effect does clinical supervision (versus peer consultation or no supervision) have on the health outcomes of patients/clients participating in supervised mental health therapy?

The secondary objectives of this review include:

Identifying the key experiences of supervisees and/or supervisors with the supervision process that help identify any reasons why clinical supervision may succeed or fail in improving clinical and/or multicultural competencies in mental health professionals and/or succeed or fail in improving health outcomes for patient/clients participating in mental health therapy.

METHODS

Criteria for including and excluding studies

Types of study designs

This review will include experimental (i.e., Randomized controlled trials: RCTs) and parallel cohort quasi-experimental designs (QED) if the treatment and control groups were assessed at the same points in time and matching or statistical controls have been employed to ensure no significant baseline differences between these groups in relation to key constructs (i.e., supervisee competence in mental health therapy/interventions, multicultural competency, nature of the supervisory relationship). Studies will be included that have a control group

that include either peer-based consultation or no supervision condition. We will include QEDs due to the likelihood that most evaluations of supervision for mental health professionals will be conducted in real world, naturalistic settings that reduce the feasibility of RCTs. All studies formally included in the review must evaluate the effectiveness of clinical supervision on at least one of the outcomes listed in the subsequent section.

In relation to the secondary objective of this review we will also review non-experimental studies. In our preliminary review of the state of the research landscape pertaining to clinical supervision, qualitative studies reveal that many studies relevant to our focus employ diverse methodologies (i.e., single-subject designs, cohort/longitudinal studies, case control and qualitative studies). Therefore, in order to gain insight into reasons why supervision may succeed or fail to enhance clinical and multicultural competencies of mental health professionals and/or health outcomes of patients, we will report findings from these studies in a tabular format and narrative summary that will provide a fuller context to our findings regarding our analysis of the RCTs and QEDs formally included in our review. Similarly, findings from relevant previously conducted review articles and/or meta-analyses will not be formally included in our review or analyses, but will be reported in the form of a narrative summary within our review. Review articles and meta-analyses identified in our literature search will also serve as a means of identifying any additional studies not previously identified within our search.

Types of participants

The participants to be included in this review must be in a professional, structured supervisory relationship either as a supervisee or supervisor, or as a student in practicum, internship, or residency, whereby the supervisor is identified as having a formal role and responsibility for the supervision process. This supervision must also be in relation to the provision of mental health therapy. Peer supervision (i.e., supervision provided without a formal mandate/scope of practice between two or more individuals practicing in a same or similar role, typically that of psychotherapist) will only be included in this review as a control condition. Types of mental health professionals included in this review are psychotherapists, psychologists, counsellors, couple/marital and family therapists, social workers, psychiatrists, nurses, pastoral counsellors, and trainees in each of these professions.

Types of interventions

This review will include studies examining clearly specified clinical supervision models intended to minimize non-purposeful activity and maximize intentionality with the goal of directly optimizing mental health therapists' competencies (i.e., knowledge, skill, attitude &

clinical behaviour) and client outcomes, ensuring quality control, and/or enhancing confidence for the end goal of optimizing care. The intervention may be delivered to a range of mental health therapists in a variety of formats including: one-to-one, triadic (1 supervisor and supervisee dyad), in a group format, live (with call-in and/or with bug in ear), consultation teams, reflecting teams, or online.

Supervision will be demarcated from theoretical-based training in models of mental health therapy. Research will be included that examines supervision in relation to a supervisee's application of mental health therapy to a specific patient/client(s) situation and/or the supervisory discourse will pertain to a unique patient/client situation. Research will be excluded that is not directly focused on investigating the supervision process as it pertains to the needs of unique patients/clients, but instead focus only on the training of individuals related to the acquisition of a theoretical understanding of a mental health model and/or the acquisition of general knowledge and/or general skills related to mental health therapy without the application of this model to specific patient(s)/client(s) situations. Research studies examining competency development using simulated-patients/clients will be included in the present review, but will be treated as a distinct sub-group for analytic purposes.

Types of outcome measures

The primary outcome(s) for this study will include the measurement of effectiveness of competencies (i.e., knowledge, skills, attitudes and clinical behaviours) in the provision of mental health therapy, including multicultural competencies. We are also interested in outcomes pertaining to the effectiveness of the supervisory relationship/process and client outcomes. Examples of these measures are included below. Additional relevant and validated measures identified through the review of study will also be considered.

Mental health professional's competencies (i.e., knowledge, skill, attitude, and clinical behaviours) in mental health therapy outcomes:

- Licensing exam scores
- Objective Structured Clinical Examination (OSCE) (Harden, Stevenson, Downie, & Wilson, 1975)
- Measures of fidelity to mental health therapy interventions, such as the Therapist Behavior Rating Scale (Hogue et al., 2008) or the Cognitive Therapy for Psychosis Adherence Scale (Startup, Jackson, & Pearce, 2002).

- Supervisor reports and evaluations

Mental health professional's multicultural competencies (i.e., knowledge, skill, attitude, and clinical behaviours) in mental health therapy outcomes:

- Cultural Awareness Tool (Multicultural Mental Health Australia, 2002)
- Cultural Competence Self-Assessment Questionnaire (CCSAQ; Mason, 1995)
- Multicultural Awareness Knowledge/Skill Survey (MAKSS; D'Andrea, Daniels, & Heck, 1990).
- Multicultural Counseling Inventory (MCI; Sadowsky, Taffe, Gutkin, & Wise, 1994).
- Multicultural Counseling Knowledge and Awareness Scale (MCKAS; Ponterotto, Gretchen, Utsey, Rieger, & Austin, 2002).
- Client, Supervisor, Psychotherapists Reports
- Sexual Orientation Counselor Competency Scale (SOCCS; Bidell, 2005)

Supervision outcomes/outcomes pertaining to supervisory relationship:

- Leeds Alliance in Supervision Scale (Wainwright, 2010)
- The Manchester Clinical Supervision Scale (MCSS; Winstanley, 2000)
- Short form Supervision Satisfaction Questionnaire (SFSSQ; Carlton, 2002).

We will also examine the effectiveness of supervision for mental health therapy related to patient/clinical care outcomes. Possible measures to be included are listed below. Additional relevant measures identified through the review of study will also be considered.

Individual Client Functioning Outcomes:

- Global Assessment of Functioning (GAF; American Psychiatric Association, 2000), Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996), Beck Anxiety Inventory (BAI; Beck & Steer, 1990), Child Behavior Checklist (ASEBA; Achenbach, 2001 & 2001), Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992), Parent Stress Index (PSI-4; Abidin, 2012), State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), Outcome Rating Scale (ORS; Miller &

Duncan, 2000), Patient Health Questionnaire, Clinically Useful Depression Outcome Scale (CUDOS; Zimmerman, Chelminski, McGlinchey, & Posternak, 2008), Centre for Epidemiological Studies Depression Scale (CESD; Radloff, 1977) and Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES; Miller & Tonigan, 1996).

Couple/Marital Functioning Outcomes:

- Dyadic Adjustment Scale (DAS; Spanier, 1989), Relationships Assessment Scale (RAS; Hendrick, 1988), Experiences of Close Relationship (Revised) (ECR-R; Fraley, Waller, & Brennan, 2000), Locke-Wallace Marital Adjustment Test (MAT; Locke & Wallace, 1959), The Quality of Marriage Index (QMI; Norton, 1983), Golombok-Rust Inventory of Marital State (GRIMS; Rust, Bennun, Crowe, & Golombok, 1986), Relationship Dynamics Scale, The Couples Satisfaction Index (CSI; Funk & Rogge, 2007).

Family Functioning Outcomes:

- Family Quality of Life Scale (FQOL; Beach Center on Disabilities, 2006), Family Assessment Device (FAD; Epstein, Baldwin, & Bishop, 1983), Family Beliefs Inventory (FBI; Roehling & Robin, 1986), Family Adaptability and Cohesion Evaluation Scale, Kansas Family Life Satisfaction Scale, Self-Report Family Instrument.

Adverse events:

- Decrease in competency of the supervisee including reports of ethical and/or competency violations to the various mental health regulator/licensing bodies
- Deterioration of the supervisor-supervisee relationship including reluctance to seek-out/engage in supervision.
- Decrease in client/patient outcomes and/or deterioration in client/patient health outcomes.

Duration of follow-up

The duration of follow up may vary due to the nature of the different types of supervision provided, and the limitations of the studies. A portion of the studies are likely to include supervision that is time limited, such as graduate program internship, practicum and/or residency supervision. Others may include clinical and professional supervision that are

likely an on-going experience throughout a helping professional's career. The follow up duration for this study will be inclusive of all of these supervisory experiences. There will be no restriction placed on the inclusion of articles based on the duration of the intervention and/or the duration of the follow-up measurement pertaining to study outcomes.

Types of settings

This proposed review will include studies of supervision conducted in settings that employ or train mental health professionals engaged in the practice of mental health therapy. Settings will include universities/colleges, child welfare services, psychiatry institutions, training programs, web-based programs, hospitals, counselling agencies, and physicians' offices. The review will include the following experiences: residencies/internships/practicums, student supervision, and clinical supervision across mental health professions. We will exclude studies of educational curriculum of supervision and workshops.

Search methods for identification of studies

We will perform electronic searches of bibliographic databases, as well as on open web-sites and in the grey literature. We will also search for on-going studies. We will have no publication date, geographic, or language restrictions. Media and software reviews will not be included. The following sources will be searched:

Ovid PsycINFO

Ovid MEDLINE

CINAHL

Social Work abstracts

EMBASE (Embase.com)

Web of Science

Social Services abstracts

Educational Resources Information Centre (ERIC)

PubMed

EBSCO ACADEMIC

JSTOR

SAGE

Science Direct

Springer

Wiley

Applied Social Sciences Index and Abstracts (ASSIA)

Cochrane Library

Campbell Library

Google Scholar

Theses and Dissertations (Proquest)

Description of methods used in primary research

The search strategy is comprised of 3 main concept categories: supervision, profession, methodology, and for the objective of addressing multicultural competencies we will add a 4th category, culture. Within each category there are textword (“free text”) terms and, where appropriate, subject headings in indexed databases. We will approach the search from broad-to-narrow in order to capture general studies on clinical supervision, while narrowing the broad search to focus on studies related to multicultural competencies.

The following is one example of the search strategy in OVID PsycINFO:

| | |
|---|--|
| 1 | ((professional adj supervis*) or (clinical adj supervis*) or internship or residency or practicum).ab,ti,id. or professional supervision/ or exp clinical methods training/ or counselor education/ or exp psychology education/ or professional consultation/ or exp postgraduate training/ or practicum supervision/ |
| 2 | ((famil* adj therap*) OR (couple* adj therap*) OR (marital adj therap*) OR (marriage adj therap*) OR (famil* adj counsel*) OR (couple* adj counsel*) OR (marital adj counsel*) OR (marriage adj counsel*) OR (social adj work*) OR (nurs* adj practitioner) OR (mental adj health adj clinician*) OR (psychiatric* adj counsel*) OR therap* OR counsel* OR psycholog* OR psychiat* OR nurse*).ab,ti,id. OR psychotherapy/ OR counseling psychology/ OR psychology/ OR psychotherapeutic counseling/ OR marriage counseling/ OR couples therapy/ OR family therapy/ OR counseling/ OR Social Workers/ |

| | |
|---|--|
| 3 | <p>((evidence adj based) OR (random* adj assign*) OR (random* adj control*) OR (random* adj control* adj trial*) OR (evaluat* adj study) OR (evaluat* adj studies) OR RCT OR (quasi adj experimen*) OR (case adj study) OR (case adj studies) OR (case adj control)OR (qualitative adj studies) OR (qualitative adj study) OR (qualitative adj analysis) OR (content adj analysis)</p> <p>OR (participat* adj observation*) OR (focus adj group*) OR</p> <p>(systematic adj review*) OR (meta adj analysis) OR (mixed adj metaanal*) OR review* OR metasynth* OR meta-synth* OR interview* OR case-study OR case-studies).ab,ti,id. OR Qualitative research/ OR Methodology/ OR Quantitative Methods/OR Quasi Experimental methods/ OR Empirical Methods/ OR Evidence based practice/</p> |
| 4 | <p>((cultur* adj competenc*) OR (cultur* adj competent*) OR (multicultur* adj competenc*) OR (multicultur* adj competent*) OR (cultur* adj responsive) OR (cultur* adj treatment) OR (cross-cultur* adj competent*) OR (cross-cultur* adj competenc*) OR (cultur* adj diver*) OR ethnic* OR cross-ethnic* OR crosscultur* OR cross-cultur* OR diver* OR immigrant OR international OR cultur*).ab,ti,id. OR multiculturalism/ OR cultural sensitivity/ OR Cross Cultural Differences/ OR Diversity/</p> |
| 5 | 1 AND 2 |
| 6 | 1 AND 2 AND 3 |
| 7 | 1 AND 2 AND 4 |
| 8 | 1 AND 2 AND 3 AND 4 |

A hand search will also be conducted on key journals within the last year to ensure early on-line publications not yet indexed are identified related to supervision and psychotherapy including: The Clinical Supervisor, Counselor Education and Supervision, Social Work Supervision, and Journal of Therapy, Consultation and Training. A search of the online table of contents of each of these journals over the last two-year period will be conducted in the event that relevant articles have not been indexed into searched databases. Hopewell et al.

(2002) noted that utilising a hand search can identify research that can be missed or not indexed ensuring the broadest inclusion possible.

The review authors will check reference lists from key articles, literature review and systematic reviews in order to identify additional references. Furthermore, leading authors in the field will be contacted requesting that they share their reference lists, and RefWorks or EndNote bibliographic management files relevant to the study's objectives. These reference materials will be examined to identify any possible studies not previously identified through our employed search strategies.

Google Scholar is a large and valuable database but searches here are difficult to track. Different from our approach with more structured databases, we therefore plan to run and record individual terms and their combinations. Since Google Scholar ranks the records by relevancy and by citation counts (Google, n.d.), we will limit inspection of citations to the top 10 pages of retrieval which may vary depending on discovery of relevant citations. For example, the search strategy, professional supervision AND family therapy AND multicultural, results in a total of over 33,000 hits out of which the first 10 pages includes about 100 of the most relevant records. Beyond these pages, the relevancy of the records diminishes significantly.

Finally, records retrieved from each resource will be organized using the Endnote bibliographic management application.

Data collection and analysis

Selection of studies

Two members of the review team (RA and AM) will independently review titles and abstracts to exclude studies that are clearly irrelevant to the current reviews objectives. We will retrieve the full text reports for those that are not obviously ineligible and for those in which reviewers disagreed. Should disagreements persist a third member of the review team (LH) will be consulted to resolve the disagreement.

In the second phase the two team members (RA and AM) will review the full text versions of the articles and determine eligibility based on inclusion criteria. Any disagreements regarding exclusion/inclusion will again be resolved via consultation with the third team member (LH). The rationale for exclusion will be documented for each study that is retrieved in full text. All studies will be uploaded to Endnote, which will be employed as our citation management tool. The overall search and screening process will be presented in a PRISMA diagram (Moher et al., 2009).

Data extraction and management

Two members of the review team (RA and AM) will independently code and extract pertinent information from each included study. A piloted extraction tool will be employed to guide the extraction process (see Appendix A for the codebook). Using this codebook in relation to RCTs, QEDs and qualitative research data will be extracted on the characteristics of the participants (i.e., age, gender, professional discipline), experience level of the supervisee, experience level of the supervisor), format/nature of the supervision (e.g., one-on-one, group, video review, live supervision), model/approach of mental health therapy (e.g., CBT, systemic therapy, psychodynamic therapy), characteristics of the study (i.e., publication date, geographic location of the study, sample size, intervention setting,) and characteristics of the patients/clients regarding objectives of the study pertaining to patient/client health outcomes (i.e., age, gender, culture, presenting concern, outcome of treatment). Additional information will be extracted for RCTs and QEDs related to intervention and control conditions, measures and results. For qualitative studies additional materials will be included on the code sheet, including the specific methodology employed in the study (i.e., phenomenology, grounded theory, ethnography, case study), along with a verbatim samples that illustrate the key themes from the findings sections that illuminate factors that contribute to the success or failure of supervision. Extracted information will be stored electronically within an Excel spreadsheet. Coding disagreements will be settled via discussion and inclusion of an independent third party (LH). Management and analysis of the extracted data will be conducted in RevMan5, NVivo11, and SPSS 21.

To achieve the primary objective of this review we will include studies that utilize experimental or quasi-experimental research designs. One example of a randomized controlled trial (RCT) conducted by Weck and colleagues (2015) who studied the effectiveness of supervision methods for master's level mental health professionals (n=46) training in CBT. Specifically, this study compared live supervision using real-time computer-messaging between supervisor (i.e., treatment condition) and supervisee versus post-intervention video-based review (i.e., control). Employing third party reviewers blinded to the conditions, Weck and colleagues concluded that the therapeutic alliance and trainee competence were significant greater for the computer enhanced live supervision. However, patient outcomes following the mental health intervention as measured using the Beck Depression Inventory-II (Beck & Steer, 1996) and the Brief Symptom Inventory (Derogatis & Melisaratos, 1983) were shown to have no significant difference between the intervention and control conditions. Another RCT likely eligible for inclusion in the proposed review was conducted by Kass and colleagues (1984). This study examined the impact of supervision methods on psychiatry residents' (n=83) competence for mental health interventions and the

impact on patient outcomes from these interventions. In this three arm RCT, supervision with a senior psychiatrist was compared to peer-based consultation among residents and with a no-supervision/consultation condition. Kass and colleagues found that psychiatry residents' self-reports indicated both supervision with a senior psychiatry resident and peer-supervision to be a positive educational experience. However, patient outcomes measured using resident-completed Likert-based surveys implemented post-intervention revealed that residents who received supervision from a senior psychiatrist rated patient outcomes more positively than those participating in peer-based consultation, which in turn was rated higher than no supervision/consultation processes.

In order to achieve the secondary objectives of this review we will include qualitative studies in order to gain insight into reasons why supervision may succeed or fail to enhance clinical and multicultural competencies of mental health professionals and/or health outcomes of patients.

Multiple time points for measurement

We anticipate that most eligible studies will include both pre-test and posttest measurement. Pretest will likely occur prior to supervision commencing while we anticipate that some posttest measurement to occur following the completion of block of supervision associated with a training milestone, such as the completion of an educational rotation/practicum or following a training regimen in a particular model of therapy (e.g., CBT). Due to the idiosyncratic nature of ongoing supervision unrelated to new trainees and/or a specific training regimen, we anticipate that studies will likely conduct posttest measurements and measurement at various follow-up time points. Therefore, they will be analysed in the following groups: post-test (within 2 weeks following the completion of supervision); medium term follow-up (3-6months following the completion of supervision) and long term follow-up (6+ months following the completion of supervision). Posttest measures will be considered distinct from follow-up measures, as their purpose is to determine the immediate effects of supervision versus follow-up conditions that are intended to measure the lasting impact of supervision. These groups (i.e., pretest, posttest and follow-up) will be considered distinct and will not be pooled for analysis.

Assessment of risk of bias in included studies

In relation to the first objective we will assess for potential risk of bias present and/or the quality of the RCTs and QEDs in our review using the Cochrane Risk of Bias Assessment Tool. Two members of the review team (RA and AM) will independently assess for risk of bias (i.e., high, low, unclear) by exploring each RCT and QED in relation to the following

conditions/factors as outlined by Higgins and Green (2011) and Mathews and colleagues (2015):

For RCTs

- Random sequence generation: Is the method employed to generate the random sequence described with sufficient detail to allow an assessment of the adequacy of the randomization of the sequence process.

For QEDs

- Sequence generation: Is the method employed to generate the allocation sequence described with sufficient detail to allow an assessment of the likelihood that a comparable baseline could be produced following a comparable process.

For RCTs and QEDs

- Allocation Concealment: Is the concealment method for allocation (i.e., treatment vs: control) described with adequate detail to allow reviewers to determine whether group allocation could have been predicted/determined by participants involved in the individual studies.
- Blinding of study participants & researchers: Are participants/researchers blinded to participant's group allocation/membership.
- Blinding of outcome assessment: Are outcome assessors (i.e., third party evaluators of competency using the OSCE) aware of the participants' group status.
- Incomplete outcome data: A rationale is provided for incomplete data (e.g., attrition, exclusions and withdraws, number-needed-to-treat analysis) and, where applicable imputation methods were articulated.
- Selective Reporting: Outcomes were reported *a priori* with sufficient details to allow reviewers to adjudicate their completeness. Our assessment for potential reporting bias will be further commented on in a subsection section of the current protocol.
- Other Biases: Are other possible biases present? For example, is there a conflict of interest bias, where the researcher also has a possible fiduciary connection to the training/supervision model or psychosocial intervention?

In relation to the second objective of the review we will employ the Critical Appraisal Skills Programme (CASP) checklist pertaining to qualitative research to guide our inclusion of

qualitative research studies. For the purpose of this study, qualitative studies will be rated as low when the study fails to receive an affirmative (i.e., Yes) response to all of the following a) the appropriateness of the research design to answers the study's question(s), b) the appropriateness of sampling strategy to achieve the research aims and c) were the analyses suitably rigorous (see Appendix A). Those studies rated as low on the CASP Qualitative research checklist will not be included into the final tables or narrative summary.

Measures of treatment effect

Continuous data/Outcomes

Continuous data are likely to arise from studies employing measures such as the Child Behavior Checklist, Behavior Assessment System for Children, Parent Stress Index and Outcome Rating Scale. Due to a high likelihood that included studies will employ a diverse array of outcome measures, we will report effects size differences as standardised mean differences (SMD). For studies of relatively small sample sizes, we will use a Hedge's \hat{g} corrected SMD. Principal investigators will be contacted by the reviewers to obtain means and standard deviations when this information is not provided. Calculation of 95% confidence intervals for the SMD will be calculated from available data (i.e., F-ratios and t-values) in order to facilitate a comparison of the effects of the different treatments (see Lipsey & Wilson, 2001).

Dichotomous data/Outcomes

For dichotomous data we will calculate risk ratio. For primary measures, such as ratings of the OSCE, outcomes can be pass/fail in relation to competency and therefore we would report on the risk of failing. In relation to the study's objectives pertaining to client/patient health outcomes, several client/patient outcome measures will produce scores separating normative levels from clinically significant levels. This will allow for a risk ratio to be expressed articulating the risk of patients/clients demonstrating clinical significant levels concern versus normal/subclinical levels of concern.

Studies with multiple interventions

Where multiple intervention groups occur within a study with a single comparison condition, such as the case where one method of supervision is compared head-to-head with a second method (e.g., one-on-one supervision versus group supervision), and then to a control condition, such as no supervision, we will include for analysis only that intervention condition deemed most relevant to the primary objective (i.e., mental health competence or culture competency). The determination of which intervention condition is most relevant will be made by the team during the data-extraction coding phase. Should both intervention

conditions be deemed relevant to the review's objectives, the intervention coded for inclusion in pooling will be that which is deemed to have the least potential for bias (i.e., low risk of bias).

Unit of analysis issues

We will be aware of and take steps to control for factors that arise from differing units of analysis. Challenges associated with units of analysis may arise as a result of the different levels of analyses associated with the different levels of supervision practice. For example, one-to-one supervision will produce individual level data, while group supervision may produce outcomes that speak to group-level processes (i.e., team cohesion). Furthermore, in relation to the possible impact of supervised mental health practice/psychotherapy on client/patient health outcomes, it is likely that some studies will move beyond the health comes of individuals to examine the well-being of parent-child dyads and/or family systems. These various levels of analyses will not be pooled, with a separate meta-analytic procedure being conducted for each level, should there be sufficient studies.

Dealing with missing data

We will assess for missing data and attrition in our review. In the event of missing or incomplete data, the reviewers will endeavour to obtain missing information from the principle investigators (PIs) of studies containing missing information. Should authors provide any missing data, this will be imputed such that all studies undergoing imputation will be documented and a sensitivity analysis will be conducted to determine any effects in these studies compared to studies without imputation.

Assessment of heterogeneity

The reviewers will examine variables related to population (e.g., professional discipline of trainee), and intervention, including medium of supervision (e.g., individual or group-based supervision) and method of supervision (e.g., case consultation, video review or live supervision) in order to determine if variation in study distributions exist. We will assess for study heterogeneity using the the Q-statistic and its related p value to determine diversity of intervention effects; Tau² with confidence interval to determine the magnitude of variation between studies; and the I² statistic with confidence intervals to estimate what proportion of variability in effect estimates may be due to chance (Higgins & Green, 2011).

Assessment of reporting biases

We will assess for reporting bias by employing funnel plots if a sufficient number of studies are identified, suggested by Higgins and Green (2011) to be at least 10, in order to examine whether symmetry in the relationship between effect estimates and standard error exists. In

the event that visual inspection reveals an asymmetrical funnel plot (i.e., effect size clustering on one side of the funnel), suggesting possible publication bias, we will proceed to utilize a trim and fill method suggested by Sutton and colleagues (2000).

Data synthesis

In relation to our first objective we will employ RevMan5 to support our data synthesis and meta-analytic procedures. ES will be entered into RevMan5 to calculate the standardized effects size, as well as standard error. For continuous outcomes we will use Hedge's \hat{g} as the measure of effect sizes (ES). For binary outcomes we will convert log odds to Cohen's d then into Hedge's \hat{g} . In order to calculate any meta-analysis we will use the Hedge's \hat{g} . f any additional conversions as required.

We anticipate that the studies included in our review will use a diverse array of outcome measures, supervision methods, mediums and models, as well as include a wide range of professional backgrounds and experience levels for the trainees. Therefore, due to these likely sources of heterogeneity across studies we will employ a random effects model for pooling results. We will examine whether heterogeneity exists across the RCTs and QEDs studies and combine only those studies that are homogeneous. Separate analyses will also be conducted for general competencies in the provision of mental health therapy, distinct from cultural competencies. Conceptually different studies, such as those examining only increases in knowledge, distinct from a clinical skills will be pooled separately with distinct analyses being conducted. Where feasible and appropriate we will also pool the results pertaining to patient/client outcomes for those involved in supervised mental health interventions. As previously discussed, separate meta-analyses will also be conducted for each distinct unit of analysis (i.e., individual versus group/family level outcomes). The results of each outcome will be reported using Forest plots with 95% confidence intervals.

Summary of Findings Tables

In relation to our primary objectives our work will be presented in a summary table format that describes the key findings from the included studies and adjudicates the quality of the evidence according to the GRADE classification system (see Guyatt et al., 2011; Guyatt et al., 2008). The GRADE systems will guide our classification of included studies according to their relative rigour into high, moderate, lower and very low quality. Specifically, GRADE criteria will be applied by examining each study's methodology/research design, risk of bias, imprecision, inconsistency, indirectness and magnitude of effect (see Guyatt et al., 2011).

In relation to the review's second objective we will present a separate table summarizing the findings from the meta-synthesis of qualitative research. This table will present a taxonomy

of themes that emerged across like-qualitative methods, including a presentation of key differences and similarities in sampling, analysis and findings from the various qualitative methods (i.e., phenomenology, ethnography, and grounded theory). In relation to the findings from the studies, differences and similarities of themes, including meta-themes (themes that run through studies of similar method) will be presenting including, where possible key quotes from the various studies will be used to illuminate the key theme being described.

Subgroup analysis and investigation of heterogeneity

There are likely meaningful subgroups within the population under investigation in the current review, making it important to estimate the effects for these groups. Although there are number of potential moderators, two in particular may be most salient including the medium of supervision (i.e., one-one-one, triadic and small group) and the supervisee training level, which may best be divided into those who have achieved their professional licensure/registration and those who are still students or residents and are not yet expected to conduct independent (i.e., unsupervised) practice as mental health professionals. Important subgroups may include studies specifically pertaining to cultural competencies and those pertaining to general clinical competencies. Findings from these moderators and subgroup analyses will be presented in tabular format including the pooled estimates of the percentage point impacts of different supervision mediums (i.e., individual, triadic and small group) by supervisee training level. Should there be sufficient numbers of studies, suggested by (Higgins & Green, 2006) to be at least 10 studies per grouping variable, there is impetus to conduct a moderator analysis on these two salient subgroups and on the influence of supervision method (e.g., case consultation, tape review, live supervision) as well as the duration of the supervision. This analysis will be conducted using a meta-regression as a direct test of the differences between subgroups and moderator influences on the mean effect (Littell, Corcoran & Pillai, 2008).

Sensitivity analysis

A sensitivity analysis will be conducted in order to test the soundness of decisions made throughout the current review. Specifically we will compare our models through re-analyses to determine whether studies with a) missing data and/or use of imputation, b) lack an intent-to-treat analysis, c) lack rigorous controls for assignment at baseline (i.e., methodologically weaker QEDs) and d) studies with high-risk of bias may impact the overall results of the analysis. These studies will systematically be dropped and not included in the re-analysis, allowing the authors to gain insight into whether potential methodological issues

and/or bias is possibly influencing the results (Littell, Corcoran, & Pillai, 2008). In such cases these differential findings will be reported.

Treatment of qualitative research

Qualitative research will be included in this review in relation to the second objective and will be informed by Sani and Shlonsky's (2012), interpretative approach to meta-synthesis for qualitative research. Following our screening process, included qualitative studies will be imported into Nvivo, with links to full-text articles. For analytic purposes, included studies will be grouped separately by specific methodology (e.g., phenomenological studies, ethnographies, grounded theory, case study) to allow us to complete a separate synthesis for each specific qualitative methodology. Our goal in using interpretative meta-synthesis will be to integrate the materials from conceptually congruent methodologies in order to identify taxonomies or central themes throughout the studies in each methodology, with particular emphasis on synthesize themes, to best illuminate reasons that clinical supervision may succeed or fail in improving clinical and/or multi-cultural competencies and/or enhanced patient outcomes. We will engage in an iterative process of developing and applying a coding framework through the analysis of the full-text articles, moving systematically through the coding of studies of like-methodology. Harnessing a constant comparative approach we will apply the coding framework to each subsequent article, allowing new codes to emerge from the qualitative findings while also ensuring that we revisit previously reviewed articles to examine the presence of any newly identified themes. After all studies from one method have been successfully coded, coding will commence on the next methods until complete. Coding will be done by two team members with expertise in qualitative research and the application of Nvivo for thematic coding. Results of themes will be presented in a summary table with narratives from each method, as well as a grand narrative related to divergent and convergent themes across the various methodologies. We will utilize this meta-synthesis as an enhancement model (Saini & Shlonsky, 2012) that will help us to interpret the findings from the quantitative meta-analytic findings conducted on RCT and QEDs. Qualitative methods will be valuable to understand the key mechanisms underpinning supervisory interventions that contribute to supervision being effective or that enhance psychotherapists' competency and improve patient outcomes.

REFERENCES

Abidin, R. R. (2012). *Parenting Stress Index* (4th ed.). Lutz, FL: PAR.

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington, VT: University of Vermont Department of Psychiatry.

- Achenbach, T.M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- American Association for Marriage and Family Therapy (2014). *Approved supervision designation: Standards handbook*. Retrieved from https://dx5br1z4f6n0k.cloudfront.net/imis15/Documents/Approved_Supervisor_Handbook_2014.pdf
- American Psychological Association. Board of Educational Affairs (2014). *Guidelines for clinical supervision for health service psychologists*. Retrieved from: <http://www.apa.org/about/policy/guidelines-supervision.pdf>
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.
- Aponte, H. J., Powell, F. D., Brooks, S., Watson, M. F., Litzke, C., Lawless, J., & Johnson, E. (2009). Training the person of the therapist in an academic setting. *Journal of Marital and Family Therapy*, 35(4), 381–394. doi: 10.1111/j.1752-0606.2009.00123.x
- Association for Counselor Education and Supervision. (2011). *Best Practices in Clinical Supervision*.
- 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(3), 317-331. doi: 10.1080/10503300500268524
- Barak, M. M. E., Travis, D. J., Pyun, H., & Xie, B. (2009). The impact of supervision on worker outcome: A metaanalysis. *Social Service Review* 83(1), 3-32. doi.org/10.1086/599028
- Beck, A. T., & Steer, R. A. (1990). *Manual for the Beck Anxiety Inventory*. San Antonio, TX: Psychological Corporation.
- Beck A. T., Steer R. A., & Brown G. K. (1996) *Manual for the Beck Depression Inventory-II*". San Antonio, TX: Psychological Corporation.

- Berger, C., & Mizrahi, T. (2001). An evolving paradigm of supervision within a changing health care environment. *Social Work in Health Care, 32*(4), 1-18. doi: 10.1300/J010v32n04_01
- Bernard, J. M., & Goodyear, R. K. (2014). *Fundamentals of clinical supervision* (5th ed.). Boston: Pearson.
- Bidell, M. P. (2005). The Sexual Orientation Counselor Competency Scale: Assessing attitudes, skills, and knowledge, of counselors working with lesbian, gay, and bisexual clients. *Counselor Education and Supervision, 44* (4), 267-279. doi: 10.1002/j.1556-6978.2005.tb01755.x
- Bogo, M., & McKnight, K. (2006). Clinical supervision in social work. *The Clinical Supervisor, 24*(1-2), 49-67. doi.org/10.1300/j001v24n01_04
- 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. doi.org/10.1111/j.1365-2850.2007.01021.x
- British Association for Counselling and Psychotherapy (2016). *What is Counselling & Psychotherapy?* Retrieved from <http://www.bacp.co.uk/crs/Training/whatiscounselling.php>
- Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: Do therapists know how competent they are?. *Behavioural Cognitive Psychotherapy, 36*(5) 581-587. doi: 10.1017/S1352465808004438
- Brunero, S., & Stein-Pabury, J. (2008). The effectiveness of clinical supervision in nursing: an evidence-based literature review. *Australian Journal of Advanced Nursing, 25* (3), 86-94.
- Butterworth, T., Bell, C., Jackson, C., Pajnkihar, M. (2007). Wicked spell or magic bullet? A review of the clinical supervision literature 2001-2007. *Nurse Education Today, 28*, 264-272. doi: 10.1016/j.nedt.2007.05.004
- Callahan, J. L., Almstrom, C. M., Swift, J. K., Borja, S. E., & Heath, C. J. (2009). Exploring the contribution of supervisors to intervention outcomes. *Training and Education in Professional Psychology, 3*(2), 72-7. doi: 10.1037/a0014294

- Carlton, M. (2002). Short Form Supervision Satisfaction Questionnaire. In C. E. Munson (Ed.), *Handbook of clinical social work supervision* third edition (p. 579). Binghamton, NY: The Haworth Social Work Practice Press.
- Carpenter, J., Webb, C. M., & Bostock, L. (2013). The surprisingly weak evidence base for supervision: Findings from a systematic review in child welfare practice. *Children and Youth services Review* 35, 1843-1853. doi.org/10.1016/j.chilyouth.2013.08.014
- Cheon, H., & Murphy, M. J. (2007). The self-of-the-therapist awakened. *Journal of Feminist Family Therapy*, 19, 1–16. doi.org/10.1300/j086v19n01_01
- Council for Accreditation of Counseling and Related Educational Programs. (2009). *2009 CACREP Accreditation Manual*. Alexandria, VA: Author. Retrieved from <http://www.cacrep.org/wp-content/uploads/2013/12/2009-Standards.pdf>
- Coulter, S. (2011). Systemic family therapy for families who have experienced trauma: A randomised controlled trial. *British Journal of Social Work*, 41(3), 502-519. doi: 10.1093/bjsw/bcq132
- Critical Appraisal Skills Programme [CASP] (2013). Qualitative Research Checklist. Oxford, Uk: CASP. Retrieved from http://media.wix.com/ugd/dded87_29c5b002d99342f788c6ac670e49f274.pdf
- Cummins, A. (2009). Clinical supervision: The way forward? A review of the literature. *Nurse Education in Practice*, 9, 215-220. doi.org/10.1016/j.nepr.2008.10.009
- D'Andrea, M., Daniels, J., & Heck, R. (1990). *The Multicultural Awareness-Knowledge-Skills Survey*. Honolulu, HI: University of Hawaii Manoa.
- Dawson, M., Phillips, B., & Leggat, S. (2013). Clinical supervision for allied health professionals: A systematic review. *Journal of Allied Health*, 42(2), 65-73.
- Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: an introductory report. *Psychological Medicine*, 13, 595–605. DOI:10.1017/S0033291700048017.
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal*, 315(7109), 629-634. doi: 10.1136/bmj.315.7109.629

- Elliott, R. (1998). Editor's introduction: A guide to the empirically supported treatments controversy. *Psychotherapy Research, 8*(2), 115-125.
doi: 10.1080/10503309812331332257
- Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). Tire McMaster Family Assessment Device. *Journal of Marital and Family Therapy, 9* (2), 171-180.
- Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). Tire McMaster Family Assessment Device. *Journal of Marital and Family Therapy, 9* (2), 171-180.
- Falender, C. A., Burnes, T. R., & Ellis, M. V. (2013). Multicultural clinical supervision and benchmarks: Empirical support informing practice and supervisor training. *The Counseling Psychologist, 41*(1), 8-27. doi: 10.1177/0011000012438417
- Falender, C. A., Ellis, M. V., & Burnes, T. (2013). Response to reactions to major contribution: Multicultural clinical supervision and Benchmarks. *The Counseling Psychologist, 41*, 140-151. doi: 10.1177/0011000012464061.
- Falender, C. A., Shafranske, E. P., & Falicov, C. (2014). Diversity and multiculturalism in supervision. In C. A. Falender, E. P. Shafranske, & C. Falicov (Eds.), *Multiculturalism and diversity in clinical supervision: A competency-based approach* (pp. 3-28). Washington, DC: American Psychological Association.
- Falicov, C.J. (2014). Psychotherapy and supervision as cultural encounters: The MECA framework. In C.A. Falender, E.P. Shafransky, & C.J. Falicov (Eds.), *Multiculturalism and diversity in clinical supervision: A competency-based approach*. Washington, DC: American Psychological Association.
- Farnan J. M., Petty L. A., Georgitis E., Martin, S., Chiu, E., Prochaska, M., & Arora, V. M. (2012). A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Academic Medicine, 87*(4), 428–442. doi: 10.1097/ACM.0b013e31824822cc
- Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item-response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78* (2), 350-365. doi: 10.1037//0022-3514.78.2.350
- Frank, G., (1984). The Boulder Model: History, rationale, and critique. *Professional Psychology: Research and Practice, 15*(3), 417–435. doi: 10.1037/0735-7028.15.3.417

- Francke, A. L. & Graaff, F. M. (2012). The effects of group supervision of nurses: A systematic review. *International Journal of Nursing Studies*, 49(9), 1165-1179. doi: 10.1016/j.ijnurstu.2011.11.012
- Funk, J. L. & Rogge, R. D. (2007). Testing the ruler with item response theory: Increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology*, 21 (4), 572-583. doi: 10.1037/0893-3200.21.4.572
- Gambrill, E. (2010). Evidence-informed practice: Antidote to propaganda in the helping professions?. *Research on Social Work Practice*, 20(3), 302-320. doi: 10.1177/1049731509347879
- Geertz, C. (1973). Thick description: toward an interpretive theory of culture. In *The interpretation of cultures: selected essays* (pp. 3–30). New York: Basic Books.
- Google. (n.d.). About Google Scholar. Retrieved from <https://scholar.google.com/scholar/about.html>
- Guyatt, G. H., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J.,...Schunemann, H. J. (2011). GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64(4), 383-394. doi: 10.1016/j.jclinepi.2010.04.026.
- Guyatt, G. H., Oxman, A. D., Vist, G., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schunemann, H.J., for the GRADE Working Group (2008). Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendation. *British Medical Journal*, 336(7650), 924-926. doi: 10.1136/bmj.39489.470347.ad
- Hansen, N. D., Randazzo, K. V., Schwartz, A., Marshall, M., Kalis, D., Frazier, R., . . . Norvig, G. (2006). Do we practice what we preach? An exploratory survey of multicultural psychotherapy competencies. *Professional Psychology, Research and Practice*, 37, 66–74. doi:10.1037/0735-7028.37.1.66
- Harden, R. M., Stevenson, M., Downie, W.W., & Wilson, G.M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, 1, 447-451.

- Hendrick, S. S. (1988). A generic measure of relationship satisfaction. *Journal of Marriage and Family*, 50 (1), 93-98. <http://www.jstor.org/stable/352430>
- Higgins, J.P.T., & Green, S. (Eds.) (2011). *Cochrane handbook for systematic reviews of interventions*. The Cochrane Collaboration. Retrieved from <http://handbook.cochrane.org/>
- Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C....Liddle, H.A. (2008). Assessing fidelity in individual and family therapy for adolescent substance abuse. *Journal of Substance Abuse Treatment*, 35(2), 137-147. doi: 10.1016/j.jsat.2007.09.002
- Holmes, D., Murray, S. J., Perron, A., & Rail, G. (2006). Deconstructing the evidence-based discourse in health sciences: Truth, power and fascism. *International Journal of Evidence-Based Healthcare*, 4(3), 180-186. doi: 10.1111/j.1479-6988.2006.00041.x
- Hopewell, S., Clarke, M., Lusher, A., Lefebvre, C., & Westby, M. (2002). A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. *Statistics in Medicine*, 21, 1625-1634.
- Kadushin, A. (2002). *Supervision in social work* (4th ed.). New York, NY: University Press.
- Kass, F., Charles, E., Eagle, P., & Yadofsky, B. (1984). Assessing the impact of supervision on outpatient evaluation in psychiatry. *Quality Review Bulletin*, 10(1), 3-5.
- Kilminster, S. M., & Jolly, B. C. (2000). Effective supervision in clinical practice settings: A literature review. *Medical Education*, 34(10), 827–840. doi: 10.1046/j.1365-2923.2000.00758.x
- Lee, R. E., & Everett, C. A. (2004). *The integrative family therapy supervisor*. New York: Routledge.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage Publications.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis. A pocket guide for social work research methods*. New York: Oxford University Press. doi: 10.1177/1049731508318552.

- Locke, H. J., & Wallace, K. M. (1959). Short Marital Adjustment and Prediction Tests: Their reliability and validity. *Marriage and Family Living*, 21 (3), 251–255. doi: <http://dx.doi.org/10.2307/348022>
- Mason, J. (1995). *Cultural Competence Self-Assessment Questionnaire: A manual for users*. Portland, OR: Portland State University, Research and Training Center on Family Support and Children's Mental Health.
- Mathews, B., Walsh, K., Coe, S., Kenny, M. C., & Vagenas, D. (2015). *Protocol for a systematic review: Child protection training for professionals to improve reporting of child abuse and neglect*. Campbell Collaboration. Retrieved from <http://www.campbellcollaboration.org/lib/project/304/>
- Maxwell, J. A. (2011). *A realist approach for qualitative research*. Thousand Oaks, CA: Sage.
- Miller, S. D., & Duncan, B. L. (2000). *The Outcome Rating Scale*. Chicago, IL: Author.
- Miller, W. R., & Tonigan, J. S. (1996). Assessing drinkers' motivations for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychology of Addictive Behaviors*, 10 (2), 81–89. doi: 10.1037/0893-164X.10.2.81
- Milne, D. L. (2007). An empirical definition of clinical supervision. *British Journal of Clinical Psychology*, 46, 437– 447. doi:10.1348/ 014466507X197415
- Milne, D. (2009). *Evidence-based clinical supervision: Principles and practices*. London, UK: Wiley-Blackwell.
- Milne, D., Aylott, H., Fitzpatrick, H., & Ellis, M. V. (2008). How does clinical supervision work? Using a “best evidence synthesis” approach to construct a basic model of supervision. *The Clinical Supervisor*, 27(2). 170-90. doi: 10.1080/07325220802487915
- Milne, D. and James, I. (2000), A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology*, 39(2), 111–127. doi: 10.1348/014466500163149
- Milne, D. L., Sheikh, A. I., Pattison, S., & Wilkinson, A. (2011). Evidence-based training for clinical supervisors: A systematic review of 11 controlled studies. *The Clinical Supervisor*, 30(1), 53-71. doi: 10.1080/07325223.2011.564955

- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal*, 339(b2535). doi: 10.1136/bmj.b2535
- Multicultural Mental Health Australia (2002). *Cultural awareness tool*. Sydney Australia: Multicultural Mental Health Australia. ISBN 0 9581735 0 8
- National Association of Social Work. (2013). *Best practice standards on social work supervision*. Retrieved from <http://www.naswdc.org/practice/naswstandards/supervisionstandards2013.pdf>
- Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family*, 45 (1), 141–151. doi: 10.2307/351302
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Ponterotto, J., Gretchen, D., Utsey, S., Rieger, B., & Austin, R. (2002). A revision of the Multicultural Counseling Awareness Scale. *Journal of Multicultural Counseling and Development*, 30 (3), 153–180. doi: 10.1002/j.2161-1912.2002.tb00489.x
- Radloff, L. S. (1977). The CES-D Scale: A self report depression scale for research in the general population. *Applied Psychological Measurements*, 1 (3), 385-401. Doi: 10.1177.014662167700100306
- Ponterotto, J., Gretchen, D., Utsey, S., Rieger, B., & Austin, R. (2002). A revision of the Multicultural Counseling Awareness Scale. *Journal of Multicultural Counseling and Development*, 30 (3), 153–180. doi: 10.1002/j.2161-1912.2002.tb00489.x
- Radloff, L. S. (1977). The CES-D Scale: A self report depression scale for research in the general population. *Applied Psychological Measurements*, 1 (3), 385-401. doi: 10.1177.014662167700100306
- Ratts, M. J., Singh, A. A., Nassar-McMillan, S., Butler, S. K., & McCullough, J. R. (2015). *Multicultural and social justice counseling competencies*. Retrieved from http://www.multiculturalcounseling.org/index.php?option=com_content&view=article&id=205:amcd-endorses-multicultural-and-social-justice-counseling-competencies&catid=1:latest&Itemid=123

- Reiser, R. P., & Milne, D. (2012). Supervising cognitive-behavioral psychotherapy: Pressing needs, impressing possibilities. *Journal of Contemporary Psychotherapy, 42*, 161-171. doi: 10.1007/s10879-011-9200-6.
- Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.
- Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children (BASC)*. Circle Pines, MN: American Guidance Services.
- Roehling, P. V., & Robin, A. L. (1986). Development and validation of the Family Beliefs Inventory: A measure of unrealistic beliefs among parents and adolescents. *Journal of Consulting and Clinical Psychology, 54* (5), 693-697. doi: 10.1037//0022-006X.54.5.693
- Royal Australian and New Zealand College of Psychiatrists (2012). Training. *2012 Fellowship Program: Regulations, policies and procedures*. Retrieved from <https://www.ranzcp.org/Files/PreFellowship/2012-Fellowship-Program/RPP-TRAINING.aspx>
- Rust, J., Bennun, I., Crowe, M., & Golombok, S. (1986). The Golombok Rust Inventory of Marital State (GRIMS). *Sexual and Marital Therapy, 1* (1), 55-60. doi: 10.1080/02674658608407680
- Saini, M., & Shlonsky, A. (2012). *Systematic synthesis of qualitative research: A pocket guide for social work research methods*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195387216.001.0001
- Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed.). London: Routledge.
- Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology, 77*(3), 410-421. doi: 10.1037/a0013788
- Slife, B. D., Wiggins, B. J., & Graham, J. T. (2005). Avoiding an EST monopoly: Toward a pluralism of philosophies and methods. *Journal of Contemporary Psychotherapy, 35*, 83-97. doi: 10.1007/s10879-005-0805-5
- Sodowsky, G. R., Taffe, R. C., Gutkin, T. B., & Wise, S. L. (1994). Development of the Multicultural Counseling Inventory: A self-report measure of multicultural

- competences. *Journal of Counseling Psychology*, 41 (2), 137–148.
doi:10.1037/0022-0167.41.2.137
- Spanier, G. B. (1989). *Manual for the Dyadic Adjustment Scale*. North Tonawanda, NY: Multi-Health Systems.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Sodowsky, G. R., Taffe, R. C., Gutkin, T. B., & Wise, S. L. (1994). Development of the Multicultural Counseling Inventory: A self-report measure of multicultural competences. *Journal of Counseling Psychology*, 41 (2), 137–148.
doi:10.1037/0022-0167.41.2.137
- Spanier, G. B. (1989). *Manual for the Dyadic Adjustment Scale*. North Tonawanda, NY: Multi-Health Systems.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Sprenkle, D. H., & Piercy, F. P. (2005) *Research methods in family therapy*. New York, NY: Guilford Press.
- Staller, K. M. (2006). Railroads, runaways, & researchers: Returning evidence rhetoric to its practice base. *Qualitative Inquiry*, 12(3), 503-522. doi: 10.1177/1077800406286524
- Startup, M., Jackson, M., & Pearce E. (2002). Assessing therapist adherence to cognitive-behaviour therapy for psychosis. *Behavioural and Cognitive Psychotherapy*, 30, 329–339. doi: 10.1017/S1352465802003077
- Sterne, J. A., Egger, M., & Smith, G. D. (2001). Systematic review in health care: investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal*, 323(7304), 101-105. doi: 10.1136/bmj.323.7304.101
- Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical test and prevalence in the literature. *Journal of Clinical Epidemiology*, 53(11), 1119-1120. doi: 10.1016/S0895-4356(00)00242-0
- Sterne, J. A., Hernan, M. A., Reeves, B. C., Savoic, J., Berkman, N. D., ... Higgins, J. P. T. (2016). *The risk of bias in non-randomized studies-of interventions (ROBINS-I) assessment tool*. Cochrane Collaboration. Available from <https://53e94a67-a62cb3a1a-sites.google.com/site/riskofbiastool/ROBINS-I%20tool%20template%207mar2016.pdf?attachauth=ANoY7cpeaL9srVTZX9eq1N>

oDkF1yddciZu6vN8sMVkgn3rxJ2RVCmgtJdrYS9Z6KiauGA4iIDx7cySviKp5y-
GFFTH8NmCc2frMCF-
skPsbIHclwtY_G6zXmx682pO_YWpI1_phjrzDQ3Cb0rUWTdV6Kt7WXGhcnEsgB
kDiTOWrkCZTJXQUwcAWWJdo1vDH9_FZ1W5Nb8dCo5P3KbdNSggw2VOeomV
Y3QaEf83nMM7jD1v_WNrmtEiNllaX5pfa--9OTQNWxVUea&attredirects=0

- Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R., & Jones, D.R. (2000). Empirical assessments of effect of publication bias on meta-analyses. *British Medical Journal*, *320*, 1574-1577. doi: 10.1136/bmj.320.7249.1574
- Timm, T. M., & Blow, A. J. (1999). Self-of-the-therapist work: A balance between removing restraints and identifying resources. *Contemporary Family Therapy*, *21*(3), 331–350. doi: 10.1023/A:1021960315503
- Todd, T. C., & Storm, C. L. (2002). Thoughts on the evolution of MFT Supervision. In T.C. Todd & C.L. Storm (Eds.) *The complete systemic supervisor: Context, philosophy, and pragmatics* (pp. 1-16). Needham Heights, MA: Allyn & Bacon.
- Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal?. *American Psychologist*, *69*(30), 218-229. doi:10.1037/a0035099
- Vallance, K. (2005). Exploring counsellor perceptions of the impact of counselling supervision on clients. *Counselling and Psychotherapy Research*, *5*(2), 107-110. doi: 10.1080/17441690500211106
- Wainwright, N.A. (2010). The development of the Leeds Alliance in Supervision Scale (LASS): A brief sessional measure of the supervisory alliance. The University of Leeds. Available from http://practicumsupport-psych.sites.olt.ubc.ca/files/2012/08/Nigel_Antony_Wainwright_DClinCpsychol_THESES_2010_-2.pdf
- Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research Therapy*, *47*(2), 119-27. doi:10.1016/j.brat.2008.10.018
- Watkins, C. E., Jr. (2011). Does psychotherapy supervision contribute to patient outcomes? Considering thirty years of research. *The Clinical Supervisor*, *30*(2), 235-256. doi: 10.1080/07325223.2011.619417
- Watkins, C. E., Jr. (2012). Psychotherapy supervision in the new millennium: Competency-based, evidence-based, particularized, and energized. *Journal of Contemporary Psychotherapy*, *42*, 193–203. doi: 10.1007/s10879-011-9202-4

- Weck, F., Jakob, M., Neng, J.M.B., Hofling, V., Grikscheit, F., & Bohus, M. (2016). The effects of bug-in-the-eye supervision on therapeutic alliance and therapist competence in cognitive-behavioural therapy: A randomized controlled trial. *Clinical Psychology and Psychotherapy*, 23(5), 386-396. doi: 10.1002/cpp.1968.
- Wendt, D. Jr. (2006). The unevaluated framework of APA's policy on evidence-based practice in psychology (EBPP). *The New School Psychology Bulletin*, 4(1), 89-99.
- Wheeler, S., & Richards, K. (2007). The impact of clinical supervision on counselors and therapists, their practice, and their clients. A systematic review of the literature. *Counseling and Psychotherapy Research*, 7(1), 54–65. doi: 10.1080/14733140601185274
- Winstanley, J. (2000). *Manchester Clinical Supervision Scale user guide*. Sydney, Australia: Osman Consulting Pty.
- Yardley, L. (2000) Dilemmas in qualitative health research. *Psychology & Health*, 15, 215-228. doi: 10.1080/08870440008400302
- Zimmerman, M., Chelminski, I., McGlinchey, J., & Posternk, M. (2008). A clinically useful depression outcome scale. *Comprehensive Psychiatry*, 49 (2), 131-140. doi:10.1016/j.comppsy.2007.10.006

REVIEW AUTHORS

Lead review author:

The lead author is the person who develops and co-ordinates the review team, discusses and assigns roles for individual members of the review team, liaises with the editorial base and takes responsibility for the on-going updates of the review.

Name: Robert Allan

Title: Assistant Professor

Affiliation: University of Colorado Denver

Address: Campus Box 106

PO Box 173364

City, State, Province or County: Denver, Colorado

Postal Code: 80217-3364

Country: United States

Phone: 303-315-4913

Email: robert.allan@ucdenver.edu

Co-author(s):

Name: Alan McLuckie

Title: Assistant Professor

Affiliation: University of Calgary

Address: 2500 University Dr, NW

City, State, Province or County: Calgary, Alberta

Postal Code: T2N 1N4

Country: Canada

Phone: 403-220-2926

Email: amclucki@ucalgary.ca

Co-author(s):

Name: Lillian Hoffecker

Title: Research Librarian

Affiliation: Health Sciences Library,
University of Colorado Anschutz Medical Campus

Address: 12950 E. Montview Blvd.
Mail Stop A003

City, State, Province or County: Aurora, Colorado

Postal Code: 80045

Country: United States

Phone: 303-724-2124

Email: Lillian.Hoffecker@ucdenver.edu

ROLES AND RESPONSIBILITIES

Please give brief description of content and methodological expertise within the review team. The recommended optimal review team composition includes at least one person on the review team who has content expertise, at least one person who has methodological expertise and at least one person who has statistical expertise. It is also recommended to have one person with information retrieval expertise.

Who is responsible for the below areas? Please list their names:

- Content: Robert Allan, Alan McLuckie
- Systematic review methods: Lillian Hoffecker working with all authors
- Statistical analysis: Alan McLuckie
- Information retrieval: Lillian Hoffecker working with all authors

Sources of Support

University support for research assistants, no external funding.

DECLARATIONS OF INTEREST

None known.

PRELIMINARY TIMEFRAME

| Activities | Date of completion |
|--|---------------------------|
| Searches for studies | October 31, 2017 |
| Pilot testing of inclusion criteria | November 30, 2017 |
| Pilot testing of study codes and data collection | January 31, 2018 |
| Extraction of data from research reports | March 31, 2018 |
| Statistical Analysis | May 31, 2018 |
| Preparation of report | October 31, 2018 |

PLANS FOR UPDATING THE REVIEW

Robert Allan, Alan McLuckie, and Lillian Hoffecker will be responsible for updating this review. We anticipate that this review will be updated as we accumulate new evidence and/or at least every three years.

AUTHOR DECLARATION

Authors' responsibilities

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group

also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

You accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group.

Publication in the Campbell Library

The support of the Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review, and subsequent updates in the Campbell Library. The Campbell Collaboration places no restrictions on publication of the findings of a Campbell systematic review in a more abbreviated form as a journal article either before or after the publication of the monograph version in *Campbell Systematic Reviews*. Some journals, however, have restrictions that preclude publication of findings that have been, or will be, reported elsewhere and authors considering publication in such a journal should be aware of possible conflict with publication of the monograph version in *Campbell Systematic Reviews*. Publication in a journal after publication or in press status in *Campbell Systematic Reviews* should acknowledge the Campbell version and include a citation to it. Note that systematic reviews published in *Campbell Systematic Reviews* and co-registered with the Cochrane Collaboration may have additional requirements or restrictions for co-publication. Review authors accept responsibility for meeting any co-publication requirements.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Robert Allan

Date: 6 July 2017

APPENDIX A. CODEBOOK FOR QUANTITATIVE AND QUALITATIVE STUDIES

| Code | Description |
|------------------------------------|--|
| Reviewer name | |
| Date coding started | |
| Date coding completed | |
| Identification of reference | |
| Study ID | |
| Title | |
| First Author | |
| Journal name | |
| Year of publication | |
| Type of publication | <i>1-Peer-reviewed journal, 2-Book chapter/book, 3-dissertation/thesis, 4-other (specify)</i> |
| Methods of data collection | |
| Types of study designs | <i>Randomized controlled trials, Quasi-experimental (will be included in relation to objective #1) and Qualitative designs will be included for objective #2</i> |
| Study duration | |
| Study date | |
| Aim of the study | |
| Experimental designs (RCTs) | |
| Random sequence generation | <i>High risk/low risk/unclear risk</i> |
| Allocation concealment | <i>High risk/low risk/unclear risk</i> |
| Blinding of participants | <i>High risk/low risk/unclear risk</i> |
| Blinding of outcome assessment | <i>High risk/low risk/unclear risk</i> |
| Incomplete outcome data | <i>High risk/low risk/unclear risk</i> |
| Selective outcome reporting | <i>High risk/low risk/unclear risk</i> |
| Other sources of bias | <i>Bias due to problems not covered above</i> |
| Statistical method | <i>Report statistical method and if this method is adequate</i> |
| Quasi-experimental designs | |
| Allocation mechanism | <i>Is the allocation mechanism appropriate to generate equivalent groups?</i> |
| Group equivalence | <i>Are all likely relevant confounders taken into account in the analysis?</i> |
| Hawthorne effects | <i>Are differences in outcomes across groups influenced by participant motivation as a result of program implementation and, or monitoring?</i> |
| Spill-over | <i>Is the program influencing the outcome of the individuals in the control group?</i> |
| Selective methods of analysis | <i>Is the method of analysis or specification model used by the author selectively chosen?</i> |
| Other sources of bias | <i>Bias due to problems not covered above</i> |
| Statistical method | <i>Report statistical method and if this method is adequate?</i> |
| Qualitative Designs | |

| | |
|---|--|
| Is the design appropriate to address the research aims? | <i>Yes/No/Can't tell</i> |
| Specify the research methodology used | <i>List the specific methodology (Phenomenology, grounded theory, ethnography, case study)</i> |
| Was recruitment appropriate | <i>Yes/No/Can't tell</i> |
| Was the data analysis sufficiently rigorous? | <i>Yes/No/Can't tell</i> |
| Study population | |
| Sample size | |
| Treatment group N | |
| Control group N | |
| Geographic location of the study | |
| Types of mental health professionals | <i>1- psychotherapists, 2-psychologists, 3-counsellors, 4-couple/marital and family therapists, 5-social workers, 6-psychiatrists, 7-nurses, 8-pastoral counsellors, and 9-trainees in each of these professions, 10-other (i.e., multidisciplinary)</i> |
| Types of settings | <i>1-Universities/colleges, 2-child welfare services, 3-psychiatry institutions, 4-training programs, 5-web-based programs, 6-hospitals, 7-counselling agencies, and 8-physicians' offices</i> |
| Types of experiences | <i>1-residencies/internships/practicums, 2-student supervision, and 3-clinical supervision across mental health professions.</i> |
| Supervisor clinical experience level and background | <i>1-Education level, 2-qualifications, 3- prior relevant experiences</i> |
| Supervisee clinical experience level and background | <i>1-Education level, 2-qualifications, 3-prior relevant experiences</i> |
| Specific supervisory training (i.e., specialist/trained supervisor versus non-specialist) | <i>1-yes, 2-no, 3-other</i> |
| Time limited supervision | <i>Graduate program internship, practicum and/or residency supervision</i> |
| On-going supervision | <i>Clinical and professional supervision that will be an on-going experience throughout a helping professional's career</i> |
| Age | |
| Gender | |
| Culture / social-location | |
| Ethical considerations | |
| Intervention | |
| Intervention of interest (clinical supervision) | <i>Aim, content, format, mixed or single intervention</i> |
| Medium of Supervision | <i>1-one-on-one supervision, 2- triadic (1 supervisor and supervisee dyad), 3-small group supervision</i> |
| Methods of clinical supervision | <i>1-peer-based consultation (reflecting teams or online), 2-facilitated team-based consultation, 3-case discussion, 4-video/audio review or 9-live presentation/demonstrations (with call-in and/or with bug in ear)</i> |
| Models of clinical supervision | <i>Developmental, psychodynamic, solution-focused, narrative</i> |

| | |
|---|---|
| Time-frame and dose of clinical supervision (i.e., regular and ongoing versus intermittent) | <i>Frequency, duration, intensity, and types/methods of supervision implemented (i.e., weekly)</i> |
| Cultural sensitivity | <i>Clinical supervision that considers the language, socio-cultural values and traditions of the supervisee</i> |
| Relationship between the supervisor and supervisee | |
| Recruitment of supervisors and supervisee's | <i>What were the procedures used to recruit participants to partake in the study?</i> |
| Attrition | <i>Were the drop-out rates or number of participants lost during the course of the intervention or follow-up process measured/recorded? Was an intent-to-treat analysis utilized? Yes/No?</i> |
| Fidelity/integrity | <i>Was the intervention delivered as intended? Was there a method for measuring/gauging fidelity?</i> |
| Adaptation | <i>Was the type/method of supervision intentionally or purposefully changed during intervention implementation, from the original standard, to enhance effectiveness</i> |
| Level of supervisee participation/expectation during supervision | |
| Outcomes | |
| Primary outcome measures | <i>1-enhance clinical competencies of mental health professionals and psychotherapists, 2-enhance multicultural clinical competencies of mental health professionals and psychotherapists, 3- the effectiveness of the supervisory relationship/process</i> |
| Secondary outcome measures | <p><i>To measure the effectiveness of supervision to improve patient/client outcomes</i></p> <ul style="list-style-type: none"> <i>Information on patient/client care outcomes may include the age, gender, and clinical state of the client/patient group and the frequency, intensity of their treatment and the presence of other concurrent and/or treatment regimens</i> |
| Effectiveness | <i>Did the study measure the effectiveness of clinical supervision? 1-yes, 2-no, 3-unclear</i> |
| Methods of assessing outcome measures | <i>1-how the researchers monitored and evaluated the supervisee level of competency, attitude, skill, knowledge and clinical behaviour pre/post supervision? 2-how the researchers monitored and evaluated therapy/treatment provided to client/patients pre/post supervisee supervision, 3-how the researchers monitored and evaluated the and client/patient outcomes pre/post supervisee supervision</i> |
| Validity and reliability of outcome measures used | |
| Methods and duration of follow-up for non-respondents | |
| Timing of outcome assessment | <i>Frequency, length of follow-up</i> |

| | |
|--|--|
| What were the intervals of measurement? | <i>Were there pretest, posttest and follow-up measures? For posttest specify the length of time following the implementation of the supervision. For follow-up specify the length of time since the completion of the supervision process.</i> |
| Unit of analysis | <i>1 - individual level, 2 – group level (related to supervision group, 3-family level (related to patient outcomes)</i> |
| Results | |
| Effect measure | <i>Effect measure, confidence interval</i> |
| Outcome measure of treatment group | |
| Outcome measure of control group | |
| How was the outcome measure reported? | 1- supervisor report, 2- supervisee report, 3- client report, 4- Other (specify) |
| How are the units of analysis assigned to conditions? | 1- Random, 2- Quasi-random, 3- Non-random 4- Unclear |
| Effect size type (<i>include page number data was found</i>) | 1- Treatment/control Post-test comparison 2- Treatment/control group Follow-up comparison |
| Does the study report means and standard deviations for each condition? | |
| Treatment group N | |
| Treatment group mean | |
| Treatment group SD | |
| Control group N | |
| Control group mean | |
| Control group SD | |
| If means and SD not reported, list what quantitative information is provided | |
| Qualitative Studies | |
| Sample of key themes pertinent to illuminating factors that help or hinder supervision | Provide a verbatim narrative of the key themes identified in the results |