Working alliance and outcome effectiveness in videoconferencing psychotherapy: a systematic review and non-inferiority meta-analysis

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Clinical Psychology &amp; Psychotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CPP-1780.R1</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Research Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>17-May-2018</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Norwood, Carl; The University of Nottingham, Division of Psychiatry and Applied Psychology, Moghaddam, Nima; University of Lincoln, College of Social Sciences, Malins, Sam; University of Nottingham, CLAHRC EM, Sabin-Farrell, Rachel; University of Nottingham, Doctorate in Clinical Psychology</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Videoconferencing psychotherapy, VCP, working alliance, meta-analysis, systematic review</td>
</tr>
</tbody>
</table>
Abstract

Videoconferencing psychotherapy (VCP) – the remote delivery of psychotherapy via secure video link – is an innovative way of delivering psychotherapy, which has the potential to overcome many of the regularly cited barriers to accessing psychological treatment.

However, some debate exists as to whether an adequate working alliance can be formed between therapist and client, when therapy is delivered through such a medium. The presented article is a systematic literature review and two meta-analyses aimed at answering the questions: is it possible to develop adequate working alliance in VCP? And is outcome equivalence possible between VCP and face-to-face delivery? Twelve studies were identified which met inclusion/exclusion criteria, all of which demonstrated good working alliance and outcome for VCP. Meta-analyses showed that working alliance in VCP was inferior to face-to-face delivery (standardised mean difference [SMD] = -0.30; 95% confidence interval [CI] [-0.67, 0.07], p = .11; with the lower-bound of the CI extending beyond the non-inferiority margin [-0.50]), but that target symptom reduction was non-inferior (SMD = -0.03; 95% CI [-0.45, 0.40], p = .90; CI within the non-inferiority margin [-.50]). These results are discussed and directions for future research recommended.

Keywords: Videoconferencing psychotherapy, VCP, working alliance, meta-analysis, systematic review
Introduction

The working alliance – defined by Bordin (1979) as the collaboration between client and therapist across the three domains of goals (agreed outcomes to work towards), tasks (the expectations and actions of client and therapist, directed towards the achievement of goals), and bonds (attachment between therapist and client) – has been extensively investigated within psychotherapy, and has been found to be related to outcome across multiple meta-analyses (Horvath, Del Re, Fluckiger, & Symonds, 2011; Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000), with meta-synthesised themes identified for creating and maintaining a strong relationship (Noyce & Simpson, 2016). These studies, however, have been conducted solely focusing on the face-to-face delivery of therapy, with much less being known about the significance of the working alliance when therapy is delivered through alternative mediums.

The use of remote psychotherapy is gaining popularity (Hollis et al., 2015) as a way for services to best meet the needs of clients, and is also recommended by the UK government (HM Government, 2011). One such remote psychotherapy delivery method is videoconferencing psychotherapy (VCP) – the delivery of psychotherapy via secure video link. This has the advantage of potentially overcoming traditionally cited barriers to accessing treatment, such as transport (Harvey & Gumport, 2015), perceived stigma (Clement et al., 2015; Sirey et al., 2001), and insufficient service and staff provision (Alvidrez & Azocar, 1999; Lousada, Weisz, Hudson, & Swain, 2015).

Despite these advantages, a view seems to remain that delivering a service via VCP would in some way hinder the working alliance. Indeed, it has been pointed out that ‘conventional wisdom’ suggests that face-to-face services will facilitate a better alliance (Bee et al., 2008), perhaps due to the fact that any alternative delivery method would reduce the interpersonal...
richness of an interaction by limiting the availability and readability of eye contact, physical expression, and body posture (Wootton, Yellowles, & McLaren, 2003) – with a poor video connection conceivably exacerbating these limitations. Perhaps it is due to these beliefs that psychologists appear reluctant to fully embrace VCP, being typically reluctant to endorse VCP as a stand-alone intervention (Mora, Nevid, & Chaplin, 2008). Reservations about security, and a lack of formal support or training were cited as reasons for their reluctance (Vincent, Barnett, Killpack, Sehgal, & Swinden, 2017).

It is possibly a consequence of a lack of interpersonal richness in the interaction, or a result of psychologists’ apparent reluctance to embrace VCP, but a study conducted by Rees and Stone (2005) found that psychologists rated sessions delivered by VCP lower for working alliance than those delivered face-to-face – even when the actual sessions were identical in nature. However, empirical evidence exists to show that therapy delivered by VCP can be effective (Vogel et al., 2014), and has comparable outcomes to face-to-face treatment (Dunstan & Tooth, 2012; Strachan, Gros, Ruggiero, Lejuez, & Acierno, 2012), with yet further research suggesting equivalence of working alliance between VCP and face-to-face delivery (Simpson & Reid, 2014).

This raises interesting questions: is working alliance actually poorer in VCP? And, if so, is it possible to have equivalent outcome through VCP delivery? The present systematic literature review and meta-analysis aimed to broadly answer these questions by achieving the following objectives: (1) to review existing VCP literature that incorporates measures of working alliance, and (2) to conduct non-inferiority meta-analyses comparing the working alliance and outcome between VCP and face-to-face delivery. It is hypothesised that non-inferiority will be seen in terms of both working alliance and outcome between VCP and face-to-face delivery.
Method

Inclusion/exclusion Criteria

Due to VCP being a relatively new area of research, it was judged that to limit searching by research design or quality would potentially exclude relevant research in the area. Therefore, no studies were excluded by research design or quality.

Studies were eligible for inclusion if they:

- Reported data from an adult population (aged ≥18)
- Used cognitive-behavioural therapy (CBT) or contextual cognitive-behavioural therapy (CCBT; Hayes, Villatte, Levin, & Hildebrandt, 2011) – such that evidence was collated from a relatively homogeneous group of psychotherapies (those building on evidence and techniques from strands of behavioural and cognitive therapy).
- Reported pre- and post-treatment data for symptom severity (a continuous outcome relating to the target difficulty of the population)
- Reported at least one measure of working alliance
- Were published in English

Studies were excluded if they:

- Used a group-based intervention (to retain a focus on dyadic therapist-client alliance, rather than introducing the potentially confounding variable of group cohesion).
- Used VCP as a supplemental intervention or peripheral component.

Searching

To identify articles relevant for review, five databases were searched (PsycINFO, PsycARTICLES, MEDLINE, CINAHL Complete, and PubMed). Three concepts were
required to be combined for the search: working alliance, videoconferencing, and contextual CBT.

The working alliance search was conducted using the prefixes ‘working’, ‘helping’, and ‘therapeutic’ with the suffixes ‘alliance’, ‘relationship’, and ‘bond’ attached to each. The video conferencing search combined individual searches on the terms ‘video conferencing’, ‘skype’, ‘video conference’, ‘video consultation’, ‘telemedicine’, ‘telehealth’, ‘telecare’, and ‘assistive technology’. Contextual CBT was searched using the terms ‘cognitive behavior therapy’, ‘CBT’, ‘cognitive behav* therap*’, ‘third wave’, ‘acceptance and commitment therapy’, ‘ACT’, ‘dialectical behav* therap*’, ‘DBT’, ‘mindfulness’, and ‘behavi* therap*’. The three concepts were combined in a final search, and all titles and abstracts were screened for inclusion. Further potential studies were identified from the reference lists of articles screened as potentially acceptable, and known literature reviews in remote psychotherapy (Backhaus et al., 2012; Richardson, Frueh, Grubaugh, Egede, & Elhai, 2009; Simpson, 2009; Simpson, & Reid, 2014). The final search was conducted on 19th April 2018 – see Figure 1 for an outline of the selection process. Of the articles accessed in full, the reasons for exclusion were: no quantifiable working alliance measure (Frueh et al., 2007; Gros, Yoder, Tuerk, Lozano, & Acierno, 2010; Simpson, 2001; Strachan et al., 2012), outcome data not reported pre- and post-intervention (Day & Schneider, 2002; Goetter, Herbert, Forman, Yuen, & Thomas, 2014; Simpson, Deans, & Brebner, 2001), intervention not delivered by VCP (Herbst et al., 2016), group therapy (Morland et al., 2014), therapy not being the main intervention of focus (Olden et al., 2017), VCP being supplemented by phone calls (Vogel et al., 2014), data reported in another study already included in the review (Bouchard et al., 2000), and the type of therapy being eclectic or unspecified (Ghosh, McLaren, & Watson, 1997; Simpson, Bell, Knox, & Mitchell, 2005). Twelve studies met the criteria and were
included in the review (Bouchard et al., 2004; Ertelt et al., 2010; Germain, Marchand, Bouchard, Drouin, & Guay, 2009; Germain, Marchand, Bouchard, Guay, & Drouin, 2010; Himle et al., 2006; Lichstein et al., 2013; Manchanda & McLaren, 1998; Mitchell et al., 2008; Morland et al., 2015; Stefan & David, 2013; Stubbings, Rees, Roberts, & Kane, 2013; Yuen et al., 2013).

Meta-analysis

Two meta-analyses were conducted aimed at answering the questions (1) is working alliance in VCP non-inferior to face-to-face delivery? And (2) is outcome in VCP non-inferior to face-to-face delivery? For the purposes of the meta-analyses, only RCTs with a face-to-face delivery control group were included. The meta-analyses were conducted using Review Manager 5 software.

For the purposes of this review and meta-analysis, the studies conducted by Mitchell and colleagues (2008), and Ertelt and colleagues (2010), have been grouped together as they report on the same dataset, with Mitchell and colleagues (2008) reporting outcome data, and Ertelt and colleagues (2010) reporting data on the working alliance. The same is also true of two other studies in the review, with Germain, Marchand, Bouchard, Drouin, and Guay (2009) reporting outcome data, and Germain, Marchand, Bouchard, Guay, and Drouin (2010) reporting working alliance data from the same dataset.

For the meta-analysis comparing working alliance between delivery methods, a total working alliance score was calculated for each condition within each study. This was done by calculating the mean working alliance score within a specific condition, with the pooled standard deviation being calculated using Cohen’s formulae:

$$\sqrt{\frac{(SD_1^2 + SD_2^2 + SD_j^2)}{j}}$$
To meta-analyse outcome, change was calculated on the primary outcome measure by subtracting the pre-intervention score from the post-intervention score, meaning that a positive number would represent an increase in scores pre- to post-intervention (deterioration) and a minus number would indicate a decrease in scores pre- to post-intervention (improvement). If a study reported multiple outcome measures, the primary outcome measure was selected for use in the meta-analysis. In papers with multiple primary outcomes (or where no primary outcome is designated), a single measure which mapped on to the difficulty experienced by the studies’ population was selected.

As studies appraised to be of lower quality present greater risk of bias, sensitivity analyses were conducted to test the potential influence of study quality on heterogeneity and pooled effect estimates (by systematically rerunning meta-analyses whilst excluding the study rated to be of lowest quality).

**Non-inferiority margin and meta-analysis.** For meta-analysis of outcome, the non-inferiority margin was set at $\Delta$ Cohen’s $d = 0.50$, which was based on the smallest of two criterion-values: (1) the total estimated outcome-effect of face-to-face delivery (as the standard of treatment), and (2) the largest clinically acceptable difference (degree of inferiority) for VCP as compared with face-to-face delivery. Criterion 2 should be a smaller value than criterion 1, such that any clinically acceptable difference can be understood to show that VCP is effective (in addition to being clinically non-inferior to the standard of face-to-face delivery). For criterion 1, a total effect estimate was derived from a meta-analytic review comparing CBT to inactive control conditions for anxiety disorders (Hofmann & Smits, 2008); this estimate was considered apt because most studies in the present review are of CBT-variant psychotherapies for anxiety-related outcomes. Hofmann and Smits (2008) estimated the lower bound of the 95% confidence interval (CI) of this total effect size to be 0.56 (standardised mean-difference
between groups; Hedges’ $g$). For criterion 2, we use recurrent evidence for the clinical meaningfulness of a standardised mean-difference of 0.50 (Norman, Sloan, & Wyrwich, 2003; Wise, 2004) – which represents a smaller value than the assumed total effect-size, as is desirable. The test criterion for non-inferiority was that the upper bound of the 95% CI of the mean difference should fall within $\Delta$; thus, with 95% probability, the standardised mean-difference between VCP and face-to-face delivery had to be smaller than 0.50. For meta-analysis of working alliance, the non-inferiority margin was again set at $\Delta$ Cohen’s $d = 0.5$ – with the same test-criterion for non-inferiority (this time applied to the lower bound, due to direction of desirable responding). In this case, the focus was on clinically acceptable difference (Criterion 2 alone), as total effect estimates (from comparison against inactive control conditions) are not logically available for working alliance measures. Use of this test-criterion provided parity with the test-criterion applied to outcome measures and is consistent with practice in previous non-inferiority trials of psychotherapy interventions (e.g., Hedman et al., 2011).

Results

Data Abstraction

For each study, data were extracted pertaining to: authors, year of publication, location, study design, population difficulty, sample size, intervention, number of sessions, working alliance measure, working alliance rater and session taken, working alliance score, and change pre- to post-treatment on the primary outcome measure. This information can be seen in Table 1.

Along with this information, each study was quality-assessed. This presented a challenge due to the present review and meta-analysis including studies of various designs. Whilst it is important to assess studies for quality and potential sources of bias, there is no widely-
accepted tool which can be used across study designs (Katrak, Bialocerkowski, Massy-Westropp, Kumar, & Grimmer, 2004) and using multiple design-specific appraisal tools can make it difficult to compare evidence across studies (Crowe & Sheppard, 2011). However, guidelines do exist for when appraising research from multiple study designs. It has been recommended this take three stages (Hawker, Payne, Kerr, Hardey, & Powell, 2002): (1) assessment of relevance to the review, (2) data extraction, and (3) appraisal of methodological rigour. As such, an appraisal tool based on these recommendations and further informed by Critical Appraisal Skills Programme checklists for specific research designs was used, with the results presented in Table 2 (data extraction is not presented in Table 2, as this has been presented separately in Table 1). In order to assess the inter-rater reliability of the applied quality assessment, 25% of the studies (selected purposively, to reflect a range of study designs) were independently rated by two authors (CN and NM). The mean kappa coefficient across items was 1.00, indicating ‘perfect’ agreement overall (Viera & Garrett, 2005).

Methodological Rigour

Of the RCTs in this review (Ertelt et al., 2010; Mitchell et al., 2008; Morland et al., 2015; Stefan & David, 2013; Stubbings, et al., 2013), two give further details of how randomisation was achieved (Ertelt et al., 2010; Mitchell et al., 2008; Stubbings et al., 2013), with one of these (Ertelt et al., 2010; Mitchell et al., 2008) further explaining randomisation was stratified by diagnosis and current antidepressant use to balance the conditions on these variables. The other two RCTs (Morland et al., 2015; Stefan & David, 2013) stated participants were randomised, but gave no further information regarding this. Further to not making clear a randomisation strategy, it is also not clear in one study (Stefan & David,
whether the groups were similar at the start, and how many clients completed the intervention.

Two of the RCTs (Morland et al., 2015; Stubbings et al., 2013) appear to have had differences in the groups at the start of the trial. Morland and colleagues (2015) randomised the allocation of war veterans without stratifying randomisation based on their service history, which resulted in a disparity between conditions on the duration of the experienced difficulty. Whereas Stubbings and colleagues (2013) did not limit their study to a single primary diagnosis (and did not stratify randomisation by presenting problem). This resulted in some participants receiving a manualised CBT treatment (if they had a difficulty lending itself to such) and some receiving an individualised intervention (if their primary diagnosis did not lend itself to a manualised treatment).

Of the two N-RCTs in this review (Bouchard et al., 2004; Germain et al., 2009; Germain et al., 2010), neither had similar samples at the start of the trial; with one (Bouchard et al., 2004) having many more comorbid diagnoses in the VCP group compared to face-to-face (10 of 11 participants, and 4 of 10 participants respectively), and the other study having a population varying by the type of trauma they experienced – with 50% of the VCP group reporting a trauma of ‘physical or sexual aggression’, compared to 28.1% of the face-to-face group (Germain et al., 2009; Germain et al., 2010).

It is also unclear in both N-RCTs whether the participants are treated equally, other than the experimental manipulation. In one study (Bouchard et al., 2004) half of the participants in each condition were randomised to a 3-month wait before treatment (however, due to small sample numbers, the data from immediate treatment and waitlist were collapsed within each condition – meaning that within each condition participants were treated differently, but this was controlled for between conditions).
al., 2010) recruited for the face-to-face condition from a ‘local’ site and for the VCP condition from both a ‘remote’ and ‘local’ site, with those recruited for VCP from the ‘local’ site having treatment in the same building as their therapist was situated. It is stated that ‘measures’ were taken to ensure they never met face-to-face, but it is not stated what these measures were or how they could impact on the treatment experience for participants.

Of the ten studies in this review, seven (Bouchard et al., 2004; Ertelt et al., 2010; Germain et al., 2009; Germain et al., 2010; Lichstein et al., 2013; Mitchell et al., 2008; Morland et al., 2015; Stubbings, et al., 2013; Yuen et al., 2013) state inclusion/exclusion criteria, and the other three (Himle et al., 2006; Manchanda & McLaren, 1998; Stefan & David, 2013) do not, whilst seven (Bouchard et al., 2004; Ertelt et al., 2010; Himle et al., 2006; Lichstein et al., 2013; Mitchell et al., 2008; Morland et al., 2015; Stubbings, et al., 2013; Yuen et al., 2013) offer follow-up data post intervention and the other three (Germain et al., 2009; Germain et al., 2010; Manchanda & McLaren, 1998; Stefan & David, 2013) do not.

**VCP delivery**

A concern regarding VCP delivery is that it opens the possibility for technical difficulties, which could potentially hinder therapy. As such it is important to acknowledge the different technologies used in the studies reviewed, particularly as the studies range from 1998-2015, and technology has advanced a great deal in this time.

Of the studies reviewed, eight state the technology used (Bouchard et al., 2004; Germain et al., 2009; Germain et al., 2010; Himle et al., 2006; Lichstein et al., 2013; Manchanda & McLaren, 1998; Stefan & David, 2013; Stubbings et al., 2013; Yuen et al., 2013) and two do not (Ertelt et al., 2010; Mitchell et al., 2008; Morland et al., 2015). Two of the studies report using extra technology: one using a fax machine to send through homework (Bouchard et al.,
2004), and another using a telephone on the ‘hands free’ setting to transmit audio
(Manchanda & McLaren, 1998) as the videoconferencing technology used in this study could
only display picture. One study (Stefan & David, 2013) used notably different technology to
deliver VCP, with a three-dimensional holographic image being produced.

In terms of VCP location in the reviewed studies, six stated VCP took place in a clinic or
research building (Bouchard et al., 2004; Ertelt et al., 2010; Germain et al., 2009; Germain et
al., 2010; Lichstein et al., 2013; Manchanda & McLaren, 1998; Mitchell et al., 2008;
Stubbings et al., 2013), three did not state where VCP took place (Himle et al., 2006;
Morland et al., 2015; Stefan & David, 2013), and only one study used a VCP intervention set
in the home (Yuen et al., 2013).

Working alliance

All studies included in this review used either the Working Alliance Inventory (WAI;
Horvath & Greenberg, 1989) or the Working Alliance Inventory – Short Form (WAI-SF;
Tracey & Kokotovic, 1989) to measure working alliance. The WAI was used in six studies
(Bouchard et al., 2004; Ertelt et al., 2010; Germain et al., 2009; Germain et al., 2010; Himle
et al., 2006; Lichstein et al., 2013; Manchanda & McLaren, 1998; Mitchell et al., 2008) and
the WAI-SF was used in the other four (Morland et al., 2015; Stefan & David, 2013;
Stubbings et al., 2013; Yuen et al., 2013). Both have high internal consistency, with the WAI
showing internal consistency of .87 to .93 dependant on the rater (Horvath & Greenberg,
1989), and the WAI-SF showing .95 to .98 dependant on the rater (Tracey & Kokotovic,
1989).

Different raters of working alliance were used across the different studies, with five studies
taking a rating of working alliance from just the client (Bouchard et al., 2004; Germain et al.,
2009; Germain et al., 2010; Himle et al., 2006; Stefan & David, 2013; Yuen et al., 2013), one
from an observer (Lichstein et al., 2013), and the remaining four from both the client and
therapist (Ertelt et al., 2010; Manchanda & McLaren, 1998; Mitchell et al., 2008; Morland et
al., 2015; Stubbings et al., 2013). This is potentially important due to a phenomenon known
as the halo effect (Horvath et al., 2011) – a trend seeing higher correlations between working
alliance and outcome if both are rated by the same person. As all the studies in this review
used self-report outcome measures, there is a risk of the halo effect impacting on the
relationship between working alliance and outcome in those five studies which just measure
working alliance from the clients’ perspective (Bouchard et al., 2004; Germain et al., 2009;
Germain et al., 2010; Himle et al., 2006; Stefan & David, 2013; Yuen et al., 2013).

Another consideration is when working alliance scores were recorded. It has been proposed
that the relation between working alliance and outcome grows in magnitude the later alliance
is recorded (Horvath et al., 2011), in such a way that working alliance scores are confounded
by prior symptom change such that people are rating therapeutic benefit at later sessions,
rather than working alliance (Crits-Christoph, Connolly Gibbons, Hamilton, Ring-Kurtz, &
Gallop, 2011). In the present review, two studies (Himle et al., 2006; Stubbings et al., 2013)
record working alliance only at the end of treatment, and so run the risk of this phenomenon.
All other studies have at least an early and late working alliance rating, with the exception of
one (Stefan & David, 2013) – which offers only one intervention session and so takes a
recording following this.

Participants

Participants from six of the studies were from a clinical population and had a clear diagnosis
(Bouchard et al., 2004; Ertelt et al., 2010; Germain et al., 2009; Germain et al., 2010;
Mitchell et al., 2008; Morland et al., 2015; Stubbings et al., 2013; Yuen et al., 2013), with a
further three stating participants had a specific difficulty and would benefit from a psychotherapy treatment (Himle et al., 2006; Lichstein et al., 2013; Manchanda & McLaren, 1998). Only one study appeared to not take clinical need into consideration (Stefan & David, 2013), with participants being recruited from a university and awarded course credits.

**Key Findings**

All studies in this review demonstrated strong working alliance in VCP. Of the six reviewed studies with a face-to-face control group, four (Bouchard et al., 2004; Germain et al., 2009; Germain et al., 2010; Stefan & David, 2013; Stubbings, et al., 2013) found that overall working alliance score was non-inferior in the VCP condition compared to face-to-face (though one study [Bouchard et al., 2004] does not explicitly state working alliance data for the face-to-face condition, rather stating in the discussion ‘the comparison between the posttreatment alliance data of the two conditions does not show any difference between the face-to-face and videoconference condition’ [p.21]), with one of these studies (Stefan & David, 2013) finding the VCP group scored significantly higher on the goal subscale of the WAI-SF (Tracey & Kokotovic, 1989). Of those reporting a higher working alliance in the face-to-face group: one study reported no difference between groups in participants’ self-reported working alliance, but a significantly higher working alliance reported by therapists in the face-to-face condition (Ertelt et al., 2010; Mitchell et al., 2008) and one study reported significantly higher working alliance in the face-to-face condition at session two but not at any other time (Morland et al., 2015). A statistically significant increase in working alliance scores over the course of VCP was seen in two studies (Ertelt et al., 2010; Germain et al., 2009; Germain et al., 2010; Mitchell et al., 2008).

Improvement in symptom severity when intervention is delivered by VCP was seen in all studies within this review. This improvement was maintained across all seven studies which
offered a post intervention follow-up (Bouchard et al., 2004; Ertelt et al., 2010; Himle et al., 2006; Lichstein et al., 2013; Mitchell et al., 2008; Morland et al., 2015; Stubbings et al., 2013; Yuen et al., 2013), and VCP symptom reduction was non-inferior to face-to-face across all six studies which offered a face-to-face comparison (Bouchard et al., 2004; Ertelt et al., 2010; Germain et al., 2009; Germain et al., 2010; Mitchell et al., 2008; Morland et al., 2015; Stefan & David, 2013; Stubbings et al., 2013).

Results of Meta-analysis

Figure 2 shows summary statistics for the mean working alliance scores, comparing VCP with face-to-face treatment. The lower limit of the 95% CI for working alliance scores (n = 4; SMD = -0.30; 95% CI [-0.67, 0.07], p = .11; random effects model) fell outside the pre-specified limit of non-inferiority (Δ = -0.50): indicating that, with respect to working alliance, VCP was inferior to face-to-face treatment. Tests of heterogeneity show low heterogeneity for the studies in this analysis (I² = 47%; p = .13).

A sensitivity analysis was conducted to ascertain whether pooled estimates were sensitive to study quality, by excluding the study with the lowest overall quality rating (based on number of criteria met: Stefan & David, 2013). In doing so, evidence of heterogeneity increased (I² = 64%), but substantive results remained the same (n = 3; SMD = -0.31; 95% CI [-0.83, 0.21], p = .24) indicating that the finding of inferior working alliance in VCP (versus face-to-face treatment) was robust to inclusion of studies with variable quality.

A second meta-analysis was conducted comparing symptom reduction across the two conditions. Figure 3 shows summary statistics for this analysis, demonstrating that people who received treatment via VCP had non-inferior symptom reduction compared to people who received a face-to-face treatment (n = 4; SMD = -0.03; 95% CI [-0.45, 0.40], p = .90);
random effects model): The upper limit of the 95% CI for outcome effect-sizes fell within the
pre-specified limit of non-inferiority ($\Delta = 0.50$). Tests of heterogeneity show moderate non-
significant heterogeneity (60%, $p = .06$). Sensitivity analysis was again conducted to
ascertain whether pooled estimates were changed by excluding the study with the lowest
quality rating (Stefan & David, 2013). In doing so, heterogeneity increased ($I^2 = 68\%$) and
VCP symptom reduction remained non-inferior to face-to-face ($n = 3$: SMD = -0.16; 95% CI
[-0.71, 0.39], $p = .56$).

Discussion

A total of 12 articles were selected for review, having met the stated criteria. All of these
articles were reviewed narratively, with data extracted from five included in two meta-
analyses comparing both working alliance and symptom reduction between VCP and face-to-
face delivery. This literature review and meta-analysis aimed to answer the questions: (1) is
working alliance in VCP non-inferior to face-to-face delivery? And (2) is outcome in VCP
non-inferior to face-to-face delivery? From the review and analysis conducted, it would seem
that, in terms of outcome, VCP is non-inferior to face-to-face delivery. However, the working
alliance appears to be inferior when therapy is delivered by video-conferencing.

These results appear to partially support the view expressed by Rees and Stone (2005),
stating that working alliance is viewed as lower in VCP – though this appeared to have little
impact on outcome, with symptom reduction in VCP being non-inferior to face-to-face
delivery. This dissociation makes very interesting reading as it seems to contradict the well-
established finding that poorer working alliance will lead to poorer outcome (Horvath &
Symonds, 1991; Horvath et al., 2011; Martin et al., 2000).
Whilst it is not being suggested that working alliance in the VCP condition was low (as all of the reviewed studies demonstrated what would be called a strong working alliance in VCP), the fact that the pooled effects for working alliance across studies were inferior in VCP is worthy of further thought – indeed, one study (Ertelt et al., 2010; Mitchell et al., 2008) actually demonstrated poorer working alliance in VCP to a statistically significant degree regardless of non-inferiority criteria. Three suggestions are offered for why working alliance may be inferior in VCP: (1) the working alliance is less important when therapy is delivered via VCP (though this seems unlikely, as the working alliance is pan-theoretical [Bordin, 1979; Horvath & Symonds, 1991] and it seems counter-intuitive to suggest that by changing the delivery medium, client and therapist no longer need to share a bond or common goal); (2) working alliance is being rated as lower due to discomfort with the delivery medium; or (3) something unique is happening when therapy is delivered by VCP which compensates for slightly lower working alliance. To that end, it could be any number of factors not measured by the WAI which is responsible for this finding (such as client engagement or motivation). However, a recent meta-synthesis (Noyce & Simpson, 2016) stated ‘empowerment through respect’ as a key aspect in the development of a relationship between client and therapist. It could well be that by accessing treatment remotely (thereby not submitting to the therapists’ will in terms of attending their building, to be seated in their room, per their wishes) client empowerment is increased, which helps facilitate equal outcome in the absence of equal working alliance – it is argued that empowerment in this way may still be seen if the remote therapy took place in a clinical building (as was the case in many of the studies reviewed), as it can be argued that more parity exists between client and therapist as both therapist and client are in independent work-spaces, communicating via their respective consoles, over which they have individual control. It is also worth noting that, were this to be the case, it is something which would not necessarily lend itself to measurement on the WAI, as questions
pertaining to goals, tasks, and bond (the three areas covered by the WAI) would not necessarily capture client empowerment.

The results obtained from the present study add to an ever growing literature pertaining to the use of remote psychotherapy interventions. Available systematic reviews (Scula et al., 2012), and narrative reviews (Berger, 2017) of internet delivered interventions (internet based treatments in which clients follow pre-set exercises typically blended with contact with a therapist, usually over a messaging service rather than using videoconferencing) suggest both equivalency of outcome and alliance is regularly seen. When considered alongside the current findings, the general theme of non-inferior outcome between face-to-face and alternative delivery methods seems consistent, though working alliance seems less clear. It seems almost counter-intuitive to suggest that non-inferior working alliance can be established with minimal therapist contact (typically done over the phone or by email, thus reducing the interpersonal richness of the interaction compared to VCP), yet not seen in VCP. One potential explanation for this difference is in the rater of working alliance. In the presented analysis all but one study collected working alliance rating from both the therapist and the client. However, studies have suggested that therapists rate working alliance lower when therapy is delivered remotely (Rees & Stone, 2005; Berger, 2017), whereas client ratings do not seem to alter (Ruwaard et al., 2007; Ruwaard et al., 2009). Therefore, perhaps the inferiority seen in the presented study could be explained by the amalgamation of the working alliance ratings used. Overall, these findings, in combination with findings of other remote psychotherapy literature, emphasise the potential opportunities to use remote delivery methods, within a menu of care, to better meet clients’ needs and overcome some regularly cited barriers to engagement.
It is also interesting to consider the potential future implications for remotely delivered therapy. As stated earlier in this paper, only adult studies were included – this was due to the judgment that to include child and adolescent studies may skew the data due to younger people stereotypically being more comfortable with the type of technology used to deliver remote therapy. If this were indeed the case, it may be expected to find a stronger working alliance formed in child and adolescent studies of VCP. Unfortunately, however, the literature for children and adolescents use of remotely delivered therapy has received less empirical attention than that of adults (Slone, Reese, & McClellan, 2012), indeed, if the inclusion criteria of the present study were changed to include child and adolescent studies it would yield no further papers for inclusion – though VCP has been found to be effective for the treatment obsessive-compulsive disorder in adolescents (Storch et al., 2011).

It can be reasonably speculated that as familiarity with the technology increases, perhaps the working alliance formed through this medium may do the same. Another consideration is that the way we measure working alliance itself may need to change. If indeed, as suggested above, VCP delivery is tapping in to a construct not currently captured in Bordin’s (1979) conceptualisation of working alliance, it may be that more nuanced and specific measures need to be developed in order to accurately measure the working alliance through different delivery mediums.

Conclusions drawn from this review must be considered in light of the following limitations: firstly, the quality of the studies included in the analysis has to be considered. As VCP is a relatively new area of research, a decision was made to not exclude studies on the basis of quality. However, from the quality assessment conducted (see Table 2) only one of the controlled studies (Ertelt et al., 2010; Mitchell et al., 2008) and one of the uncontrolled studies met all applicable quality criteria (Yuen et al., 2013). This demonstrates the need for
higher quality studies in the area – particularly RCTs, due to their limited number. Attempts were made to at least partially overcome this by conducting a sensitivity analysis (excluding the study with the lowest overall quality rating), which demonstrated consistency of results – inferiority of working alliance in VCP, and non-inferiority of outcome.

Another noted limitation is that of external validity. When reviewing the studies, an interesting paradox became apparent. One of the advantages of using VCP is that it can be delivered in the home. However, from a research perspective, delivery in the home environment potentially introduces lots of confounding variables and so is not always desirable. Interestingly, only one study actually used VCP in the home (though without a control group; Yuen et al., 2013), with the other studies all delivering VCP in a clinical building. Whilst this may improve the internal validity of the studies (by conducting VCP in a controlled environment), it compromises external validity as it is not how the intervention will be delivered in practice.

Added to this point, as the purpose of VCP is to treat clinical populations remotely, this is what the research must reflect. In the present review one study does not treat a clinical population (Stefan & David, 2013), and it is unclear with a further three (Himle et al., 2006; Lichstein et al., 2013; Manchanda & McLaren, 1998). This makes it difficult to draw firm conclusions as to the clinical effectiveness of VCP – although only one of these studies (Stefan & David, 2013) was used in the meta-analyses.

For future research, the following recommendations are made: (1) future studies should endeavour to deliver VCP in the home, to a clinical population, so as to mimic clinical use and improve external validity; (2) future studies would benefit from incorporating alliance-outcome correlations (and other process-outcome associations) to enable more nuanced analysis of factors relating to outcome effectiveness – none of the studies reviewed here
included such data; and (3) multiple measures of working alliance (from different raters) should be obtained across the course of the intervention to ensure alliance is actually being measured (rather than being confounded by prior symptom change or the halo effect).

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
References


outcome and comparison to treatment delivered in person. Behavior Therapy, 42(2), 276-83. https://doi.org/10.1016/j.beth.2010.07.005


https://dx.doi.org/10.1016/j.brat.2005.12.010


https://dx.doi.org/10.1192/bjp.bp.113.142612


http://dx.doi.org/10.1037/0022-0167.36.2.223

http://dx.doi.org/10.1037/0022-0167.38.2.139


https://doi.org/10.1016/S2215-0366(15)00082-6

http://dx.doi.org/10.1258/1357633981931452


http://dx.doi.org/10.1016/j.brat.2008.02.004

http://dx.doi.org/10.1016/j.chb.2008.05.011
http://dx.doi.org/10.4088/JCP.13m08842


http://dx.doi.org/10.1097/01.MLR.0000062554.74615.4C

http://dx.doi.org/10.1080/10503307.2016.1208373

http://dx.doi.org/10.1097/NMD.0000000000000563

http://dx.doi.org/10.1037-7028.36.6.649


http://dx.doi.org/10.1002/cpp.279


http://dx.doi.org/10.1111/ajr.12149


https://doi.org/10.1176/appi.ps.52.12.1615


http://dx.doi.org/10.2196/jmir.2564

http://dx.doi.org/10.2196/jmir.2084.


https://doi.org/10.1016/j.brat.2014.10.007


<table>
<thead>
<tr>
<th>Author(s) and location</th>
<th>Study design</th>
<th>Sample</th>
<th>Intervention</th>
<th>Alliance</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TABLE 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevance information from identified literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Author(s) and location</strong></td>
<td><strong>Study design</strong></td>
<td><strong>Sample</strong></td>
<td><strong>Intervention</strong></td>
<td><strong>Alliance</strong></td>
<td><strong>Outcome</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Bouchard, et al. (2004)</td>
<td>N-RCT</td>
<td>Panic disorder with agoraphobia</td>
<td>VCP: 11</td>
<td>CBT</td>
<td>12</td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Authors</td>
<td>Design</td>
<td>Condition</td>
<td>Treatment</td>
<td>Session</td>
<td>Working Alliance</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------</td>
<td>------------------</td>
</tr>
<tr>
<td>Morland, et al. (2015)</td>
<td>America</td>
<td>PTSD</td>
<td>CPT</td>
<td>2, 6, 12</td>
<td>62.61 (6.95)</td>
</tr>
<tr>
<td>Stefan &amp; David (2013)</td>
<td>America</td>
<td>Mood or anxiety disorder</td>
<td>REBT</td>
<td>1</td>
<td>64.37 (7.85)</td>
</tr>
<tr>
<td>Stubbings, Rees, Roberts, Kane (2013)</td>
<td>Australia</td>
<td>Mood or anxiety disorder</td>
<td>CBT</td>
<td>12</td>
<td>72.42 (10.33)</td>
</tr>
<tr>
<td>Yuen, et al. (2013)</td>
<td>America</td>
<td>Social anxiety</td>
<td>ABBT</td>
<td>12</td>
<td>65.8 (7.24)</td>
</tr>
</tbody>
</table>

As no single measure takes into account insomnia and depression (and the participants in the study were comorbid), change was calculated by summing the pre-intervention outcome measures on the ISI and HRSD, and then subtracted the summed post-intervention outcome measures.


2 Standard deviation estimates calculated from standard error reported in the original article.

3 Standard deviation estimates calculated from confidence intervals reported in the original article.
**TABLE 2**
Quality assessment table for studies identified for review

<table>
<thead>
<tr>
<th>First author (year)</th>
<th>Relevance to review</th>
<th>Clinical population</th>
<th>Design</th>
<th>Recruitment</th>
<th>Inclusion/exclusion stated</th>
<th>Control group (randomised)</th>
<th>Were the groups similar at the start of the trial?</th>
<th>Were the controls selected in an acceptable way?</th>
<th>Aside from the experimental intervention, were the groups treated equally?</th>
<th>Are all outcomes reported?</th>
<th>Were all participants accounted for at the end of the trial?</th>
<th>Post-treatment follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bouchard (2004)</td>
<td>Y</td>
<td>Y</td>
<td>N-RCT</td>
<td>Referral from mental health professional</td>
<td>Y</td>
<td>Y (N)</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Germain (2009)</td>
<td>Y</td>
<td>Y</td>
<td>N-RCT</td>
<td>Treatment waitlists, collaborating psychiatrists, local media</td>
<td>Y</td>
<td>Y (N)</td>
<td>N</td>
<td>UC</td>
<td>UC</td>
<td>Y</td>
<td>UC</td>
<td>N</td>
</tr>
<tr>
<td>Germain (2010)</td>
<td>Y</td>
<td>Y</td>
<td>N-RCT</td>
<td>Treatment waitlists, collaborating psychiatrists, local media</td>
<td>Y</td>
<td>Y (N)</td>
<td>N</td>
<td>UC</td>
<td>UC</td>
<td>Y</td>
<td>UC</td>
<td>N</td>
</tr>
<tr>
<td>Himle (2006)</td>
<td>Y</td>
<td>UC</td>
<td>Case series</td>
<td>University anxiety disorder program</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Lichstein (2013)</td>
<td>Y</td>
<td>UC</td>
<td>Series of case studies</td>
<td>Primary care clinics</td>
<td>Y</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Manchanda (1998)</td>
<td>Y</td>
<td>UC</td>
<td>Case study</td>
<td>General Practitioner</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Mitchell (2008)</td>
<td>Y</td>
<td>Y</td>
<td>RCT</td>
<td>Local physicians &amp; psychologists, local media</td>
<td>Y</td>
<td>Y (Y)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Ertelt (2010)</td>
<td>Y</td>
<td>Y</td>
<td>RCT</td>
<td>Local physicians &amp; psychologists, local media</td>
<td>Y</td>
<td>Y (Y)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Morland (2015)</td>
<td>Y</td>
<td>Y</td>
<td>RCT</td>
<td>Local service providers, local media</td>
<td>Y</td>
<td>Y (Y)</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Stefan (2013)</td>
<td>Y</td>
<td>N</td>
<td>RCT</td>
<td>Undergrad psychology students</td>
<td>N</td>
<td>Y (Y)</td>
<td>UC</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Stubbings (2013)</td>
<td>Y</td>
<td>Y</td>
<td>RCT</td>
<td>Self-referral, or referral from health clinics</td>
<td>Y</td>
<td>Y (Y)</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Yuen (2013)</td>
<td>Y</td>
<td>Y</td>
<td>UCT</td>
<td>Local media and professional</td>
<td>Y</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
referrals

Note. Questions pertaining to blinding were omitted as it would be impossible to blind groups given the nature of the intervention delivery. Y = Yes, N = No, UC = Unclear, N/A = Not applicable. Design: N-RCT = non-randomised controlled trial, RCT = randomised controlled trial, UCT = uncontrolled trial.


FIGURE 1 PRISMA diagram outlining the selection process
FIGURE 2  Forest plot of comparison (sensitivity analysis): VCP versus F2F, outcome: working alliance

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>VCP</th>
<th>F2F</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitchell 2008/Ertelt 2010</td>
<td>219.95</td>
<td>224.63</td>
<td>39</td>
</tr>
<tr>
<td>Morland 2015</td>
<td>68.6</td>
<td>70</td>
<td>49</td>
</tr>
<tr>
<td>Stefan 2013</td>
<td>62.61</td>
<td>64.37</td>
<td>27</td>
</tr>
<tr>
<td>Stubbings 2013</td>
<td>72.42</td>
<td>72.18</td>
<td>10</td>
</tr>
</tbody>
</table>

Total (95% CI): 121

Heterogeneity: Tau² = 0.06; Chi² = 5.71, df = 3 (P = 0.13); I² = 47%

Test for overall effect: Z = 1.61 (P = 0.11)

Note. VCP = videoconferencing psychotherapy; F2F = face-to-face; CI = confidence interval. Ertelt 2010 reports working alliance data on the same dataset as Mitchell 2008 reports outcome data.
FIGURE 3 Forest plot of comparison (sensitivity analysis): VCP versus F2F, outcome: symptom reduction.

Note. VCP = videoconferencing psychotherapy; F2F = face-to-face; CI = confidence interval. Mitchell 2008 reports outcome data on the same dataset as Ertelt 2010 reports working alliance data.